

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

Hao Luo¹, Yan Lin¹, Tao Liu¹, Fei-Liao Lai¹, Chun-Ting Zhang¹, Feng Gao^{1,2,*} and Ren Zhang^{3,*}

¹Department of Physics, School of Science, Tianjin University, Tianjin 300072, China, ²Frontiers Science Center for Synthetic Biology and Key Laboratory of Systems Bioengineering (Ministry of Education), Tianjin University, Tianjin 300072, China and ³Center 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 re-

search 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-

*To whom correspondence should be addressed. Tel: +1 313 577 0027; Fax: +1 313 577 5218; Email: rzhang@med.wayne.edu
Correspondence may also be addressed to Feng Gao. Tel: +86 22 2740 2987; Fax: +86 22 2740 2697; Email: fgao@tju.edu.cn

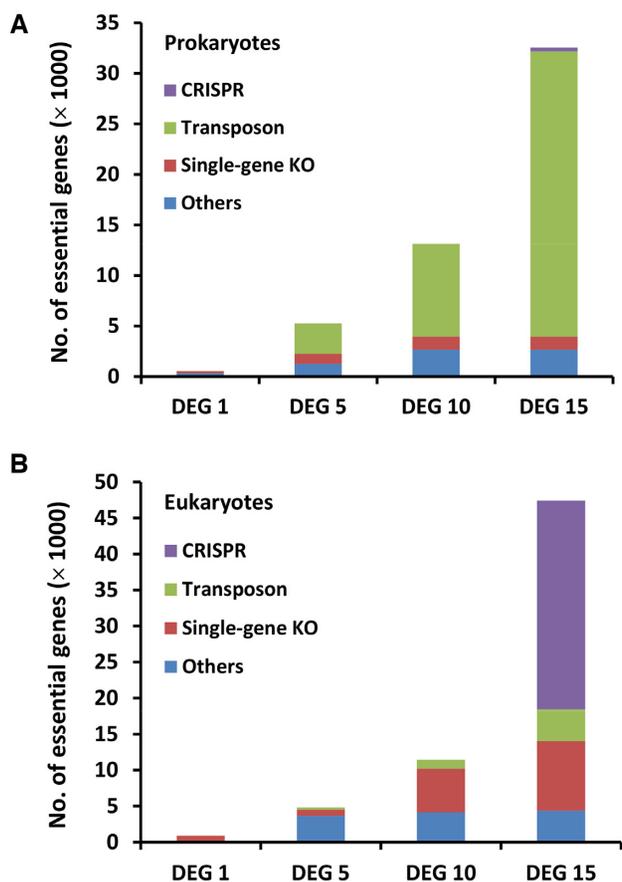


Figure 1. The development of DEG. The number of essential-gene records for (A) prokaryotes and (B) eukaryotes in DEG with different versions. The stacked bars show the number of records according to the experimental methods.

tology and KEGG pathway enrichment analysis, and generation of Venn diagrams to compare and contrast gene sets between experiments.

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 DEG 15 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. In com-

Table 1. Contents of DEG 15

Domain of life	Organism	No. of essential genomic elements		Method	Saturated	Reference	Note ^a
		Coding	Noncoding				
Bacteria	<i>Acinetobacter baumannii</i> ATCC 17978	453	59	INSeq	Yes	(24)	
	<i>Acinetobacter baumannii</i> ATCC 17978	157	1	INSeq	Yes	(24)	In the mouse lung
	<i>Acinetobacter baylyi</i>	499		Single-gene knockout	Yes	(57)	Minimal medium
	<i>Aggregatibacter actinomycetemcomitans</i>	59		Tn-seq ^b	Yes	(58)	For coinfection with sympatric and allopatric microbes
	<i>Agrobacterium fabrum</i> str. C58	361	11	Tn-seq	Yes	(25)	
	<i>Bacillus subtilis</i>	261	2	Single-gene knockout	Yes	(59)	
	<i>Bacillus thuringiensis</i> BMB171	516		Tn-seq	Yes	(60)	
	<i>Bacteroides fragilis</i>	550		Tn-seq	Yes	(61)	
	<i>Bacteroides thetaiotaomicron</i>	325		INSeq	Yes	(21)	
	<i>Bifidobacterium breve</i>	453		TraDIS	Yes	(62)	
	<i>Brevundimonas subvibrioides</i>	448		Tn-seq	Yes	(25)	
	<i>Brevundimonas subvibrioides</i> ATCC 15264	412	35	Tn-seq	Yes	(25)	
	<i>Burkholderia cenocepacia</i> J2315	383		TraDIS	Yes	(63)	
	<i>Burkholderia cenocepacia</i> K56-2	508		Tn-seq	Yes	(64)	
	<i>Burkholderia pseudomallei</i> K96243	505		TraDIS	Yes	(65)	
	<i>Burkholderia thailandensis</i>	406		Tn-seq	Yes	(66)	
	<i>Campylobacter jejuni</i>	233		Tn-seq	Yes	(67)	
	<i>Campylobacter jejuni</i> subsp. <i>jejuni</i> 81-176	384		Tn-seq	Yes	(68)	
	<i>Campylobacter jejuni</i> subsp. <i>jejuni</i> NCTC 11168	166		Tn-seq	Yes	(68)	
	<i>Caulobacter crescentus</i>	480	532	Tn-seq	Yes	(69)	
	<i>Escherichia coli</i>	620		Genetic footprinting	Yes	(70)	
	<i>Escherichia coli</i>	303		Single-gene knockout	Yes	(71)	
	<i>Escherichia coli</i>	379		CRISPR	Yes	(72)	
	<i>Escherichia coli</i> O157:H7	1265	37	Tn-seq	Yes	(26)	
	<i>Escherichia coli</i> ST131 strain EC958	315		TraDIS	Yes	(73)	
	<i>Francisella novicida</i>	396		Tn-seq	Yes	(74)	
	<i>Francisella tularensis</i> Schu S4	453		TraDIS	Yes	(75)	
	<i>Haemophilus influenzae</i>	667		Genetic footprinting	Yes	(76)	
	<i>Helicobacter pylori</i>	344		MATT	Yes	(77)	
	<i>Mycobacterium avium</i> subsp. <i>hominissuis</i> strain MAC109	230		Tn-seq	Yes	(78)	
	<i>Mycobacterium tuberculosis</i>	614		TraSH	Yes	(79)	
	<i>Mycobacterium tuberculosis</i>	774		Tn-seq	Yes	(80)	
<i>Mycobacterium tuberculosis</i>	742	35	Tn-seq	Yes	(27)		
<i>Mycobacterium tuberculosis</i>	461		Tn-seq	Yes	(81)		
<i>Mycobacterium tuberculosis</i>	601		Tn-seq	Yes	(82)		
<i>Mycoplasma genitalium</i>	382		Tn-seq	Yes	(19,83)		
<i>Mycoplasma pneumoniae</i>	342	34	Tn-seq	Yes	(28)		
<i>Mycoplasma pulmonis</i>	321		Tn-seq	Yes	(84)		
<i>Neisseria gonorrhoeae</i> MS11	751		Tn-seq	Yes	(85)		
<i>Porphyromonas gingivalis</i>	463		Tn-seq	Yes	(86)		
<i>Porphyromonas gingivalis</i> ATCC 33277	281		Tn-seq	Yes	(87)		
<i>Providencia stuartii</i> strain BE2467	496	25	Tn-seq	Yes	(88)		
<i>Pseudomonas aeruginosa</i>	335		TraSH	Yes	(89)		
<i>Pseudomonas aeruginosa</i>	117		Tn-seq	Yes	(23)		
<i>Pseudomonas aeruginosa</i>	321		Tn-seq	Yes	(90)		
<i>Pseudomonas aeruginosa</i> PAO1	336		Tn-seq	Yes	(91)		
<i>Pseudomonas aeruginosa</i> PAO1	551		Tn-seq	Yes	(92)		

Table 1. Continued

Domain of life	Organism	No. of essential genomic elements		Method	Saturated	Reference	Note ^a
		Coding	Noncoding				
	<i>Ralstonia solanacearum</i> GMI1000	465		Tn-seq	Yes	(93)	
	<i>Rhodobacter sphaeroides</i>	493		Tn-seq	Yes	(94)	
	<i>Rhodospseudomonas palustris</i> CGA009	522		Tn-seq	Yes	(95)	
	<i>Salmonella enterica</i> Typhimurium	306	15	TraDIS	Yes	(29)	
	<i>Salmonella entericaseroovar</i> Typhi	356		TraDIS	Yes	(20)	
	<i>Salmonella entericaseroovar</i> Typhi Ty2	358	24	TraDIS	Yes	(29)	
	<i>Salmonella entericaseroovar</i> Typhimurium	105		Tn-seq	Yes	(96)	
	<i>Salmonella entericaseroovar</i> Typhimurium SL1344	353	23	TraDIS	Yes	(29)	
	<i>Salmonella typhimurium</i>	490		Insertion-duplication	Yes	(97)	
	<i>Shewanella oneidensis</i>	403		Transposon mutagenesis	Yes	(98)	
	<i>Sphingomonas wittichii</i>	579	32	Tn-seq	Yes	(30)	
	<i>Staphylococcus aureus</i>	302		Antisense RNA	No	(99,100)	
	<i>Staphylococcus aureus</i>	351		TMDH	Yes	(101)	
	<i>Staphylococcus aureus</i> subsp. <i>aureus</i> MRSA252	295		Tn-seq	Yes	(102)	
	<i>Staphylococcus aureus</i> subsp. <i>aureus</i> MSSA476	305		Tn-seq	Yes	(102)	
	<i>Staphylococcus aureus</i> subsp. <i>aureus</i> MW2	256		Tn-seq	Yes	(102)	
	<i>Staphylococcus aureus</i> subsp. <i>aureus</i> NCTC 8325	288		Tn-seq	Yes	(102)	
	<i>Staphylococcus aureus</i> subsp. <i>aureus</i> USA300 TCH1516	295		Tn-seq	Yes	(102)	
	<i>Streptococcus agalactiae</i> A909	317		Tn-seq	Yes	(103)	
	<i>Streptococcus mutans</i> UA159	197	6	Tn-seq	Yes	(104)	
	<i>Streptococcus pneumoniae</i>	113		Insertion-duplication	No	(105)	
	<i>Streptococcus pneumoniae</i>	133		allelic replacement mutagenesis	No	(106)	
	<i>Streptococcus pneumoniae</i>		72	Tn-seq	Yes	(31)	
	<i>Streptococcus pyogenes</i> MGAS5448	227		Tn-seq	Yes	(107)	
	<i>Streptococcus pyogenes</i> NZ131	241		Tn-seq	Yes	(107)	
	<i>Streptococcus sanguinis</i>	218		Single-gene knockout	Yes	(108)	
	<i>Streptococcus suis</i>	361		Tn-seq	Yes	(109)	
	<i>Synechococcus elongatus</i> PCC 7942	682	34	Tn-seq	Yes	(110)	
	<i>Vibrio cholerae</i>	789		Tn-seq	Yes	(111)	
	<i>Vibrio cholerae</i> C6706	343		Tn-seq	Yes	(112)	
	<i>Vibrio vulnificus</i>	316		Tn-seq	Yes	(113)	
Archaea	<i>Methanococcus maripaludis</i>	519		Tn-seq	Yes	(32)	
	<i>Sulfolobus islandicus</i> M.16.4	441		Tn-seq	Yes	(33)	
Eukaryotes	<i>Arabidopsis thaliana</i>	358		Single-gene knockout	No	(54)	
	<i>Aspergillus fumigatus</i>	35		Conditional promoter replacement	No	(114)	
	<i>Bombyx mori</i>	1006		CRISPR	Yes	(115)	
	<i>Caenorhabditis elegans</i>	44		Genetic mapping	No	(116)	
	<i>Caenorhabditis elegans</i>	294		RNA interference	No	(56)	
	<i>Danio rerio</i>	315		Insertional mutagenesis	No	(117)	
	<i>Drosophila melanogaster</i>	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				
	<i>Homo sapiens</i>	2452		OMIM annotation ^c	No	(119)	
	<i>Homo sapiens</i>	1562		CRISPR	Yes	(14)	Stem cells
	<i>Homo sapiens</i>	1593		CRISPR	Yes	(14)	HAP1 cells
	<i>Homo sapiens</i>	1690		CRISPR	Yes	(120)	Core essential genes among 17 cell lines
	<i>Homo sapiens</i>	3230		Exome sequencing	Yes	(18)	
	<i>Homo sapiens</i>	2054		CRISPR	Yes	(12)	KBM7 cells
	<i>Homo sapiens</i>	2181		CRISPR	Yes	(12)	HAP1 cells
	<i>Homo sapiens</i>	1878		CRISPR	Yes	(11)	KBM7 cells
	<i>Homo sapiens</i>	1660		CRISPR	Yes	(11)	K562 cells
	<i>Homo sapiens</i>	1630		CRISPR	Yes	(11)	Jiyoye cells
	<i>Homo sapiens</i>	1461		CRISPR	Yes	(11)	Raji cells
	<i>Homo sapiens</i>	1196		CRISPR	Yes	(13)	A375 cells
	<i>Homo sapiens</i>	1892		CRISPR	Yes	(13)	DLD1 cells
	<i>Homo sapiens</i>	2196		CRISPR	Yes	(13)	GBM cells
	<i>Homo sapiens</i>	2073		CRISPR	Yes	(13)	HCT116 cells
	<i>Homo sapiens</i>	386		shRNA	Yes	(13)	HCT116 cells
	<i>Homo sapiens</i>	1696		CRISPR	Yes	(13)	HeLa cells
	<i>Homo sapiens</i>	2038		CRISPR	Yes	(13)	RPE1 cells
	<i>Homo sapiens</i>	92		Functional genomics	No	(15)	Podocytes
	<i>Homo sapiens</i>	79		CRISPR	Yes	(16)	Hepatocellular carcinoma
	<i>Homo sapiens</i>	191		CRISPR	Yes	(17)	K562 cells
	<i>Komagataella phaffii</i> GS115	753		Tn-seq	Yes	(121)	
	<i>Mus musculus</i>	435		Single-gene knockout	No	(53)	Embryonic lethality
	<i>Mus musculus</i>	1933		Single-gene knockout	No	(52)	Prewaning lethality
	<i>Mus musculus</i>	2136		MGI annotation ^d	No	(122)	
	<i>Plasmodium falciparum</i>	2680		transposon mutagenesis	Yes	(34)	
	<i>Saccharomyces cerevisiae</i>	1110		Single-gene knockout	Yes	(123)	Six conditions including minimal medium
	<i>Schizosaccharomyces pombe</i>	1260		Single-gene knockout	Yes	(124)	Rich medium

^aBacteria 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 (HITS), 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 insertion 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), HITS (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 baumannii* (24), *Brevundimonas subvibrioides* (25), *Escherichia coli* O157:H7 (26), *Mycobacterium tuberculosis* (27), *Mycoplasma pneumonia* (28), *Salmonella entericaserovar* 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-

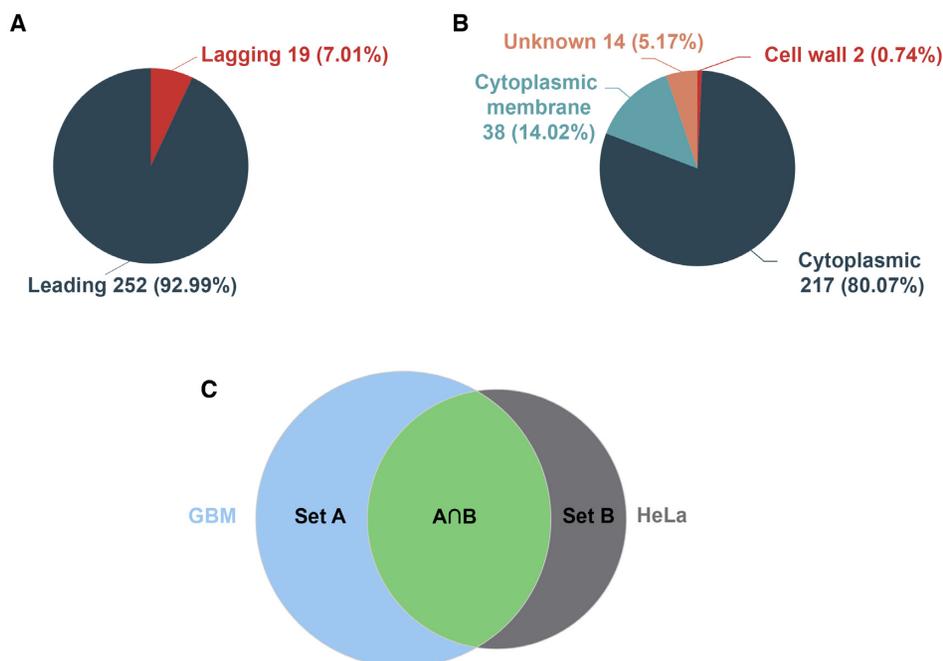


Figure 2. Screenshots of some analysis modules in DEG 15. (A) Distribution of essential genes between leading and lagging strands and (B) distribution of sub-cellular localizations of essential genes in the *Bacillus subtilis* genome. (C) A Venn diagram showing the intersection and the union between two datasets (GBM and HeLa cells). The diagrams are clickable to show a list of genes with detailed information.

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 GOATOOLS (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

- Koonin, E.V. (2000) How many genes can make a cell: the minimal-gene-set concept. *Annu. Rev. Genomics Hum. Genet.*, **1**, 99–116.
- Bartha, I., di Iulio, J., Venter, J.C. and Telenti, A. (2018) Human gene essentiality. *Nat. Rev. Genet.*, **19**, 51–62.
- Zhang, R., Ou, H.Y. and Zhang, C.T. (2004) DEG: a database of essential genes. *Nucleic Acids Res.*, **32**, D271–D272.
- Zhang, R. and Lin, Y. (2009) DEG 5.0, a database of essential genes in both prokaryotes and eukaryotes. *Nucleic Acids Res.*, **37**, D455–D458.
- 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.
- 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.
- 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.
- Barquist, L., Boinett, C.J. and Cain, A.K. (2013) Approaches to querying bacterial genomes with transposon-insertion sequencing. *RNA Biol.*, **10**, 1161–1169.
- 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.
- Hsu, P.D., Lander, E.S. and Zhang, F. (2014) Development and applications of CRISPR-Cas9 for genome engineering. *Cell*, **157**, 1262–1278.
- 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.
- Blomen, V.A., Majek, P., Jae, L.T., Bigenzahn, J.W., Nieuwenhuis, J., Staring, J., Sacco, R., van Diemen, F.R., Olk, N., Stukalov, A. *et al.* (2015) Gene essentiality and synthetic lethality in haploid human cells. *Science*, **350**, 1092–1096.
- Hart, T., Chandrashekar, M., Aregger, M., Steinhart, Z., Brown, K.R., MacLeod, G., Mis, M., Zimmermann, M., Fradet-Turcotte, A., Sun, S. *et al.* (2015) High-resolution CRISPR screens reveal fitness genes and genotype-specific cancer liabilities. *Cell*, **163**, 1515–1526.
- Mair, B., 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.
- 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.
- Wang, Y., Gao, B., Tan, P.Y., Handoko, Y.A., Sekar, K., Deivasigamani, A., Seshachalam, V.P., OuYang, H.Y., Shi, M., Xie, C. *et al.* (2019) Genome-wide CRISPR knockout screens identify NCAPG as an essential oncogene for hepatocellular carcinoma tumor growth. *FASEB J.*, **33**, 8759–8770.
- Arroyo, J.D., Jourdain, A.A., Calvo, S.E., Ballarano, C.A., Doench, J.G., Root, D.E. and Mootha, V.K. (2016) A Genome-wide CRISPR death screen identifies genes essential for oxidative phosphorylation. *Cell Metab.*, **24**, 875–885.
- Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B. *et al.* (2016) Analysis of protein-coding genetic variation in 60,706 humans. *Nature*, **536**, 285–291.
- Hutchison, C.A., Peterson, S.N., Gill, S.R., Cline, R.T., White, O., Fraser, C.M., Smith, H.O. and Venter, J.C. (1999) Global transposon mutagenesis and a minimal Mycoplasma genome. *Science*, **286**, 2165–2169.
- 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 every Salmonella Typhi gene using one million transposon mutants. *Genome Res.*, **19**, 2308–2316.
- Goodman, A.L., McNulty, N.P., Zhao, Y., Leip, D., Mitra, R.D., Lozupone, 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.
- Gawronski, J.D., Wong, S.M., Giannoukos, G., Ward, D.V. and Akerley, 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.
- Gallagher, L.A., Shendure, J. and Manoil, C. (2011) Genome-scale identification of resistance functions in *Pseudomonas aeruginosa* using Tn-seq. *mBio*, **2**, e00315-10.
- 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**, e01163-14.
- 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.

26. Warr, A.R., Hubbard, T.P., Munera, D., Blondel, C.J., Abel Zur Wiesch, P., Abel, S., Wang, X., Davis, B.M. and Waldor, M.K. (2019) Transposon-insertion sequencing screens unveil requirements for EHEC growth and intestinal colonization. *PLoS Pathog.*, **15**, e1007652.
27. Zhang, Y.J., Ioerger, T.R., Huttenhower, C., Long, J.E., Sasseti, C.M., Sacchettini, J.C. and Rubin, E.J. (2012) Global assessment of genomic regions required for growth in *Mycobacterium tuberculosis*. *PLoS Pathog.*, **8**, e1002946.
28. Lluch-Senar, M., Delgado, J., Chen, W.H., Llorens-Rico, V., O'Reilly, F.J., Wodke, J.A., Unal, E.B., Yus, E., Martinez, S., Nichols, R.J. *et al.* (2015) Defining a minimal cell: essentiality of small ORFs and ncRNAs in a genome-reduced bacterium. *Mol. Syst. Biol.*, **11**, 780.
29. Barquist, L., Langridge, G.C., Turner, D.J., Phan, M.D., Turner, A.K., Bateman, A., Parkhill, J., Wain, J. and Gardner, P.P. (2013) A comparison of dense transposon insertion libraries in the *Salmonella* serovars Typhi and Typhimurium. *Nucleic Acids Res.*, **41**, 4549–4564.
30. Roggo, C., Coronado, E., Moreno-Forero, S.K., Harshman, K., Weber, J. and van der Meer, J.R. (2013) Genome-wide transposon insertion scanning of environmental survival functions in the polycyclic aromatic hydrocarbon degrading bacterium *Sphingomonas wittichii* RW1. *Environ Microbiol.*, **15**, 2681–2695.
31. Mann, B., van Opijnen, T., Wang, J., Obert, C., Wang, Y.D., Carter, R., McGoldrick, D.J., Ridout, G., Camilli, A., Tuomanen, E.I. *et al.* (2012) Control of virulence by small RNAs in *Streptococcus pneumoniae*. *PLoS Pathog.*, **8**, e1002788.
32. Sarmiento, F., Mrazek, J. and Whitman, W.B. (2013) Genome-scale analysis of gene function in the hydrogenotrophic methanogenic archaeon *Methanococcus maripaludis*. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 4726–4731.
33. Zhang, C., Phillips, A.P.R., Wipfler, R.L., Olsen, G.J. and Whitaker, R.J. (2018) The essential genome of the crenarchaeal model *Sulfolobus islandicus*. *Nat. Commun.*, **9**, 4908.
34. Zhang, M., Wang, C., Otto, T.D., Oberstaller, J., Liao, X., Adapa, S.R., Udenze, K., Bronner, I.F., Casandra, D., Mayho, M. *et al.* (2018) Uncovering the essential genes of the human malaria parasite *Plasmodium falciparum* by saturation mutagenesis. *Science*, **360**, eaap7847.
35. Lin, Y., Gao, F. and Zhang, C.T. (2010) Functionality of essential genes drives gene strand-bias in bacterial genomes. *Biochem. Biophys. Res. Commun.*, **396**, 472–476.
36. Lang, K.S., Hall, A.N., Merrikh, C.N., Ragheb, M., Tabakh, H., Pollock, A.J., Woodward, J.J., Dreifus, J.E. and Merrikh, H. (2017) Replication-transcription conflicts generate R-loops that orchestrate bacterial stress survival and pathogenesis. *Cell*, **170**, 787–799.
37. Luo, H. and Gao, F. (2019) DoriC 10.0: an updated database of replication origins in prokaryotic genomes including chromosomes and plasmids. *Nucleic Acids Res.*, **47**, D74–D77.
38. Gao, F., Luo, H. and Zhang, C.T. (2013) DoriC 5.0: an updated database of oriC regions in both bacterial and archaeal genomes. *Nucleic Acids Res.*, **41**, D90–D93.
39. Yu, N.Y., Wagner, J.R., Laird, M.R., Melli, G., Rey, S., Lo, R., Dao, P., Sahinalp, S.C., Ester, M., Foster, L.J. *et al.* (2010) PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics*, **26**, 1608–1615.
40. Mao, X., Ma, Q., Zhou, C., Chen, X., Zhang, H., Yang, J., Mao, F., Lai, W. and Xu, Y. (2014) DOOR 2.0: presenting operons and their functions through dynamic and integrated views. *Nucleic Acids Res.*, **42**, D654–D659.
41. Jeske, L., Placzek, S., Schomburg, I., Chang, A. and Schomburg, D. (2019) BRENDA in 2019: a European ELIXIR core data resource. *Nucleic Acids Res.*, **47**, D542–D549.
42. Kanehisa, M., Sato, Y., Furumichi, M., Morishima, K. and Tanabe, M. (2019) New approach for understanding genome variations in KEGG. *Nucleic Acids Res.*, **47**, D590–D595.
43. The Gene Ontology, C. (2019) The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.*, **47**, D330–D338.
44. Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernandez-Plaza, A., Forslund, S.K., Cook, H., Mende, D.R., Letunic, I., Rattei, T., Jensen, L.J. *et al.* (2019) eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.*, **47**, D309–D314.
45. Klopfenstein, D.V., Zhang, L., Pedersen, B.S., Ramirez, F., Warwick Vesztrocy, A., Naldi, A., Mungall, C.J., Yunes, J.M., Botvinnik, O., Weigel, M. *et al.* (2018) GOATOOLS: a Python library for Gene Ontology analyses. *Sci. Rep.*, **8**, 10872.
46. Yu, G., Wang, L.G., Han, Y. and He, Q.Y. (2012) clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*, **16**, 284–287.
47. Li, D., Mei, H., Shen, Y., Su, S., Zhang, W., Wang, J., Zu, M. and Chen, W. (2018) ECharts: a declarative framework for rapid construction of web-based visualization. *Visual Informatics*, **2**, 136–146.
48. Peters, J.M., Colavin, A., Shi, H., Czarny, T.L., Larson, M.H., Wong, S., Hawkins, J.S., Lu, C.H.S., Koo, B.M., Marta, E. *et al.* (2016) A comprehensive, CRISPR-based functional analysis of essential genes in bacteria. *Cell*, **165**, 1493–1506.
49. Juhas, M., Eberl, L. and Glass, J.I. (2011) Essence of life: essential genes of minimal genomes. *Trends Cell Biol.*, **21**, 562–568.
50. Pertesi, M., Ekdahl, L., Palm, A., Johnsson, E., Jarvstrat, L., Wihlborg, A.K. and Nilsson, B. (2019) Essential genes shape cancer genomes through linear limitation of homozygous deletions. *Commun. Biol.*, **2**, 262.
51. Magnusdottir, S., Heinken, A., Kutt, L., Ravcheev, D.A., Bauer, E., Noronha, A., Greenhalgh, K., Jager, C., Baginska, J., Wilmes, P. *et al.* (2017) Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. *Nat. Biotechnol.*, **35**, 81–89.
52. Cacheiro, P., Munoz-Fuentes, V., Murray, S.A., Dickinson, M.E., Bucan, M., Nutter, L.M.J., Peterson, K.A., Haselimashhadi, H., Flenniken, A.M., Morgan, H. *et al.* (2020) Human and mouse essentiality screens as a resource for disease gene discovery. *Nat. Commun.*, **11**, 655.
53. Dickinson, M.E., Flenniken, A.M., Ji, X., Teboul, L., Wong, M.D., White, J.K., Meehan, T.F., Weninger, W.J., Westerberg, H., Adissu, H. *et al.* (2016) High-throughput discovery of novel developmental phenotypes. *Nature*, **537**, 508–514.
54. Meinke, D., Muralla, R., Sweeney, C. and Dickerman, A. (2008) Identifying essential genes in *Arabidopsis thaliana*. *Trends Plant. Sci.*, **13**, 483–491.
55. Boldrin, F., Degiacomi, G., Serafini, A., Kolly, G.S., Ventura, M., Sala, C., Proveddi, R., Palu, G., Cole, S.T. and Manganelli, R. (2018) Promoter mutagenesis for fine-tuning expression of essential genes in *Mycobacterium tuberculosis*. *Microb. Biotechnol.*, **11**, 238–247.
56. Kamath, R.S., Fraser, A.G., Dong, Y., Poulin, G., Durbin, R., Gotta, M., Kanapin, A., Le Bot, N., Moreno, S., Sohrmann, M. *et al.* (2003) Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature*, **421**, 231–237.
57. de Berardinis, V., Vallenet, D., Castelli, V., Besnard, M., Pinet, A., Cruaud, C., Samair, S., Lechaplais, C., Gyapay, G., Richez, C. *et al.* (2008) A complete collection of single-gene deletion mutants of *Acinetobacter baylyi* ADP1. *Mol. Syst. Biol.*, **4**, 174.
58. Lewin, G.R., Stacy, A., Michie, K.L., Lamont, R.J. and Whiteley, M. (2019) Large-scale identification of pathogen essential genes during coinfection with sympatric and allopatric microbes. *Proc. Natl. Acad. Sci. U.S.A.*, **116**, 19685–19694.
59. Kobayashi, K., Ehrlich, S.D., Albertini, A., Amati, G., Andersen, K.K., Arnaud, M., Asai, K., Ashikaga, S., Aymerich, S., Bessieres, P. *et al.* (2003) Essential *Bacillus subtilis* genes. *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 4678–4683.
60. Bishop, A.H., Rachwal, P.A. and Vaid, A. (2014) Identification of genes required by *Bacillus thuringiensis* for survival in soil by transposon-directed insertion site sequencing. *Curr. Microbiol.*, **68**, 477–485.
61. Veeranagouda, Y., Husain, F., Tenorio, E.L. and Wexler, H.M. (2014) Identification of genes required for the survival of *B. fragilis* using massive parallel sequencing of a saturated transposon mutant library. *BMC Genomics*, **15**, 429.
62. Ruiz, L., Bottacini, F., Boinett, C.J., Cain, A.K., O'Connell-Motherway, M., Lawley, T.D. and van Sinderen, D. (2017) The essential genomic landscape of the commensal *Bifidobacterium breve* UCC2003. *Sci. Rep.*, **7**, 5648.
63. Wong, Y.C., Abd El Ghany, M., Naem, R., Lee, K.W., Tan, Y.C., Pain, A. and Nathan, S. (2016) Candidate essential genes in

- Burkholderia cenocepacia J2315 identified by genome-wide TraDIS. *Front Microbiol.*, **7**, 1288.
64. Gislason, A.S., Turner, K., Domaratzi, M. and Cardona, S.T. (2017) Comparative analysis of the Burkholderia cenocepacia K56-2 essential genome reveals cell envelope functions that are uniquely required for survival in species of the genus Burkholderia. *Microb Genom.*, **3**, e000140.
 65. Moule, M.G., Hemsley, C.M., Seet, Q., Guerra-Assuncao, J.A., Lim, J., Sarkar-Tyson, M., Clark, T.G., Tan, P.B., Titball, R.W., Cuccui, J. *et al.* (2014) Genome-wide saturation mutagenesis of Burkholderia pseudomallei K96243 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 struome. *PLoS One*, **8**, e53851.
 67. Metris, A., Reuter, M., Gaskin, D.J., Baranyi, J. and van Vliet, A.H. (2011) In vivo and in silico determination of essential genes of Campylobacter jejuni. *BMC Genomics*, **12**, 535.
 68. Mandal, R.K., Jiang, T. and Kwon, Y.M. (2017) Essential genome of Campylobacter jejuni. *BMC Genomics*, **18**, 616.
 69. Christen, B., Abeliuk, E., Collier, J.M., Kalogeraki, V.S., Passarelli, B., Coller, J.A., Fero, M.J., McAdams, H.H. and Shapiro, L. (2011) The essential genome of a bacterium. *Mol. Syst. Biol.*, **7**, 528.
 70. Gerdes, S.Y., Scholle, M.D., Campbell, J.W., Balazsi, G., Ravasz, E., Daugherty, M.D., Somera, A.L., Kyrpides, N.C., Anderson, I., Gelfand, M.S. *et al.* (2003) Experimental determination and system level analysis of essential genes in Escherichia coli MG1655. *J. Bacteriol.*, **185**, 5673–5684.
 71. Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L. and Mori, H. (2006) Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.*, **2**, doi:10.1038/msb4100050.
 72. Rousset, F., Cui, L., Siouve, E., Becavin, C., Depardieu, F. and Bikard, D. (2018) Genome-wide CRISPR-dCas9 screens in E. coli identify essential genes and phage host factors. *PLoS Genet.*, **14**, e1007749.
 73. Phan, M.D., Peters, K.M., Sarkar, S., Lukowski, S.W., Allsopp, L.P., Gomes Moriel, D., Achard, M.E., Totsika, M., Marshall, V.M., Upton, M. *et al.* (2013) The serum resistome of a globally disseminated multidrug resistant uropathogenic Escherichia coli clone. *PLoS Genet.*, **9**, e1003834.
 74. Gallagher, L.A., Ramage, E., Jacobs, M.A., Kaul, R., Brittnacher, M. and Manoil, C. (2007) A comprehensive transposon mutant library of Francisella novicida, a bioweapon surrogate. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 1009–1014.
 75. Ireland, P.M., Bullifent, H.L., Senior, N.J., Southern, S.J., Yang, Z.R., Ireland, R.E., Nelson, M., Atkins, H.S., Titball, R.W. and Scott, A.E. (2019) Global analysis of genes essential for Francisella tularensis Schu S4 growth in vitro and for fitness during competitive infection of Fischer 344 rats. *J. Bacteriol.*, **201**, e00630-18.
 76. Akerley, B.J., Rubin, E.J., Novick, V.L., Amaya, K., Judson, N. and Mekalanos, J.J. (2002) A genome-scale analysis for identification of genes required for growth or survival of Haemophilus influenzae. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 966–971.
 77. Salama, N.R., Shepherd, B. and Falkow, S. (2004) Global transposon mutagenesis and essential gene analysis of Helicobacter pylori. *J. Bacteriol.*, **186**, 7926–7935.
 78. Matern, W.M., Jenquin, R.L., Bader, J.S. and Karakousis, P.C. (2020) Identifying the essential genes of Mycobacterium avium subsp. hominissuis with Tn-Seq using a rank-based filter procedure. *Sci. Rep.*, **10**, 1095.
 79. Sassetti, C.M., Boyd, D.H. and Rubin, E.J. (2003) Genes required for mycobacterial growth defined by high density mutagenesis. *Mol. Microbiol.*, **48**, 77–84.
 80. Griffin, J.E., Gawronski, J.D., Dejesus, M.A., Ioerger, T.R., Akerley, B.J. and Sassetti, C.M. (2011) High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. *PLoS Pathog.*, **7**, e1002251.
 81. DeJesus, M.A., Gerrick, E.R., Xu, W., Park, S.W., Long, J.E., Boutte, C.C., Rubin, E.J., Schnappinger, D., Ehrst, S., Fortune, S.M. *et al.* (2017) Comprehensive essentiality analysis of the Mycobacterium tuberculosis genome via saturating transposon mutagenesis. *mBio*, **8**, e02133-16.
 82. Minato, Y., Gohl, D.M., Thiede, J.M., Chacon, J.M., Harcombe, W.R., Maruyama, F. and Baughn, A.D. (2019) Genomewide assessment of Mycobacterium tuberculosis conditionally essential metabolic pathways. *mSystems*, **4**, e00070-19.
 83. Glass, J.L., Assad-Garcia, N., Alperovich, N., Yooseph, S., Lewis, M.R., Maruf, M., Hutchison, C.A. 3rd, Smith, H.O. and Venter, J.C. (2006) Essential genes of a minimal bacterium. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 425–430.
 84. French, C.T., Lao, P., Loraine, A.E., Matthews, B.T., Yu, H. and Dybvig, K. (2008) Large-scale transposon mutagenesis of Mycoplasma pulmonis. *Mol. Microbiol.*, **69**, 67–76.
 85. Remmele, C.W., Xian, Y., Albrecht, M., Faulstich, M., Fraunholz, M., Heinrichs, E., Dittrich, M.T., Muller, T., Reinhardt, R. and Rudel, T. (2014) Transcriptional landscape and essential genes of Neisseria gonorrhoeae. *Nucleic Acids Res.*, **42**, 10579–10595.
 86. 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.
 87. Hutcherson, J.A., Gogeneni, H., Yoder-Himes, D., Hendrickson, E.L., Hackett, M., Whiteley, M., Lamont, R.J. and Scott, D.A. (2016) Comparison of inherently essential genes of Porphyromonas gingivalis identified in two transposon-sequencing libraries. *Mol. Oral Microbiol.*, **31**, 354–364.
 88. Johnson, A.O., Forsyth, V., Smith, S.N., Learman, B.S., Brauer, A.L., White, A.N., Zhao, L., Wu, W., Mobley, H.L.T. and Armbruster, C.E. (2020) Transposon insertion site sequencing of Providencia stuartii: essential genes, fitness factors for catheter-associated urinary tract infection, and the impact of polymicrobial infection on fitness requirements. *mSphere*, **5**, e00412-20.
 89. Liberati, N.T., Urbach, J.M., Miyata, S., Lee, D.G., Drenkard, E., Wu, G., Villanueva, J., Wei, T. and Ausubel, F.M. (2006) An ordered, nonredundant library of Pseudomonas aeruginosa strain PA14 transposon insertion mutants. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 2833–2838.
 90. Poulsen, B.E., Yang, R., Clatworthy, A.E., White, T., Osmulski, S.J., Li, L., Penaranda, C., Lander, E.S., Shoresh, N. and Hung, D.T. (2019) Defining the core essential genome of Pseudomonas aeruginosa. *Proc. Natl. Acad. Sci. U.S.A.*, **116**, 10072–10080.
 91. 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.
 92. 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.
 93. Su, Y., Xu, Y., Li, Q., Yuan, G. and Zheng, D. (2020) The essential genome of Ralstonia solanacearum. *Microbiol. Res.*, **238**, 126500.
 94. 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**, e00015-17.
 95. Pechter, K.B., Gallagher, L., Pyles, H., Manoel, C.S. and Harwood, C.S. (2015) Essential Genome of the Metabolically Versatile Alphaproteobacterium Rhodospseudomonas palustris. *J. Bacteriol.*, **198**, 867–876.
 96. Khatiwara, A., Jiang, T., Sung, S.S., Dawoud, T., Kim, J.N., Bhattacharya, D., Kim, H.B., Ricke, S.C. and Kwon, Y.M. (2012) Genome scanning for conditionally essential genes in Salmonella enterica Serotype Typhimurium. *Appl. Environ. Microbiol.*, **78**, 3098–3107.
 97. Knuth, K., Niesalla, H., Hueck, C.J. and Fuchs, T.M. (2004) Large-scale identification of essential Salmonella genes by trapping lethal insertions. *Mol. Microbiol.*, **51**, 1729–1744.
 98. Deutschbauer, A., Price, M.N., Wetmore, K.M., Shao, W., Baumohl, J.K., Xu, Z., Nguyen, M., Tamsen, 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. *PLoS Genet.*, **7**, e1002385.
 99. Forsyth, R.A., Haselbeck, R.J., Ohlsen, 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

- 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, I., 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 Burne, R.A. (2018) Genomewide 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, V., 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., Bootsma, H.J., De Jonge, M.I., Stockhofe-Zurwieden, N., Smith, H.E. and De Greeff, 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., Lowe, 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-Dieiguez, 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., Linteau, 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, R., 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 abiotic 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., Swindell, 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., Rhem, 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., Chandrashekar, M., Hustedt, 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., Hayles, 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.