DEG 15, an update of the Database of Essential Genes that includes built-in analysis tools

Hao Luo \(^1\), Yan Lin \(^1\), Tao Liu \(^1\), Fei-Liao Lai \(^1\), Chun-Ting Zhang \(^1\), Feng Gao \(^{1,2,*}\) and Ren Zhang \(^{3,*}\)

\(^1\)Department of Physics, School of Science, Tianjin University, Tianjin 300072, China, \(^2\)Frontiers Science Center for Synthetic Biology and Key Laboratory of Systems Bioengineering (Ministry of Education), Tianjin University, Tianjin 300072, China and \(^3\)Center for Molecular Medicine and Genetics, School of Medicine, Wayne State University, Detroit, MI 48201, USA

Received September 15, 2020; Revised September 30, 2020; Editorial Decision October 01, 2020; Accepted October 06, 2020

ABSTRACT

Essential genes refer to genes that are required by an organism to survive under specific conditions. Studies of the minimal-gene-set for bacteria have elucidated fundamental cellular processes that sustain life. The past five years have seen a significant progress in identifying human essential genes, primarily due to the successful use of CRISPR/Cas9 in various types of human cells. DEG 15, a new release of the Database of Essential Genes (www.essentialgene.org), has provided major advancements, compared to DEG 10. Specifically, the number of eukaryotic essential genes has increased by more than fourfold, and that of prokaryotic ones has more than doubled. Of note, the human essential-gene number has increased by more than tenfold. Moreover, we have developed built-in analysis modules by which users can perform various analyses, such as essential-gene distributions between bacterial leading and lagging strands, sub-cellular localization distribution, enrichment analysis of gene ontology and KEGG pathways, and generation of Venn diagrams to compare and contrast gene sets between experiments. Additionally, the database offers customizable BLAST tools for performing species- and experiment-specific BLAST searches. Therefore, DEG comprehensively harbors updated human-curated essential-gene records among prokaryotes and eukaryotes with built-in tools to enhance essential-gene analysis.

INTRODUCTION

Essential genes refer to genes required for a cell or an organism to survive under certain conditions (1,2). The research on the determination of essential genes has attracted significant attention in the past decade, due to its theoretical implications and practical uses. Studies of genome-wide gene essentiality screenings have elucidated fundamental cellular processes that sustain life (2). We created DEG, a Database of Essential Genes in 2003 (3), a time when the genome-scale gene essentiality screening was still not available. The development of DEG parallels with the development of the essential-gene field. Significant progress has been made in performing genome-wide essentiality screenings among diverse species, primarily due to technological developments. We subsequently published DEG 5, which included essential genes of both bacteria and eukaryotes (4), and DEG 10, which included both protein-coding genes and non-coding genomic elements (5). Since 2014, when DEG 10 was published (5), significant progress has been made mainly owing to the invention of CRISPR/Cas9 (6,7) and the widespread use of Tn-seq (8,9). To accommodate the progress in essential-gene studies, we created DEG 15, which, compared to DEG 10, provides two major updates:

1. The number of essential-gene entries has significantly increased. Specifically, compared to DEG 10, the number of eukaryotic essential genes has increased by more than fourfold, and that of prokaryotic ones has more than doubled. Of note, the human essential-gene number has increased by more than ten-fold. Figure 1 shows the number of essential gene records in different versions of DEG, as well as the methods used to determine the gene essentiality. It is shown that the increase in prokaryotic records and eukaryotic records are mainly due to the widespread use of Tn-seq and CRISPR/Cas9, respectively (Figure 1).

2. We have developed built-in analysis modules by which users can perform various analyses, such as essential-gene distributions between bacterial leading and lagging strands, sub-cellular localization distribution, gene on-
DETERMINATION OF ESSENTIAL GENES IN HUMANS

Genome-wide essentiality screenings have elucidated the molecular underpinnings of many biological processes in prokaryotes. However, limited knowledge has been gained regarding essential genes in human cells. Large-scale gene essentiality screenings across human cell types can reveal genes that encode factors for regulating tissue-specific cellular processes, and such screenings in cancer cells can disclose factors that determine cancer phenotypes, thus revealing important targets for cancer therapies. However, genome-wide inactivation of genes in human cells and the analysis of lethal phenotypes have been hampered by technical barriers.

One of the major breakthroughs in biotechnology has been the invention of CRISPR/Cas9 (CRISPR-associated RNA-guided endonuclease Cas9), which is a simple yet powerful tool for editing genomes (6,7). Cas9, an endonuclease, can be guided to specific locations within complex genomes by a guide RNA (gRNA). Cas9-mediated gene editing is simple and scalable, enabling the examination of gene functions at the systems level. Because of the ease and efficient targeting, CRISPR/Cas9 is described as being analogous to the ‘search’ function in a modern word processor (10). The invention of CRISPR/Cas9 has revolutionized the biological research in many fields, with essentiality screenings in human cells being no exception.

In 2015, three papers were published simultaneously reporting the genome-wide identification of essential genes among diverse human cell types (11–13). Wang et al. used the CRISPR-based approaches in analyzing multiple cell lines, and found tumor-specific dependencies on particular genes. The core-essential genes among these cell lines are enriched for genes with evolutionarily conserved pathways, with high expression levels, and with few detrimental polymorphisms in the human population (11). Analysis by Blomen et al. revealed a synthetic lethality map in human cells (12). Hart et al. used CRISPR-based approaches to screen for fitness genes among five cell lines, and consequently discovered 1580 human core fitness genes, and context-dependent fitness genes, that is, genes conferring pathway-specific genetic vulnerabilities in cancer cells (13).

The technology of CRISPR can be used on various cell types. Mair et al. used the CRISPR system to catalogue essential genes that are indispensable for human pluripotent stem cell fitness (14). Lu et al. determined genes essential for podocyte cytoskeletons based on single-cell RNA sequencing (15). Wang et al. used CRISPR in identifying essential oncogenes for hepatocellular carcinoma tumor growth (16). Arroyo et al. used a CRISPR-based screen and consequently identified essential genes for oxidative phosphorylation (17).

There is a major difference between cell-specific and organism-specific gene essentiality. That is, essential gene sets for human cells can be significantly different from those for human development. CRISPR technology, despite being powerful, cannot be used as a reverse genetics approach in humans for gene essentiality studies. Nevertheless, exome sequencing, another recent breakthrough, enables the identification of human essential genes in vivo (18).

Exome sequencing is considerably less expensive than whole-genome sequencing, and most Mendelian diseases are caused by genetic variations in protein-coding regions (exomes). The Exome Aggregation Consortium (ExAC) reported the exome sequences of 60,706 individuals, and the genetic diversity represents an average of one variant of every 8 bases of the exomes. Thus these variations are analogous to a genome-wide mutagenesis screening conducted in nature, similar to a transposon mutagenesis screening performed in the lab. Strikingly, 3230 genes contain near-complete depletion of protein-truncating variants, representing candidate human organism-level essential genes (18). Therefore, the number of essential gene in humans in DEG15 has increased by >10-fold, primarily due to the use of CRISPR and exome sequencing technology (Table 1).

THE WIDESPREAD USE OF Tn-seq

Tn-seq technology has been successfully used in identifying essential genes in a large number of bacteria, and it has also been used in archaea and even a eukaryote.
Table 1. Contents of DEG 15

| Domain of life | Organism | No. of essential genomic elements | Method | Saturated | Reference | Note* |
|----------------|----------|----------------------------------|--------|-----------|-----------|-------|
|                |          | Coding                           | Noncoding |           |           |       |
|                | Acinetobacter baumannii ATCC 17978 | 453 | 59 | INSeq | Yes | (24) |
|                | Acinetobacter baumannii ATCC 17978 | 157 | 1 | INSeq | Yes | (24) | In the mouse lung |
|                | Acinetobacter baylyi | 499 | Single-gene knockout | Trn-seq^ | Yes | (57) | Minimal medium |
|                | Aggregatibacter actinomycetemcomitans | 59 | Trn-seq^ | Yes | (58) | For coinfection with sympatric and allopatric microbes |
|                | Agrobacterium fabrum str. C58 | 361 | 11 | Trn-seq | Yes | (25) |
| Bacteria       | Bacillus subtilis | 261 | 2 | Single-gene knockout | Yes | (59) |
|                | Bacillus thuringiensis BMB171 | 516 | Tn-seq | Yes | (60) |
|                | Bacteroides fragilis | 550 | Tn-seq | Yes | (61) |
|                | Bacteroides thetaiotaomicron | 325 | INSeq | Yes | (21) |
|                | Bifidobacterium breve | 453 | TraDIS | Yes | (62) |
|                | Brevundimonas subtilis | 448 | Tn-seq | Yes | (25) |
|                | Burkholderia cenocepacia J2315 | 383 | TraDIS | Yes | (63) |
|                | Burkholderia cenocepacia K56–2 | 508 | Tn-seq | Yes | (64) |
|                | Burkholderia pseudomallei K96243 | 505 | TraDIS | Yes | (65) |
|                | Burkholderia thailandensis | 406 | Tn-seq | Yes | (66) |
|                | Campylobacter jejuni | 233 | Tn-seq | Yes | (67) |
|                | Campylobacter jejuni subsp. jejuni 81–176 | 384 | Tn-seq | Yes | (68) |
|                | Campylobacter jejuni subsp. jejuni NCTC 11168 | 166 | Tn-seq | Yes | (68) |
|                | Caulobacter crescentus | 480 | 532 | Tn-seq | Yes | (69) |
|                | Escherichia coli | 620 | Single-gene knockout | Genetic footprinting | Yes | (70) |
|                | Escherichia coli | 303 | MATT | Yes | (71) |
|                | Escherichia coli | 379 | CRISPR | Yes | (72) |
|                | Escherichia coli O157:H7 | 1265 | 37 | Tn-seq | Yes | (73) |
|                | Escherichia coli ST131 strain EC958 | 315 | TraDIS | Yes | (74) |
|                | Francisella novicida | 396 | Tn-seq | Yes | (75) |
|                | Francisella tularensis Schu S4 | 453 | TraDIS | Yes | (76) |
|                | Haemophilus influenzae | 667 | Single-gene knockout | Genetic footprinting | Yes | (77) |
|                | Helicobacter pylori | 344 | MATT | Yes | (78) |
|                | Mycobacterium avium subsp. hominis strain MAC109 | 230 | Tn-seq | Yes | (79) |
|                | Mycobacterium tuberculosis | 614 | Tn-seq | Yes | (80) |
|                | Mycobacterium tuberculosis | 774 | Tn-seq | Yes | (81) |
|                | Mycobacterium tuberculosis | 742 | 35 | Tn-seq | Yes | (82) |
|                | Mycoplasma genitalium | 461 | Tn-seq | Yes | (83) |
|                | Mycoplasma pneumoniae | 382 | Tn-seq | Yes | (84) |
|                | Mycoplasma pulmonis | 342 | 34 | Tn-seq | Yes | (85) |
|                | Neisseria gonorrhoeae MS11 | 751 | Tn-seq | Yes | (86) |
|                | Porphyromonas gingivalis | 463 | Tn-seq | Yes | (87) |
|                | Porphyromonas gingivalis | 281 | Tn-seq | Yes | (88) |
|                | ATCC 33277 | 496 | 25 | Tn-seq | Yes | (89) |
|                | Providencia stuartii strain BE3467 | 335 | Tn-seq | Yes | (90) |
|                | Pseudomonas aeruginosa | 117 | Tn-seq | Yes | (91) |
|                | Pseudomonas aeruginosa | 321 | Tn-seq | Yes | (92) |
|                | Pseudomonas aeruginosa | 336 | Tn-seq | Yes | (93) |
| Domain of life | Organism | No. of essential genomic elements | Method | Saturated | Reference | Note(s) |
|---------------|----------|----------------------------------|--------|-----------|-----------|---------|
|               |          | Coding | Noncoding |          |           |         |
|                | *Ralstonia solanacearum* | 465 | Tn-seq | Yes | (93) |
|                | GM11000 |        |         |      |           |         |
|                | *Rhodobacter sphaeroides* | 493 | Tn-seq | Yes | (94) |
|                | CGA009 |        |         |      |           |         |
|                | *Salmonella enterica Typhimurium* | 306 | 15 | TraDIS | Yes | (29) |
|                | *Salmonella enterica serovar Typhi* | 356 |      | TraDIS | Yes | (20) |
|                | *Salmonella enterica serovar Typhi Ty2* | 358 | 24 | TraDIS | Yes | (29) |
|                | *Salmonella enterica serovar Typhimurium* | 105 |      | Tn-seq | Yes | (96) |
|                | *Salmonella enterica serovar Typhimurium SL1344* | 353 | 23 | TraDIS | Yes | (29) |
|                | *Salmonella typhimurium* | 490 | Insertion-duplication | No | (97) |
|                | *Shewanella oneidensis* | 403 | Transposon mutagenesis | Yes | (98) |
|                | *Sphingomonas wittichii* | 579 | 32 | Tn-seq | Yes | (30) |
|                | *Staphylococcus aureus* | 302 | Antisense RNA | No | (99, 100) |
|                | *Staphylococcus aureus* | 351 | TMDH | Yes | (101) |
|                | *Staphylococcus aureus subsp. aureus MRS2* | 295 | Tn-seq | Yes | (102) |
|                | *Staphylococcus aureus subsp. aureus MSA476* | 305 | Tn-seq | Yes | (102) |
|                | *Staphylococcus aureus subsp. aureus MW2* | 256 | Tn-seq | Yes | (102) |
|                | *Staphylococcus aureus subsp. aureus NCTC 8325* | 288 | Tn-seq | Yes | (102) |
|                | *Staphylococcus aureus subsp. aureus USA300 TCH1516* | 385 | Tn-seq | Yes | (102) |
|                | *Streptococcus agalactiae A909* | 317 | Tn-seq | Yes | (103) |
|                | *Streptococcus mitis UA159* | 197 | 6 | Tn-seq | Yes | (104) |
|                | *Streptococcus pneumoniae* | 113 | InsertionDuplication | No | (105) |
|                | *Streptococcus pneumoniae* | 133 | Allelic replacement mutagenesis | Yes | (106) |
|                | *Streptococcus pneumoniae* | 72 | Tn-seq | Yes | (31) |
|                | *Streptococcus pyogenes* | 227 | Tn-seq | Yes | (107) |
|                | *Streptococcus pyogenes MGAS448* | 241 | Tn-seq | Yes | (107) |
|                | *Streptococcus pyogenes NZ131* | 218 | Single-gene knockout | Yes | (108) |
|                | *Streptococcus suis* | 361 | Tn-seq | Yes | (109) |
|                | *Streptococcus suis PCC 7942* | 682 | 34 | Tn-seq | Yes | (110) |
|                | *Vibrio cholerae* | 789 | Tn-seq | Yes | (111) |
|                | *Vibrio cholerae C6706* | 343 | Tn-seq | Yes | (112) |
|                | *Vibrio vulnificus* | 316 | Tn-seq | Yes | (113) |
|                | *Methanococcus maripaludis* | 519 | Tn-seq | Yes | (32) |
| Archaea        | *Sulfobobas islandicus M.16.4* | 441 | Tn-seq | Yes | (33) |
| Eukaryotes     | *Arabidopsis thaliana* | 358 | Single-gene knockout | No | (54) |
|                | *Aspergillus fumigatus* | 35 | Conditional promoter replacement | No | (114) |
|                | *Bombyx mori* | 1006 | CRISPR | Yes | (115) |
|                | *Caenorhabditis elegans* | 44 | Genetic mapping | No | (116) |
|                | *Caenorhabditis elegans* | 294 | RNA interference | No | (56) |
|                | *Danio rerio* | 315 | Insertional mutagenesis | No | (117) |
|                | *Drosophila melanogaster* | 376 | P element insertion | No | (118) |
Table 1. Continued

| Domain of life | Organism     | No. of essential genomic elements | Method                          | Saturated | Reference | Note a |
|---------------|--------------|-----------------------------------|---------------------------------|-----------|-----------|--------|
|               |              | Coding                            | Noncoding                       |           |           |        |
| Homo sapiens  | 2452         | OMIM annotation                    | No                               | (119)     |           |        |
| Homo sapiens  | 1562         | CRISPR                           | Yes                              | (14)      | Stem cells|        |
| Homo sapiens  | 1593         | CRISPR                           | Yes                              | (14)      | HAP1 cells|        |
| Homo sapiens  | 1690         | CRISPR                           | Yes                              | (120)     | Core essential genes among 17 cell lines| |
| Homo sapiens  | 3230         | Exome sequencing                  | Yes                              | (18)      |           |        |
| Homo sapiens  | 2054         | CRISPR                           | Yes                              | (12)      | KBM7 cells|        |
| Homo sapiens  | 2181         | CRISPR                           | Yes                              | (12)      | HAP1 cells|        |
| Homo sapiens  | 1878         | CRISPR                           | Yes                              | (13)      | KBM7 cells|        |
| Homo sapiens  | 1660         | CRISPR                           | Yes                              | (11)      | K562 cells|        |
| Homo sapiens  | 1630         | CRISPR                           | Yes                              | (11)      | Jiyo cells|        |
| Homo sapiens  | 1461         | CRISPR                           | Yes                              | (11)      | Raji cells|        |
| Homo sapiens  | 1196         | CRISPR                           | Yes                              | (13)      | A375 cells|        |
| Homo sapiens  | 1892         | CRISPR                           | Yes                              | (13)      | DLD1 cells|        |
| Homo sapiens  | 2196         | CRISPR                           | Yes                              | (13)      | G8M cells |        |
| Homo sapiens  | 2073         | CRISPR                           | Yes                              | (13)      | HCT116 cells| |
| Homo sapiens  | 386          | shRNA                             | Yes                              | (13)      | HCT116 cells| |
| Homo sapiens  | 1696         | CRISPR                           | Yes                              | (13)      | HeLa cells|        |
| Homo sapiens  | 2038         | CRISPR                           | Yes                              | (15)      | RPE1 cells|        |
| Homo sapiens  | 92           | Functional genomics               | No                               | (15)      | Podocytes |        |
| Homo sapiens  | 79           | CRISPR                           | Yes                              | (16)      | Hepatocellular carcinoma| |
| Homo sapiens  | 191          | CRISPR                           | Yes                              | (17)      | K562 cells|        |
| Komagatella phaffii GS15 | 753 | Tn-seq                            | Yes                              | (121)     |           |        |
| Mus musculus  | 435          | Single-gene knockout              | No                               | (53)      | Embryonic lethality| |
| Mus musculus  | 1933         | Single-gene knockout              | No                               | (52)      | Preweaning lethality| |
| Mus musculus  | 2136         | MGI annotation                    | No                               | (122)     |           |        |
| Plasmodium falciparum | 2680 | transposon mutagenesis            | Yes                              | (34)      |           |        |
| Saccharomyces cerevisiae | 1110 | Single-gene knockout               | Yes                              | (123)     | Six conditions including minimal medium| |
| Schizosaccharomyces pombe | 1260 | Single-gene knockout               | Yes                              | (124)     | Rich medium| |

*a Bacteria were cultured in rich media, unless otherwise indicated.

*bTn-seq is a method that performs saturated transposon mutagenesis followed by parallel sequencing to determine the transposon integration sites. Tn-seq has many variants under different names, such as insertion sequencing (INSeq), Transposon Directed Insertion Sequencing (TraDIS), high-throughput insertion tracking by deep sequencing (HTTS), transposon sequencing, Microarray tracking of transposon mutants (MATT), Transposon site hybridization (TraSH), transposon mutagenesis followed by Sanger sequencing, transposon mutagenesis followed by genetic footprinting, transposon-site hybridization, Transposon-Mediated Differential Hybridisation (TMDH).

cOMIM: Online Mendelian Inheritance in Man (125).

dMGI: Mouse Genome Informatics (126).

parison to the single gene knockout method, Tn-seq is less time-consuming and labor-intensive, because of the parallel nature in mutagenesis and insertion site determination. The invention of the Tn-seq method can date back to a study in which Venter and coworkers performed Sanger sequencing to determine transposon integration sites (19) in 1999. In 2009, two technologies, high-density transposon-mediated mutagenesis and high-throughput sequencing, were mature, creating conditions that enabled Tn-seq to be invented (9).

Many variants of Tn-seq were proposed, such as TraDIS (20), INSeq (21), HTTS (22) and Tn-seq Circle (23). Here, we refer to these methods collectively as Tn-seq since they all involve transposon mutagenesis and sequencing.

Tn-seq has been widely used in identifying essential genes in bacteria. Figure 1A shows that since 2009, when DEG 5 was published (4), most bacterial essential genes have been determined by Tn-seq, and the proportion of essential genes that are determined by Tn-seq has been increasing ever since. This is not surprising given the powerfulness, ease of use, and the efficiency of Tn-seq in performing essentiality screening. Another advantage of Tn-seq is that it identifies not only essential protein-coding genes, but also non-coding genomic elements. For instance, by using Tn-seq, a large number of non-coding genomic elements have been determined in Acinetobacter baumannti (24), Brevundimonas subvibrioides (25), Escherichia coli O157:H7 (26), Mycobacterium tuberculosis (27), Mycoplasma pneumonia (28), Salmonella enterica serovar Typhimurium (29), Sphingomonas wittichii (30) and Streptococcus pneumonia (31).

In addition, Tn-seq has been used to determine essential genes in species other than bacteria. The methanogenic archaeon Methanococcus maripaludis S2 is an obligate anaer-
obic prokaryote that lives in oxygen-free environments. Sarmiento et al. used the Tn-seq method and identified 526 essential genes required for growth in rich medium, representing the first genome-wide gene essentiality screening in archaea (32). The second essentiality screening in archaea was conducted in Sulfolobus islandicus, and some archaea specific essential genes were identified (33). Moreover, Tn-seq was also used in identifying essential genes in a eukaryote. Severe malaria is caused by the apicomplexan parasite Plasmodium falciparum, a unicellular protozoan parasite of humans, and 680 genes were identified as essential for optimal growth of this parasite (34). Because of the widespread use of Tn-seq, the number of prokaryotic essential genes in DEG 15 has more than doubled compared to that of DEG 10 (Figure 1A).

ANALYSIS MODULES

To facilitate the use of DEG, we developed a set of analysis modules in the current release. Essential genes are preferentially situated in the leading strand, rather than the lagging strand (35), mainly because of the decreased mutagenesis pressure resulting from the head-on collisions of transcription and replication machineries in the leading strand (36). We obtained replication origins and determined leading vs. lagging strands using the DoriC database (37,38). Users can examine essential gene distributions between leading and lagging strands, and clicking the pie graph will display a list of genes in leading or lagging strands (Figure 2A).

Sub-cellular localization and operon information were obtained from the PSORTb v3.0 tool and the DOOR database, respectively (39,40). Clicking a species name, e.g. Bacillus subtilis, will display sub-cellular localization distributions of essential genes, and detailed gene information can be further examined by clicking on a particular cell compartment (Figure 2B). Other information includes orthologous groups, EC number (41), KEGG pathway (42) and GO (43), as determined by eggNOG-mapper (44). Users can analyze the GO distributions, and enriched GO terms powered by GOATool (45), and enriched KEGG pathways, obtained using clusterProfiler package in R language (46). The analysis results, including strand bias distribution, sub-cellular distribution, and enrichment analysis of GO and KEGG pathways, are visualized with ECharts (47).

To analyze human essential genes, we developed a tool by which users can compare and contrast the essential gene sets between experiments, generate Venn diagrams to visualize the comparison, and obtain unions and intersections for the two gene sets by clicking the corresponding graph (Figure 2C). Furthermore, DEG 15 continues to provide customizable BLAST tools that allow users to perform species- and experiment-specific searches for a single gene, a list of genes, annotated or un-annotated genomes.

FUTURE PERSPECTIVE

The identification of essential genes in both prokaryotes and eukaryotes has attracted significant attention over the past decade, largely because of the practical implications of these studies (2). Bacterial essential genes are attractive drug targets, as inhibiting these genes can suppress bacterial survival (48). Interest on essentiality screenings has also been boosted by synthetic biology, which aims to make an artificial self-sustainable living cell (49). The minimal gene set of a bacterium is considered a chassis for further addi-
tion of other parts with desirable traits. An increasing number of essentiality screens are being performed in a context-specific manner. For instance, essential genes for cancer cells can reveal cancer-specific cellular processes, which are targets for cancer drugs (50). Determination of essential genes of A. baumannii revealed genes required for its infection and survival in the lung (24). Moreover, another important direction is the prediction of gene essentiality using bioinformatic approaches, e.g., based on metabolic models (51). Therefore, because of the theoretical implications of the minimal-gene-set concept and its practical uses, it is expected that the essential gene identifications will continue to be further advanced.

Reverse genetics will continue to be indispensable for pinpointing gene functions. It is expected that single-gene knockout projects for the model organisms, such as mice (52,53) and Arabidopsis thaliana (54), will soon be completed. Multiple ways to manipulate gene expression are available, such as those based on TetR/Pip-OFF repressible promoter system (55) and RNA interference (56). From the aspect of technology, this is a golden era for essential-gene research, because of the availability of Tn-seq and CRISPR/Cas9. The two technologies enable the gene essentiality screenings in a wide range of cell types and species under diverse conditions. Therefore, we anticipate that the increase in the number of essential genes for many cell types under various conditions will be accelerated in the future. Therefore, we will continue to update DEG with high-quality human-curated data in a timely manner to keep pace with this rapidly developing field.

DATA AVAILABILITY

DEG is accessible from essentialgene.org or tubic.org/deg. All DEG data is freely available to download.

FUNDING

National Key Research and Development Program of China [2018YFA0903700 to F.G.] (in part); National Natural Science Foundation of China [31801104 to H.L., 31571358 to F.G., 31200991 to Y.L.]. Funding for open access charge: National Natural Science Foundation of China [31571358 to F.G.].

Conflict of interest statement. None declared.

REFERENCES

1. Koonin,E.V. (2000) How many genes can make a cell: the minimal-gene-set concept. *Annu. Rev. Genomics Hum. Genet.*, 1, 99–116.
2. Bartha,I., di Iulio,J., Venter,J.C. and Teleni,A. (2018) Human gene essentiality. *Nat. Rev. Genet.*, 19, 51–62.
3. Zhang,R., Ou,H.Y. and Zhang,C.T. (2004) DEG: a database of essential genes. *Nucleic Acids Res.*, 32, D271–D272.
4. Zhang,R. and Lin.Y. (2009) DEG 3.0, a database of essential genes in both prokaryotes and eukaryotes. *Nucleic Acids Res.*, 37, D455–D458.
5. Luo,H., Lin,Y., Gao,F., Zhang,C.T. and Zhang,R. (2014) DEG 10, an update of the database of essential genes that includes both protein-coding genes and noncoding genomic elements. *Nucleic Acids Res.*, 42, D574–D580.
6. Cong,L., Ran,F.A., Cox,D., Lin,S., Barretto,R., Habib,N., Hsu,P.D., Wu,X., Jiang,W., Marraffini,L.A. et al. (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science*, 339, 819–823.
7. Jinek,M., Chylinski,K., Fonfara,I., Hauer,M., Doudna,J.A. and Charpentier,E. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, 337, 816–821.
8. Barquist,L., Boineau,C.J. and Cain,A.K. (2013) Approaches to querying bacterial genomes with transposon-insertion sequencing. *RNA Biol.*, 10, 1161–1169.
9. van Opijnen,T., Bodi,K.L. and Camilli,A. (2009) Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nat. Methods*, 6, 767–772.
10. Hsu,P.D., Lander,E.S. and Zhang,F. (2014) Development and applications of CRISPR-Cas9 for genome engineering. *Cell*, 157, 1262–1278.
11. Wang,T., Birsoy,K., Hughes,N.W., Krupczak,K.M., Post,Y., Wei,J.J., Lander,E.S. and Sabatini,D.M. (2015) Identification and characterization of essential genes in the human genome. *Science*, 350, 1096–1101.
12. Blomen,V.A., Majek,P., Jae,L.T., Bignenzahn,J.W., Nieuwenhuis,J., Staring,J., Sacco,R., van Diemen,F.R., Okt,N., Stukalov,A. et al. (2015) Gene essentiality and synthetic lethality in haploid human cells. *Science*, 350, 1092–1096.
13. Hart,T., Chandrashekhar,M., Aregger,M., Steinhart,Z., Rawls,J.F., Krug,M., Brehm,R., Karaman,L. and Bickel,J. (2017) Genome-wide CRISPR knockout screens identify NAPCEP1 and TP53 as essential genes for enterohemorrhagic E. coli persistence in the murine gut. *Cell Host. Microbe.*, 21, 432–445.
14. Lu,Y., Ye,Y., Bao,W., Yang,Q., Wang,J., Liu,Z. and Shi,S. (2017) Genome-wide identification of genes essential for podocyte cytoskeletons based on single-cell RNA sequencing. *Kidney Int.*, 92, 1119–1129.
15. Wang,Y., Gao,B., Tan,P.Y., Handoko,Y.A., Sekar,K., Devisagiamani,A., Seshachalam,V.P., OuYang,H.Y., Shi,M., Xie,C. et al. (2019) Genome-wide CRISPR knockout screens identify NCAPO as an essential oncogene for hepatocellular carcinoma tumor growth. *FASEB J.*, 33, 8759–8770.
16. Arroyo,J.D., Jourdain,A.A., Calvo,S.E., Ballarano,C.A., Cummings,B.B. and Andersen,J.T. (2019) Genome-wide CRISPR death screen identifies genes essential for oxidative phosphorylation. *Cell Metab.*, 24, 875–885.
17. Mair,R., Tomic,J., Masud,S.N., Tonge,P., Weiss,A., Usaj,M., Tong,A.H.Y., Kwan,J.J., Brown,K.R., Titus,E. et al. (2019) Essential gene profiles for human pluripotent stem cells identify uncharacterized genes and substrate dependencies. *Cell Rep.*, 27, 599–615.
18. Wang,Y., Gao,B., Tan,P.Y., Handoko,Y.A., Sekar,K., Devisagiamani,A., Seshachalam,V.P., OuYang,H.Y., Shi,M., Xie,C. et al. (2019) Genome-wide CRISPR knockout screens identify NCAPO as an essential oncogene for hepatocellular carcinoma tumor growth. *FASEB J.*, 33, 8759–8770.
19. Kulbokas,E.J., Lander,E.S. and Miller,W.E. (2010) Mapping and quantifying expression levels of transcription factors using whole-genome tiling arrays. *Proc. Natl. Acad. Sci. U.S.A.*, 107, 12016–12021.
20. Langridge,G.C., Phan,M.D., Turner,D.J., Perkins,T.T., Parts,L., Haase,J., Charles,I., Maskell,D.J., Peters,S.E., Dougan,G. et al. (2009) Simultaneous assay of all human genes using DdSR to identify essential genes. *Nature*, 460, 106–110.
21. Goodman,A.L., McNulty,N.P., Zhao,Y., Leip,D., Mitra,R.D., Louzopone,C.A., Knight,R. and Gordon,J.I. (2009) Identifying genetic determinants needed to establish a human gut symbiont in its habitat. *Cell Host. Microbe.*, 6, 279–289.
22. Gawronski,J.D., Wong,S.M., Giannoukos,G., Ward,D.V. and Akeryler,B.J. (2009) Tracking insertion mutants within libraries by deep sequencing and a genome-wide screen for Haemophilus genes required in the lung. *Proc. Natl. Acad. Sci. U.S.A.*, 106, 16422–16427.
23. Walling,G., Stirling,J., Sacco,R., van Diemen,F.R., Okt,N., Stukalov,A. et al. (2009) Simultaneous assay of every Salmonella Typhi gene using one million transposon mutants. *Science*, 325, 713–735.
24. Wang,N., Ozer,E.A., Mandel,M.J. and Hauser,A.R. (2014) Genome-wide identification of Acinetobacter baumannii genes necessary for persistence in the lung. *mBio*, 5, e00315-10.
25. Curtis,P.D. and Brun,Y.V. (2014) Identification of essential alphaproteobacterial genes reveals operational variability in conserved developmental and cell cycle systems. *Mol. Microbiol.*, 93, 713–735.

Downloaded from https://academic.oup.com/nar/article/49/D1/D677/5937083 by guest on 22 July 2021
Burkholderia cepaciae J2315 identified by genome-wide TraDIS. Front Microbiol., 7, 1289.

64. Gislason, A.S., Turner, K., Domaratzki, M. and Cardona, S.T. (2017) Comparative analysis of the Burkholderia cepaciae K56-2 essential genome reveals cell envelope functions that are uniquely required for survival in species of the genus Burkholderia. Microb Genom., 3, e001160.

65. Moule, M.G., Hemsley, C.M., Seet, Q., Guerra-Assuncao, J.A., Lim, J., Sarkar-Taylor, M., Clark, T.G., Tan, P.B., Titball, R.W., Cucuji, et al. (2014) Genome-wide saturation mutagenesis of Burkholderia pseudomallei K92643 predicts essential genes and novel targets for antimicrobial development. mBio, 5, e00926-13.

66. Baugh, L., Gallagher, L.A., Patrapuvich, R., Clifton, M.C., Gardberg, A.S., Edwards, T.E., Armour, B., Begley, D.W., Dieterich, S.H., Dranow, D.M., et al. (2013) Combining functional and structural genomics to sample the essential Burkholderia structure. PLoS One, 8, e53851.

67. Metris, A., Reuter, M., Gaskin, D.J., Baranyi, J. and van Vliet, A.H. (2016) Essentiality analysis of the Metcoccus bacterium. BMC Genomics, 17, 616.

68. Christen, B., Abeliuk, E., Collier, J.M., Kalogeraki, V.S., Passarelli, B., Christen, B., Abeliuk, E., Collier, J.M., Kalogeraki, V.S., Passarelli, B., et al. (2007) A comprehensive transposon mutant library clone. Proc. Natl. Acad. Sci. U.S.A., 104, 1009–1014.

69. Ireland, P.M., Bullifent, H.L., Senior, N.J., Southern, S.J., Yang, Z.R., Gallagher, L.A., Ramage, E., Jacobs, M.A., Kaul, R., Brittnacher, M., et al. (2013) The serum resistome of a globally disseminated multidrug resistant uropathogenic Escherichia coli clone. Proc. Natl. Acad. Sci. U.S.A., 110, 10097–10102.

70. Akerley, B.J., Rubin, E.J., Novick, V.L., Osmulski, S.J., White, A.N., Zhao, L., Wu, M., Whiteley, M., Lamont, R.J. and Scott, D.A. (2016) Comparison of inherently essential genes of Porphromonas gingivalis identified in two transposon-sequencing libraries. Mol. Microbiol., 99, 354–364.

71. Gallagher, L.A., Ramage, E., Jacobs, M.A., Kaup, R., Brittnacher, M. and Manoil, C.S. (2007) A high-throughput essentiality test to identify essential genes, fitness factors for catheter-associated urinary tract infection, and the impact of microbial infection on fitness requirements. mSphere, 5, e00142-20.

72. Lee, S.A., Gallagher, L.A., Thongdee, M., Staudinger, B.J., Lippman, S., Singh, P.K. and Manoil, C. (2015) General and condition-specific essential functions of Pseudomonas aeruginosa. Proc. Natl. Acad. Sci. U.S.A., 112, 4110–4115.

73. Lee, S.A., Gallagher, L.A., Thongdee, M., Staudinger, B.J., Lippman, S., Singh, P.K. and Manoil, C. (2015) General and condition-specific essential functions of Pseudomonas aeruginosa. Proc. Natl. Acad. Sci. U.S.A., 112, 5189–5194.

74. Arkin, A.P. (2011) Evidence-based annotation of gene function in Mycobacterium tuberculosis. Proc. Natl. Acad. Sci. U.S.A., 108, 535–540.

75. Shewanella oneidensis MR-1 using genome-wide fitness profiling mSphere, 3, 3098–3107.

76. Burger, B.T., Imam, S., Scarborough, M.J., Noguera, D.R. and Donohue, T.J. (2017) Combining genome-scale experimental and computational methods to identify essential genes in Rhodobacter sphaeroides. mSystems, 2, e00213-16.

77. Knuth, K., Niesalla, H., Hueck, C.J. and Fuchs, T.M. (2004) Essentiality and condition-specific essential functions of Pseudomonas aeruginosa in cystic fibrosis sputum. Proc. Natl. Acad. Sci. U.S.A., 101, 1729–1744.

78. McFadden, A., Price, M.N., Wetmore, K.M., Shao, W., Baumohl, J.K., Xu, Z., Nguyen, M., Tanase, R., Davis, R.W. and Arkin, A.P. (2011) Evidence-based annotation of gene function in Shewanella oneidensis MR-1 using genome-wide fitness profiling across 121 conditions. mSystems, 7, e00235-16.

79. Forsyth, R.A., Haselbeck, R.J., Olshen, K.L., Yamamoto, R.T., Xu, H., Trawick, J.D., Wall, D., Wang, L., Brown-Driver, V., Froelich, J.M. et al. (2002) A genome-wide strategy for the Mycobacterium tuberculosis genome via saturating transposon mutagenesis. mBio, 8, e02133-16.

80. Minato, Y., Gohi, D.M., Thiede, J.M., Chacon, J.M., Harcombe, W.R., Maruyama, F. and Baughn, A.D. (2019) Genome-wide assessment of Mycobacterium tuberculosis conditionally essential metabolic pathways. mSystems, 4, e00070-19.

81. Glass, J.J., Assad-Garcia, N., Alperovich, N., Vooosph, S., Lewis, M.R., Maruf, M., Hutchison, C.A., Smith, H.O. and Venter, J.C. (2006) Essential genes of a minimal bacterium. Proc. Natl. Acad. Sci. U.S.A., 103, 425–430.

82. French, C.T., Lao, P., Loraine, A.E., Matthews, B.T., Yu, H. and Dybvig, K. (2008) Large-scale transposon mutagenesis of Mycobacterium tuberculosis identified by transcriptional and essential gene analysis of Neisseria gonorrhoeae. Nucleic Acids Res., 42, 10579–10595.

83. Klein, B.A., Tenorio, E.L., Lazinski, D.W., Camilli, A., Duncan, M.J. and Hu, L.T. (2012) Identification of essential genes of the periodontal pathogen Porphyromonas gingivalis. BMC Genomics, 13, 578.

84. Remmele, C.W., Xian, Y., Abrecht, M., Faulstich, M., Fraunholz, J., Heinrichs, E., Dittrich, M.T., Muller, T., Reinhardt, R. and Rudel, T. (2014) Transcriptional landscape and essential genes of Fischer 344 rats. J. Bacteriol., 195, 5673–5684.

85. Ireland, P.M., Bullifent, H.L., Senior, N.J., Southern, S.J., Yang, Z.R., Gallagher, L.A., Ramage, E., Jacobs, M.A., Kaul, R., Brittnacher, M., et al. (2013) The serum resistome of a globally disseminated multidrug resistant uropathogenic Escherichia coli clone. Proc. Natl. Acad. Sci. U.S.A., 110, 10097–10102.

86. Turner, K.H., Wessel, A.K., Palmer, G.C., Murray, J.L. and Whiteley, M. (2015) Essential genome of Pseudomonas aeruginosa in cystic fibrosis sputum. Proc. Natl. Acad. Sci. U.S.A., 112, 4110–4115.

87. Poole, L.K., Cheong, E., Clatworthy, A.E., White, T., Osmulski, S.J., Li, L., Penaranda, C., Lander, E.S., Shores, N. and Hung, D.T. (2019) Essential genome of the core essential gene of Pseudomonas aeruginosa. Proc. Natl. Acad. Sci. U.S.A., 116, 10072–10080.

88. Seet, Q., Guerra-Assuncao, J.A., Lim, J., Sarkar-Taylor, M., Clark, T.G., Tan, P.B., Titball, R.W., Cucuji, et al. (2014) Genome-wide saturation mutagenesis of Burkholderia pseudomallei K92643 predicts essential genes and novel targets for antimicrobial development. mBio, 5, e00926-13.

89. Gislason, A.S., Turner, K., Cardona, S.T. and van Vliet, A.H. (2017) Essential genome of Campylobacter jejuni. BMC Genomics, 18, 616.

90. Daugherty, M.D., Somera, A.L., Kyrpides, N.C., Anderson, I., Dieterich, S.H., Dranow, D.M., Gardberg, A.S., Edwards, T.E., Armour, B., Begley, D.W., et al. (2011) The essential genome of a bacterium. Mol. Syst. Biol., 7, 528.

91. Gelfand, M.S., Collier, J.A., Fero, M.J., McAdams, H.H. and Shapiro, L. (2011) The knockout mutants: the Keio collection.

92. Arkin, A.P. (2011) Evidence-based annotation of gene function in Mycobacterium tuberculosis. Proc. Natl. Acad. Sci. U.S.A., 108, 535–540.

93. Brassil, R.P., O’Toole, G.A., Geoghegan, D.A., Bresnahan, M.J., McErlain, D., et al. (2013) Genomics and Hu,L.T. (2012) Identification of essential genes of the periodontal pathogen Porphyromonas gingivalis. BMC Genomics, 13, 578.
identification of essential genes in Staphylococcus aureus. Mol. Microbiol., 43, 1387–1400.

100. Ji,Y., Zhang,B., Van,S.F., Horn Warren,P., Woodnutt,G., Burnham,M.K. and Rosenberg,M. (2001) Identification of critical Staphylococcal genes using conditional phenotypes generated by antisense RNA. Science, 293, 2266–2269.

101. Chaudhuri,R.R., Allen,A.G., Owen,P.J., Shalom,G., Stone,K., Harrison,M., Burgis,T.A., Lockyer,M., Garcia-Lara,J., Foster,S.J. et al. (2009) Comprehensive identification of essential Staphylococcus aureus genes using Transposon-Mediated Differential Hybridisation (TMDH). BMC Genomics, 10, 291.

102. Coe,K.A., Lee,W., Stone,M.C., Komazin-Meredith,G., Meredith,T.C., Grad,Y.H. and Walker,S. (2019) Multi-strain Tn-Seq reveals common daptomycin resistance determinants in Staphylococcus aureus. PLoS Pathog., 15, e1007862.

103. Hooven,T.A., Catomeris,A.J., Akabas,L.H., Randis,T.M., Maskell,D.J., Peters,S.E., Ott,S., Santana-Cruz,J., Tallon,L.J., Tettelin,H. et al. (2016) The essential genome of Streptococcus agalactiae. BMC Genomics, 17, 406.

104. Shields,R.C., Zeng,L., Culp,D.J. and Burme,R.A. (2018) Genowide identification of essential genes and fitness determinants of Streptococcus mutans UA159. mSphere, 3, e00031-18.

105. Thanassi,J.A., Hartman-Neumann,S.L., Dougherty,T.J., Dougherty,B.A. and Pucci,M.J. (2002) Identification of 113 conserved essential genes using a high-throughput gene disruption system in Streptococcus pneumoniae. Nucleic Acids Res., 30, 3152–3162.

106. Song,J.H., Ko,K.S., Lee,J.Y., Baek,J.Y., Oh,W.S., Yoon,H.S., Jeong,J.Y. and Chun,J. (2005) Identification of essential genes in Streptococcus pneumoniae by allelic replacement mutagenesis. Mol Cells, 19, 365–374.

107. Le Breton,Y., Belew,A.T., Valdes,K.M., Islam,E., Curry,P., Tettelin,H., Shirliff,M.E., El-Sayed,N.M. and Mciver,K.S. (2015) Essential genes in the core genome of the human pathogen Streptococcus pyogenes. Sci. Rep., 5, 9838.

108. Xu,P., Ge,X., Chen,L., Wang,X., Dou,Y., Xu,J.Z., Patel,J.R., Stone,Y., Trinh,M., Evans,K. et al. (2011) Genome-wide essential gene identification in Streptococcus sanguinis. Sci. Rep., 1, 125.

109. Arenas,J., Zomer,A., Harders-Westerveen,J., Bootma,H.J., De Jonge,M.I., Stockhofe-Zurwieden,N., Smith,H.E. and De Greaff,A. (2020) Identification of conditionally essential genes for Streptococcus suis infection in pigs. Virulence, 11, 446–464.

110. Rubin,B.E., Wetmore,K.M., Price,M.N., Diamond,S., Shultzaberger,R.K., Love,L.C., Curtin,G., Arkin,A.P., Deutschbauer,A. and Golden,S.S. (2015) The essential gene set of a photosynthetic organism. Proc. Natl. Acad. Sci. U.S.A., 112, E6634–E6643.

111. Cameron,D.E., Urbach,J.M. and Mekalanos,J.J. (2008) A defined transposon mutant library and its use in identifying motility genes in Vibrio cholerae. Proc. Natl. Acad. Sci. U.S.A., 105, 8736–8741.

112. Chao,M.C., Pritchard,J.R., Zhang,Y.J., Rubin,E.J., Livny,J., Davis,B.M. and Waldor,M.K. (2013) High-resolution definition of the Vibrio cholerae essential gene set with hidden Markov model-based analyses of transposon-insertion sequencing data. Nucleic Acids Res., 41, 9033–9048.

113. Carda-Diegeuz,M., Silva-Hernandez,F.X., Hubbard,T.P., Chao,M.C., Waldor,M.K. and Amaro,C. (2018) Comprehensive identification of Vibrio vulnificus genes required for growth in human serum. Virulence, 9, 981–993.

114. Hu,W., Sillaots,S., Lemieux,S., Davison,J., Kauffman,S., Breton,A., Lintean,A., Xin,C., Bowman,J., Becker,J. et al. (2007) Essential gene identification and drug target prioritization in Aspergillus fumigatus. PLoS Pathog., 3, e24.

115. Chang,J., Wang,W., Yu,K., Zhang,T., Chen,X., Liu,Y., Shi,R., Wang,X., Xia,Q. and Ma,S. (2020) Genome-wide CRISPR screening reveals genes essential for cell viability and resistance to antiobiotic and biotic stresses in Bombyx mori. Genome Res., 30, 757–767.

116. Yu,S., Zheng,C., Zhou,F., Baillie,D.L., Rose,A.M., Deng,Z. and Chu,J.S. (2018) Genomic identification and functional analysis of essential genes in Caenorhabditis elegans. BMC Genomics, 19, 871.

117. Amsterdam,A., Nissen,R.M., Sun,Z., Swindle,E.C., Farrington,S. and Hopkins,N. (2004) Identification of 315 genes essential for early zebrafish development. Proc. Natl. Acad. Sci. U.S.A., 101, 12792–12797.

118. Spradling,A.C., Stern,D., Beaton,A., Rhein,E.J., Laverty,T., Mozden,N., Misra,S. and Rubin,G.M. (1999) The Berkeley Drosophila Genome Project gene disruption project: single P-element insertions mutating 25% of vital Drosophila genes. Genetics, 153, 135–177.

119. Liao,B.Y. and Zhang,J. (2008) Null mutations in human and mouse orthologs frequently result in different phenotypes. Proc. Natl. Acad. Sci. U.S.A., 105, 6987–6992.

120. Hart,T., Tong,A.H.Y., Chan,K., Van Leeuwen,J., Seetharaman,A., Aregger,M., Chandrashekhar,M., Hustedi,N., Seth,S., Noonan,A. et al. (2017) Evaluation and design of genome-wide CRISPR/SpCas9 knockout screens. G3 (Bethesda), 7, 2719–2727.

121. Zhu,J., Gong,R., Zhu,Q., He,Q., Xu,N., Xu,Y., Cai,M., Zhou,X., Zhang,Y. and Zhou,M. (2018) Genome-wide determination of gene essentiality by transposon insertion sequencing in yeast Pichia pastoris. Sci. Rep., 8, 10223.

122. Liao,B.Y. and Zhang,J. (2007) Mouse duplicate genes are as essential as singletons. Trends Genet., 23, 378–381.

123. Giaever,G., Chu,A.M., Ni,L., Connelly,C., Riles,L., Veronneau,S., Dow,S., Lucau-Danila,A., Anderson,K., Andre,B. et al. (2002) Functional profiling of the Saccharomyces cerevisiae genome. Nature, 418, 387–391.

124. Kim,D.U., Hoyes,J., Kim,D., Wood,V., Park,H.O., Won,M., Yoo,H.S., Duhig,T., Nam,M., Palmer,G. et al. (2010) Analysis of a genome-wide set of gene deletions in the fission yeast Schizosaccharomyces pombe. Nat. Biotechnol., 28, 617–623.

125. Amberger,J.S., Bocchini,C.A., Scott,A.F. and Hamosh,A. (2019) OMIM.org: leveraging knowledge across phenotype-gene relationships. Nucleic Acids Res., 47, D1038–D1043.

126. Smith,C.L., Blake,J.A., Kadin,J.A., Richardson,J.E., Bult,C.J. and Mouse Genome Database, G. (2018) Mouse Genome Database (MGD)-2018: knowledgebase for the laboratory mouse. Nucleic Acids Res., 46, D836–D842.