

The 26th annual Nucleic Acids Research database issue and Molecular Biology Database Collection

Daniel J. Rigden^{1,*} and Xosé M. Fernández²

¹Institute of Integrative Biology, University of Liverpool, Crown Street, Liverpool L69 7ZB, UK and ²Institut Curie, 25 rue d'Ulm, 75005 Paris, France

ABSTRACT

The 2019 Nucleic Acids Research (NAR) Database Issue contains 168 papers spanning molecular biology. Among them, 64 are new and another 92 are updates describing resources that appeared in the Issue previously. The remaining 12 are updates on databases most recently published elsewhere. This Issue contains two Breakthrough articles, on the Virtual Metabolic Human (VMH) database which links human and gut microbiota metabolism with diet and disease, and Vibrism DB, a database of mouse brain anatomy and gene (co-)expression with sophisticated visualization and session sharing. Major returning nucleic acid databases include RNA-central, miRBase and LncRNA2Target. Protein sequence databases include UniProtKB, InterPro and Pfam, while wwPDB and RCSB cover protein structure. STRING and KEGG update in the section on metabolism and pathways. Microbial genomes are covered by IMG/M and resources for human and model organism genomics include Ensembl, UCSC Genome Browser, GENCODE and Flybase. Genomic variation and disease are well-covered by GWAS Catalog, PopHumanScan, OMIM and COSMIC, CADD being another major newcomer. Major new proteomics resources reporting here include iProX and jPOSTdb. The entire database issue is freely available online on the NAR website (<https://academic.oup.com/nar>). The NAR online Molecular Biology Database Collection has been updated, reviewing 506 entries, adding 66 new resources and eliminating 147 discontinued URLs, bringing the current total to 1613 databases. It is available at <http://www.oxfordjournals.org/nar/database/c>.

NEW AND UPDATED DATABASES

The Nucleic Acids Research (NAR) Database Issue reaches its 26th annual issue in 2019. As ever, the 168 papers within cover the full range of biological research. Among them,

entirely new databases account for 64 (Table 1) while 92 cover resources that have previously appeared in the Issue and now return with updates. The remaining 12 papers are updated on databases last published elsewhere (Table 2). The usual categorization is again used: after reports from the major resource collections at the U.S. National Center for Biotechnology Information (NCBI), the European Bioinformatics Institute (EBI) and the BIG Data Center at the Beijing Institute of Genomics, Chinese Academy of Sciences there are these groupings: (i) nucleic acid sequence and structure, transcriptional regulation; (ii) protein sequence and structure; (iii) metabolic and signaling pathways, enzymes and networks; (iv) genomics of viruses, bacteria, protozoa and fungi; (v) genomics of human and model organisms plus comparative genomics; (vi) human genomic variation, diseases and drugs; (vii) plants and (viii) other topics, such as proteomics databases. Many interdisciplinary databases defy easy categorization, encouraging readers to browse the whole issue. The NAR online Molecular Biology Database Collection, classifies databases more finely using 15 categories and 41 subcategories, and can be found at <http://www.oxfordjournals.org/nar/database/c>.

Among the major global centers, the NCBI (1) reports on new and expanded literature resources, including PubMed Labs (2) a new interface to PubMed, and new sequence database search options. The EBI paper (3) reports on the new databases Single Cell Expression Atlas and PDBe-Knowledgebase. The latter encompasses FunPDBe, an initiative to better harness structural bioinformatics methods and international collaborators to annotate the protein structural data in PDBe. An interesting facility reported by the BIG Data Center paper (4) is their BIG Search which not only scans across the Center's many resources but accesses indexes from non-Center partner databases on topics as diverse as lncRNAs, plant transcription factors and autophagy-related proteins.

Major returning resources in the 'Nucleic acid databases' section include miRBase (5) which focuses on criteria to assess the reliability of microRNA entries and functional annotation from linked target predictions, external manual curation and text mining. For long non-coding RNAs and their targets, LNCipedia (6) contributes an update, also with a major focus on text mining and manual curation. The

*To whom correspondence should be addressed. Email: nardatabase@gmail.com

Table 1. Descriptions of new online databases in the 2019 NAR Database issue

Database	URL	Brief description ^a
AleDB	http://aledb.org	Mutations from Adaptive Laboratory Evolution experiments
AlloMAPS	http://allomaps.bii.a-star.edu.sg	Allosteric signaling and mutations in proteins
AmtDB	https://amtdb.org	Ancient mitochondrial DNA
Ancestral Genomes	http://ancestralgenomes.org	Reconstructed ancestral genomes
AWESOME	http://www.awesome-hust.com	Impact of SNPs on post-translational modifications
Bactome	https://bactome.helmholtz-hzi.de	Sequences, transcriptomes and phenotypes of clinical isolates of <i>Pseudomonas auruginosa</i>
CAGm	http://www.cagmdb.org	Catalog of all germline microsatellites
CancerSEA	http://biocc.hrbmu.edu.cn/CancerSEA	Cancer single-cell state atlas
CancerSplicingQTL	http://www.cancersplicingqtl-hust.com	Splicing quantitative trait loci in cancer
CellMarker	http://biocc.hrbmu.edu.cn/CellMarker	Cell markers in human and mouse
Cell Model Passports	https://cellmodelpassports.sanger.ac.uk	Human cancer cell models
ChIPprimersDB	http://www.chipprimers.com	qPCR oligonucleotide primers for chromatin immunoprecipitation (ChIP)
CMAUP	http://bidd2.nus.edu.sg/CMAUP	Collective molecular activities of useful plants
CoevDB	http://phylobd.unil.ch/CoevDB	Pairwise nucleotide coevolution
CRISPRInc	http://crisprinc.xtbg.ac.cn	Validated CRISPR/Cas9 sgRNAs for model organism lncRNAs
Cucurbit Genomics Database	http://cucurbitgenomics.org	Genomics of the Cucurbitaceae family
dbAMP	http://csb.cse.yzu.edu.tw/dbAMP	Antimicrobial peptide sequences, structures and properties
DSMNC	http://dsmnc.big.ac.cn	Database of somatic mutations in normal cells
EDK	http://bigd.big.ac.cn/edk	Editome disease knowledgebase
EncoMPASS	https://encompass.ninds.nih.gov	Encyclopedia of membrane proteins analyzed by structure and symmetry.
EndoDB	https://vibcancer.be/software-tools/endodb	Gene expression in endothelial cells
ENPD	http://qinlab.sls.cuhk.edu.hk/ENPD	Eukaryotic nucleic acid binding proteins
ETCM	http://www.ehbio.com/ETCM	Encyclopedia of Traditional Chinese Medicine
EVmiRNA	http://bioinfo.life.hust.edu.cn/EVmiRNA	miRNA in extracellular vesicles
EWAS Atlas	http://bigd.big.ac.cn/ewas	Epigenome-Wide Association Studies
EWASdb	http://www.bioapp.org/ewas	Epigenome-Wide Association Studies Atlas
FusionGDB	https://ccsm.uth.edu/FusionGDB	Fusion Gene annotation DataBase
gcMeta	https://gcmeta.wdcm.org	Microbiome research data
Genome Properties	https://www.ebi.ac.uk/interpro/genomeproperties	Pathways and other properties represented by sets of protein families
HACER	http://bioinfo.vanderbilt.edu/AE/HACER	Human ACtive Enhancer to interpret Regulatory variants
HmtVar	https://www.hmtvar.uniba.it	Human mitochondrial DNA variants
iDOG	http://bigd.big.ac.cn/idoq	Dog data on genes, SNPs, diseases etc.
iProX	http://www.iprox.org	Proteomics datasets
jPOSTdb	https://jpostdb.org	Japan ProteOme STandard environment
KinaMetrix	http://kinametrix.com	Protein kinase models, conformations and ligands
liqDB	http://bioinfo5.ugr.es/liqdb	Small RNA expression profiles in biofluids
LncBook	http://bigd.big.ac.cn/lncbook	Human lncRNA knowledgebase
MemProtMD	http://memprotmd.bioch.ox.ac.uk	Membrane Proteins Embedded in Lipid Bilayers
MethMotif	http://bioinfo-csi.nus.edu.sg/methmotif	Transcription factor binding motifs coupled with DNA methylation profiles
NucMap	http://bigd.big.ac.cn/NucMap	Nucleosome positioning map across species
OncoBase	http://www.oncobase.biols.ac.cn	Regulatory somatic mutations in human cancers
OpenProt	http://www.openprot.org	Eukaryotic coding potential and proteomes
Pancan-meQTL	http://bioinfo.life.hust.edu.cn/Pancan-meQTL	Methylation quantitative trait loci in cancer
PDX Finder	http://www.pdxfinder.org	Patient-derived xenograft models
PED	http://bigd.big.ac.cn/ped	Plant Editosome Database
PreMedKB	http://fudan-pgx.org/premedkb	Precision medicine knowledgebase
piRTarBase	http://cosbi5.ee.ncku.edu.tw/piRTarBase	piRNA targeting sites
Plasmid Atlas	http://www.patlas.site	Plasmid annotations and metadata
PLSDB	https://ccb-microbe.cs.uni-saarland.de/plsdb	Plasmid annotations and metadata
PopHumanScan	https://pophumanscan.uab.cat	Positively selected regions of the human genome
qPhos	http://qphos.cancerbio.info	Protein phosphorylation dynamics
RetroRules	http://retrorules.org	Reaction rules for synthetic biology
RNact	http://rnect.crg.eu	Experimental and predicted protein-RNA interactions in human and mouse
SAGD	http://bioinfo.life.hust.edu.cn/SAGD	Sex-associated gene database
SEdb	http://www.licpathway.net/sedb	Super-enhancer database
SSRome	http://mggm-lab.easymomics.org	Microsatellites in all organisms
SymMap	http://www.symmap.org	Traditional chinese medicine including symptom mapping
Translocatome	http://translocatome.linkgroup.hu	Translocating human proteins
Trips-Viz	http://trips.ucc.ie	Transcriptome browser
UniLectin	http://www.unilectin.eu	Lectin structure and function
ViBriSM DB	https://vibrism.neuroinf.jp	Tomographic transcriptome and co-expression networks in mouse brain
Victors	http://www.phidias.us/victors	Virulence factors
VMH	http://vmh.life	Virtual metabolic human
YeastRGB	http://www.yeastRGB.org	Expression and localization of yeast proteins

^aFor full references to the databases featured in this issue, please see the Table of Contents.

Table 2. Updated descriptions of databases most recently published elsewhere

Database	URL	Brief description ^a
CADD	http://cadd.gs.washington.edu	Combined annotation-dependent depletion
GENCODE	http://www.genecodegenes.org	Reference annotation for the human and mouse genomes.
glycosciences.DB	http://www.glycosciences.de/database	Glyco-related databases
Haemopedia	http://haemosphere.org	Haematopoietic expression data
HumanNet	https://www.inetbio.org/humannet	Human gene functional network
LncACTdb	http://www.bio-bigdata.net/LncACTdb	lncRNA–miRNA–gene interactions
MoonDB	http://moondb.hb.univ-amu.fr	Known and predicted multifunctional proteins in model organisms
OrthoInspector	http://lbgf.fr/orthoinspectorv3	Orthologous relations and phylogenetic profiling
piRBase	http://www.regulatoryrna.org	piRNA function and annotation
Stemformatics	http://stemformatics.org	Stem cell and other cell-specific gene expression
UNITE	https://unite.ut.ee	Internal transcribed spacer sequences for fungal identification
Vesiclepedia	http://www.microvesicles.org	Extracellular vesicles

^aFor full references to the databases featured in this issue, please see the Table of Contents.

popular LncRNA2Target database (7) reports a new release with major increases in lncRNAs, targets and lncRNA–target associations. Two papers address piRNAs (the returning piRBase; 8) or their targets (the newcomer piRTarBase; 9). The RNACentral (10) hub now retrieves ncRNA data from a total of 28 databases. Important progress since the last paper is reported in mapping genome locations, quality control using the Rfam database (11) and functional annotation. Elsewhere two new databases—Plasmid Atlas (12) and PLSDB (13)—allow easy analyses and searches against the ever-increasing number of bacterial plasmid sequences. In resources for transcription factors (TFs), AnimalTFDB (14) now covers 97 animal genomes and includes a variety of new data such as links from TF-SNP pairs to GWAS data, TF gene expression data and protein-protein interaction networks involving TFs. An interesting new database MethMotif (15) integrates TF binding sites with data on DNA methylation, demonstrating the cell type specificity of many TFs in terms of both binding site sequences and methylation profiles.

In the section on protein sequence and structure databases, UniProtKB (16) reports continued exponential growth of much data, growth made manageable by a focus on Reference Proteomes. There is an interesting discussion on the importance of primary manual curation of the Swiss-Prot portion of the database, especially in cases that computational methods would struggle with, and mention of new methods that might contribute to better propagation of that information to the unreviewed UniProtKB/TrEMBL section. Some of these methods use domain assignments from InterPro (17), also contributing an update here describing, among other improvements, annotation with ‘flavors’ of intrinsic disorder and better treatment of discontinuous domains. A contributor to InterPro and major resource in its own right Pfam also has an update paper (18). It reports refinement of many existing entries and the generation of over 800 new protein families using the ECOD structural database (19). This exercise helped refine the description of some Domains of Unknown Function, while the paper also explains how useful annotations for these can flow back from InterPro work to integrate and rationalize content across its multiple contributing databases.

For protein structure both the wwPDB consortium (20) and the RCSB (21) report updates, the former pointing out the deposition and validation challenges brought by cryo-EM and serial femtosecond crystallography, the latter listing the impressive variety of external resources integrated into the webpages and describing the incorporation of a new method for description of biological assemblies. Protein post-translational modifications are well-covered by the returning PhosphoSitePlus[®] database (22), now 15 years old and with exciting new integration of disease-related mutations and protein isoform data, and the iEKPD resource (23) for phosphorylation-related protein domains. They are joined by the new database qPhos (24) covering the dynamics of protein phosphorylation. Among other interesting new arrivals are the Ancestral Genomes database (25), providing reconstructed proteomes for 78 extinct ancestors of current species and two new resources for transmembrane proteins: EncoMPASS (26) focusing on structural similarities and symmetries, and MemProtMD (27) which provides the results of embedding the structures in a lipid bilayer and subsequent coarse-grained dynamics simulations.

A major new arrival in the metabolic and signaling section is the Issue’s first Breakthrough Article. It is increasingly apparent that the gut microbiota and human diet interact in complex ways with human host metabolism to influence health and disease. The Virtual Metabolic Human (VMH) database (28) is an impressively ambitious resource that seeks to capture that complexity through linking together modules containing genes, reactions and chemical compounds within the human cell (both as a whole and in compartments) and gut microbes (more than 600 species). Further resources cover nutrition, both in terms of typical diets and in mapping dietary components onto VMH metabolites. As well as inter-connectedness between these modules, VMH links out to more than 50 other databases. The authors envisage that simulations employing VMH with different diets and different microbial abundances can be used to generate testable hypotheses, for example, regarding correlations between microbiota composition and disease states.

Elsewhere, two returning manually curated databases focus on protein complexes. CORUM (29) covers mammalian

complexes and new features include a network-style visualization of subunit interactions and, most interestingly, a recognition and accounting of the impact alternative splicing can have on protein complex function. Complex Portal (30) presents a new interface incorporating visualizations of data from other databases on metabolism, protein structure and gene expression. The very popular database of protein–protein functional associations STRING reports (31) an update to version 11.0. Not only has the number of species covered doubled since the previous version, but the database now allows genome-scale expression dataset uploads and annotation of the resulting networks according to gene-set enrichment analysis. The well-used HumanNet (32), comprising a network of human gene associations with data weighted in a Bayesian framework, reports an update. Already employed data such as protein–protein interactions increased significantly in volume and were supplemented by two novel sources of data, pathway annotations and co-essentiality data. New candidates for involvement in disease can be identified by network-based expansion from a set of known guide genes. Finally, the KEGG database, particularly valued for its pathway reconstructions, reports important new developments in the shape of the KEGG NETWORK and KEGG VARIANT components (33). These human-specific elements allow the integration of variants such as cancer-related mutations of signaling proteins into networks (derived from KEGG's original pathways) in order to visualize the effects of perturbation on disease-related pathways. A network variation map summarizes the impacts of a set of perturbants—which may also include viral proteins, environmental factors and drugs—on a given pathway.

The microbial genomics section contains updates from twin Joint Genome Institute databases. The IMG/M database of genomes, metagenomes and metatranscriptomes reports growth of around 60% in just 2 years (34). Its interface is improved in a variety of ways, including a powerful new search capability, BLAST searches against specific and bespoke databases, and powerful statistical comparisons of gene function between groups. IMG/M links to genomic metadata in the longstanding GOLD database, also reporting an update here (35). The IMG/VR viral genomics resource has tripled in size in the same period and now includes improved viral host prediction and geographic mapping for uncultivated viral genomes (36). Two virulence factor (VF) databases are included this year. The first, the well-established VFDB (37) reports a new automated tool, VFAnalyzer, for the identification of VFs in complete or draft bacterial genomes. It combines sequence similarity searches with a consideration of genomic context to achieve performance that is reportedly comparable with human curation. The new Victors database (38) has manually curated information from over 5000 VFs and a broader scope—encompassing bacterial, viral, parasitic and fungal VFs—than comparable resources.

Human and model organism genomics again has a strong presence, starting with the second of the Issue's Breakthrough Articles. VibriDB (39) contains anatomical, gene expression and co-expression data at different ages of the mouse brain presenting the information in ways including interactive 3D visualization in a browser. More than

170 000 individual expression maps covering coding and non-coding transcripts are included and co-expression can be viewed in anatomical context or in a network format. The database covers many more transcripts than similar resources, requires few brains to profile expression and cleverly allows users to share URLs that encode scene-setting parameters enabling easy sharing of visualizations. Elsewhere in the section, the major resources Ensembl (40) and the UCSC Genome Browser (41) present their usual updates and are joined by GENCODE (42). A new arrival, the Trips-Viz transcriptome browser (43) allows for mapping of Ribo-Seq and mRNA-seq data onto individual transcripts from seven model organisms. Popular model organism databases such as FlyBase (44), ZFIN (45) and PlanMine (46) are joined by the new resource iDOG (47) which contains an impressive variety of data relating to domestic dogs and other canids, and which is designed to appeal to interested lay people as well as expert researchers. The ArrayExpress functional genomics database marks 15 years since its first appearance in this journal with an update (48) reporting the increasing proportion of submissions coming from single cell studies and the challenges of capturing appropriate metadata.

A large number of databases in the areas of human genomic variation, diseases and drugs are included in this Issue. The popular returning database GWAS Catalog (49) now includes studies from over 3500 publications, while a pair of new databases—EWAS Atlas (50) and EWASdb (51)—each service the fast-growing area of Epigenome-Wide Association Studies. The new database PopHumanScan (52) processes population genomics data from last year's PopHuman database (53) revealing almost 3000 candidate human genome regions under positive selection. They are linked to relevant literature in the database while users are also invited to contribute their own candidate regions for curation. CADD (54) is a major new arrival and is a very popular measure for predicting the deleteriousness of genome variants. Its classification model is trained using a large number of features, including protein, genome and epigenome information. The deleteriousness of both single nucleotide variants and short insertions or deletions can be predicted. A new database, AWESOME (55), specifically addresses the impact of single nucleotide variants on protein post-translational modifications, considering the effects of around 1 million variants from dbSNP (56) on six different modifications. The venerable OMIM database, linking genes and phenotypes such as inherited disorders, reports continued strong growth of around 300 new phenotypes per year such that over 6000 phenotypes are now linked to around 4000 genes (57). Several cancer databases are covered, including the popular curated catalogue of cancer mutations COSMIC (58). The 3D protein structural consequences of cancer-related mutations are covered in COSMIC-3D and in the dedicated resource Cancer3D (59). The roles of long non-coding RNAs in cancer are the subject of the returning database Lnc2Cancer (60), while disease in general is covered by the similarly popular and hugely expanded LncRNADisease (61). *In vivo* study of cancers will be facilitated by the new Cell Model Passports database (62) which contains standardized infor-

mation for over 1200 cell models enabling a rational choice of cell model for a particular purpose to be made.

Plant databases include the return of the Genome Database for Rosaceae, celebrating 15 years (63), and the new arrival of the Cucurbit Genomics Database (64), covering important crops such as pumpkin, melon and cucumber. Another new arrival, CMAUP (65), aims to catalogue the chemical compounds present in plants of different kinds (edible, medicinal, garden etc.) and distributions, linking them to effects on human proteins, pathways and diseases. It includes over 5000 plants and almost 50 000 ingredients. Two important glycobiology databases are included in the last section. The returning glycosciences.DB (66) reports a new interface and new search options, including the ability to search with glycan (sub-)structures. The newcomer UniLectin3D (67) contains structural information on lectins, their ligands and their interactions. It features a wide range of visualization options and ample links out to related resources. This section also contains two major proteomics databases: iProX (68), a new Chinese contributor to the ProteomeXchange consortium and jPOSTdb (69). Following on from jPOSTrepo, also published in the Database Issue (70), jPOSTdb subjects the original raw data to a standardized protocol. Post-translational modifications and protein isoforms can be visualized with further options allowing differential expression analysis and protein set enrichment with respect to KEGG pathways and Gene Ontology terms. Finally, it is always a pleasure to welcome databases addressing entirely new kinds of data. Such is the case of AleDB (71) which records the results of Adaptive Laboratory Evolution experiments whereby microbes are grown in defined conditions and the mutations responsible for improved phenotypes are tracked.

NAR ONLINE MOLECULAR BIOLOGY DATABASE COLLECTION

With the 26th release of the NAR online Molecular Biology Database Collection (which is freely available at <http://www.oxfordjournals.org/nar/database/c>), we feature 66 new resources. We continue to monitor the collection to ensure the information is still relevant and resources are running, contacting authors when repeated downtime is detected. Thanks to this verification process we have updated 506 database entries, removing 147 obsolete or discontinued databases.

We are happy to include new databases in the Collection and we encourage authors of resources published elsewhere to contact us. Such suggestions should be addressed to XMF at xose.m.fernandez@gmail.com and should include database summaries in plain text, organized in accordance with the <http://www.oxfordjournals.org/nar/database/summary/1> template.

ACKNOWLEDGEMENTS

We thank Dr Martine Bernardes-Silva, especially, and the rest of the Oxford University Press team led by Joanna Ventikos and Elisabeth Waelkens for their help in compiling this issue.

FUNDING

Funding for open access charge: Oxford University Press. *Conflict of interest statement.* The authors' opinions do not necessarily reflect the views of their respective institutions.

REFERENCES

- Sayers, E.W., Agarwala, R., Bolton, E.E., Rodney Brister, J., Canese, K., Clark, K., Connor, R., Fiorini, N., Funk, K., Hefferon, T. *et al.* (2018) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, doi: 10.1093/nar/gky1069.
- Fiorini, N., Canese, K., Bryzgunov, R., Radetska, I., Gindulyte, A., Lattner, M., Miller, V., Osipov, M., Kholodov, M., Starchenko, G. *et al.* (2018) PubMed labs: an experimental system for improving biomedical literature search. *Database (Oxford)*, **2018**, doi: 10.1093/database/bay094.
- Cook, C.E., Lopez, R., Stroe, O., Cochrane, G., Brooksbank, C., Birney, E. and Apweiler, R. (2018) The European Bioinformatics Institute in 2018: tools, infrastructure and training. *Nucleic Acids Res.*, doi: 10.1093/nar/gky1124.
- BIG Data Center Members. (2018) Database Resources of the BIG Data Center in 2019. *Nucleic Acids Res.*, doi: 10.1093/nar/gky993.
- Kozomara, A., Birgaoanu, M. and Griffiths-Jones, S. (2018) miRBase: from microRNA sequences to function. *Nucleic Acids Res.*, doi: 10.1093/nar/gky1141.
- Volders, P.-J., Anckaert, J., Verheggen, K., Nuytens, J., Martens, L., Mestdagh, P. and Vandesompele, J. (2018) LNCipedia 5: towards a reference set of human long non-coding RNAs. *Nucleic Acids Res.*, doi: 10.1093/nar/gky1031.
- Cheng, L., Wang, P., Tian, R., Wang, S., Guo, Q., Luo, M., Zhou, W., Liu, G., Jiang, H. and Jiang, Q. (2018) LncRNA2Target v2.0: a comprehensive database for target genes of lncRNAs in human and mouse. *Nucleic Acids Res.*, doi: 10.1093/nar/gky1051.
- Wang, J., Zhang, P., Lu, Y., Li, Y., Zheng, Y., Kan, Y., Chen, R. and He, S. (2018) piRBase: a comprehensive database of piRNA sequences. *Nucleic Acids Res.*, doi: 10.1093/nar/gky1043.
- Wu, W.-S., Brown, J.S., Chen, T.-T., Chu, Y.-H., Huang, W.-C., Tu, S. and Lee, H.-C. (2018) piRTarBase: a database of piRNA targeting sites and their roles in gene regulation. *Nucleic Acids Res.*, doi: 10.1093/nar/gky956.
- The RNAcentral Consortium (2018) RNAcentral: a hub of information for non-coding RNA sequences. *Nucleic Acids Res.*, doi: 10.1093/nar/gky1034.
- Kalvari, I., Argasinska, J., Quinones-Olvera, N., Nawrocki, E.P., Rivas, E., Eddy, S.R., Bateman, A., Finn, R.D. and Petrov, A.I. (2018) Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. *Nucleic Acids Res.*, **46**, D335–D342.
- Jesus, T.F., Ribeiro-Gonçalves, B., Silva, D.N., Bortolai, V., Ramirez, M. and Carriço, J.A. (2018) Plasmid ATLAS: plasmid visual analytics and identification in high-throughput sequencing data. *Nucleic Acids Res.*, doi: 10.1093/nar/gky1073.
- Galata, V., Fehlmann, T., Backes, C. and Keller, A. (2018) PLSDB: a resource of complete bacterial plasmids. *Nucleic Acids Res.*, doi: 10.1093/nar/gky1050.
- Hu, H., Miao, Y.-R., Jia, L.-H., Yu, Q.-Y., Zhang, Q. and Guo, A.-Y. (2018) AnimalTFDB 3.0: a comprehensive resource for annotation and prediction of animal transcription factors. *Nucleic Acids Res.*, doi: 10.1093/nar/gky822.
- Lin, Q.X.X., Sian, S., An, O., Thieffry, D., Jha, S. and Benoukraf, T. (2018) MethMotif: an integrative cell specific database of transcription factor binding motifs coupled with DNA methylation profiles. *Nucleic Acids Res.*, doi: 10.1093/nar/gky1005.
- The UniProt Consortium. (2018) UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.*, doi: 10.1093/nar/gky1049.
- Mitchell, A.L., Attwood, T.K., Babbitt, P.C., Blum, M., Bork, P., Bridge, A., Brown, S.D., Chang, H.-Y., El-Gebali, S., Fraser, M.I. *et al.* (2018) InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res.*, doi: 10.1093/nar/gky1100.
- El-Gebali, S., Mistry, J., Bateman, A., Eddy, S., Luciani, A., Potter, S.C., Qureshi, M., Richardson, L.J., Salazar, G.A. and, A. Smart (2018) The

- Pfam protein families database in 2019. *Nucleic Acids Res.*, doi: 10.1093/nar/gky995.
19. Schaeffer, R.D., Liao, Y., Cheng, H. and Grishin, N.V. (2017) ECOD: New developments in the evolutionary classification of domains. *Nucleic Acids Res.*, **45**, D296–D302.
 20. wwPDB consortium. (2018) Protein Data Bank: the single global archive for 3D macromolecular structure data. *Nucleic Acids Res.*, doi: 10.1093/nar/gky949.
 21. Burley, S.K., Berman, H.M., Bhikadiya, C., Bi, C., Chen, L., Costanzo, L.D., Christie, C., Dalenberg, K., Duarte, J.M. and Dutta, S. (2018) RCSB Protein Data Bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. *Nucleic Acids Res.*, doi: 10.1093/nar/gky1004.
 22. Hornbeck, P.V., Kornhauser, J.M., Latham, V., Murray, B., Nandhikonda, V., Nord, A., Skrzypek, E., Wheeler, T., Zhang, B. and Gnad, F. (2018) 15 years of PhosphoSitePlus®: integrating post-translationally modified sites, disease variants and isoforms. *Nucleic Acids Res.*, doi: 10.1093/nar/gky1159.
 23. Guo, Y., Peng, D., Zhou, J., Lin, S., Wang, C., Ning, W., Xu, H., Deng, W. and Xue, Y. (2018) iEKPD 2.0: an update with rich annotations for eukaryotic protein kinases, protein phosphatases and proteins containing phosphoprotein-binding domains. *Nucleic Acids Res.*, doi: 10.1093/nar/gky1063.
 24. Yu, K., Zhang, Q., Liu, Z., Zhao, Q., Zhang, X., Wang, Y., Wang, Z.-X., Jin, Y., Li, X., Liu, Z.-X. *et al.* (2018) qPhos: a database of protein phosphorylation dynamics in humans. *Nucleic Acids Res.*, doi: 10.1093/nar/gky1052.
 25. Huang, X., Albou, L.-P., Mushayama, T., Muruganujan, A., Tang, H. and Thomas, P.D. (2018) Ancestral Genomes: a resource for reconstructed ancestral genes and genomes across the tree of life. *Nucleic Acids Res.*, doi: 10.1093/nar/gky1009.
 26. Sarti, E., Aleksandrova, A.A., Ganta, S.K., Yavatkar, A.S. and Forrest, L.R. (2018) EnCoMPASS: an online database for analyzing structure and symmetry in membrane proteins. *Nucleic Acids Res.*, doi: 10.1093/nar/gky952.
 27. Newport, T.D., Sansom, M.S.P. and Stansfeld, P.J. (2018) The MemProtMD database: a resource for membrane-embedded protein structures and their lipid interactions. *Nucleic Acids Res.*, doi: 10.1093/nar/gky1047.
 28. Noronha, A., Modamio, J., Jarosz, Y., Guerard, E., Sompairac, N., Preciat, G., Danielsdóttir, A.D., Krecke, M., Merten, D., Haraldsdóttir, H.S. *et al.* (2018) The Virtual Metabolic Human database: integrating human and gut microbiome metabolism with nutrition and disease. *Nucleic Acids Res.*, doi: 10.1093/nar/gky992.
 29. Giurgiu, M., Reinhard, J., Brauner, B., Dunger-Kaltenbach, I., Fobo, G., Frishman, G., Montrone, C. and Ruepp, A. (2018) CORUM: the comprehensive resource of mammalian protein complexes—2019. *Nucleic Acids Res.*, doi: 10.1093/nar/gky973.
 30. Meldal, B.H.M., Bye-A-Jee, H., Gajdoš, L., Hammerová, Z., Horáčková, A., Melicher, F., Perfetto, L., Pokorný, D., Lopez, M.R., Türková, A. *et al.* (2018) Complex Portal 2018: extended content and enhanced visualization tools for macromolecular complexes. *Nucleic Acids Res.*, doi: 10.1093/nar/gky1001.
 31. Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N.T., Morris, J.H., Bork, P. *et al.* (2018) STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.*, doi: 10.1093/nar/gky1131.
 32. Hwang, S., Kim, C.Y., Yang, S., Kim, E., Hart, T., Marcotte, E.M. and Lee, I. (2018) HumanNet v2: human gene networks for disease research. *Nucleic Acids Res.*, doi: 10.1093/nar/gky1126.
 33. Kanehisa, M., Sato, Y., Furumichi, M., Morishima, K. and Tanabe, M. (2018) New approach for understanding genome variations in KEGG. *Nucleic Acids Res.*, doi: 10.1093/nar/gky962.
 34. Chen, I.-Min, A., Chu, K., Palaniappan, K., Pillay, M., Ratner, A., Huang, J., Huntemann, M., Varghese, N., White, J.R., Seshadri, R. *et al.* (2018) IMG/M v5.0: an integrated data management and comparative analysis system for microbial genomes and microbiomes. *Nucleic Acids Res.*, doi: 10.1093/nar/gky901.
 35. Mukherjee, S., Stamatis, D., Bertsch, J., Ovchinnikova, G., Katta, H.Y., Mojica, A., Chen, I.-Min, A., Kyripides, N.C. and Reddy, T.B.K. (2018) Genomes OnLine database (GOLD) v7: updates and new features. *Nucleic Acids Res.*, doi: 10.1093/nar/gky977.
 36. Paez-Espino, D., Roux, S., Chen, I.-Min, A., Palaniappan, K., Ratner, A., Chu, K., Huntemann, M., Reddy, T.B.K., Pons, J.C., Llabrés, M. *et al.* (2018) IMG/VR v2.0: an integrated data management and analysis system for cultivated and environmental viral genomes. *Nucleic Acids Res.*, doi: 10.1093/nar/gky1127.
 37. Liu, B., Cheng, D., Jin, Q., Chen, L. and Yang, J. (2018) VFDB 2019: a comparative pathogenomic platform with an interactive web interface. *Nucleic Acids Res.*, doi: 10.1093/nar/gky1080.
 38. Sayers, S., Li, L., Ong, E., Deng, S., Fu, G., Lin, Y., Yang, B., Zhang, S., Fa, Z., Zhao, B., Xiang, Z. *et al.* (2018) Victors: a web-based knowledge base of virulence factors in human and animal pathogens. *Nucleic Acids Res.*, doi: 10.1093/nar/gky999.
 39. Morita, M., Shimokawa, K., Nishimura, M., Nakamura, S., Tsujimura, Y., Takemoto, S., Tawara, T., Yokota, H., Wemler, S., Miyamoto, D. *et al.* (2018) ViBrism DB: an interactive search and viewer platform for 2D/3D anatomical images of gene expression and co-expression networks. *Nucleic Acids Res.*, doi: 10.1093/nar/gky951.
 40. Cunningham, F., Achuthan, P., Akanni, W., Allen, J., Amode, M., Armean, I.M., Bennett, R., Bhai, J., Billis, K. and Boddu, S. (2018) Ensembl 2019. *Nucleic Acids Res.*, doi: 10.1093/nar/gky1113.
 41. Haussler, M., Zweig, A.S., Tyner, C., Speir, M.L., Rosenbloom, K.R., Raney, B.J., Lee, C.M., Lee, B.T., Hinrichs, A.S., Gonzalez, J.N. *et al.* (2018) The UCSC Genome Browser database: 2019 update. *Nucleic Acids Res.*, doi: 10.1093/nar/gky1095.
 42. Frankish, A., Diekhans, M., Ferreira, A.-M., Johnson, R., Jungreis, I., Loveland, J., Mudge, S., Sis, C., Wright, J., Armstrong, J. *et al.* (2018) GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res.*, doi: 10.1093/nar/gky955.
 43. Kiniry, S.J., O'Connor, P.B.F., Michel, A.M. and Baranov, P.V. (2018) Trips-Viz: a transcriptome browser for exploring Ribo-Seq data. *Nucleic Acids Res.*, doi: 10.1093/nar/gky842.
 44. Thurmond, J., Goodman, J.L., Strelets, V.B., Attrill, H., Sian Gramates, L., Marygold, S.J., Matthews, B.B., Millburn, G., Antonazzo, G., Trovisco, V. *et al.* (2018) FlyBase 2.0: the next generation. *Nucleic Acids Res.*, doi: 10.1093/nar/gky1003.
 45. Ruzicka, L., Howe, D.G., Ramachandran, S., Toro, S., Slyke, C.E., Bradford, Y.M., Eagle, A., Fashena, D., Frazer, K., Kalita, P. *et al.* (2018) The Zebrafish Information Network: new support for non-coding genes, richer Gene Ontology annotations and the Alliance of Genome Resources. *Nucleic Acids Res.*, doi: 10.1093/nar/gky1090.
 46. Rozanski, A., Moon, H., Brandl, H., Martín-Durán, J.M., Grohne, M.A., Hüttner, K., Bartscherer, K., Henry, I. and Rink, J.C. (2018) PlanMine 3.0—improvements to a mineable resource of flatworm biology and biodiversity. *Nucleic Acids Res.*, doi: 10.1093/nar/gky1070.
 47. Tang, B., Zhou, Q., Dong, L., Li, W., Zhang, X., Lan, L., Zhai, S., Xiao, J., Zhang, Z., Bao, Y. *et al.* (2018) iDog: an integrated resource for domestic dogs and wild canids. *Nucleic Acids Res.*, doi: 10.1093/nar/gky1041.
 48. Athar, A., Füllgrabe, A., George, N., Iqbal, H., Huerta, L., Ali, A., Snow, C., Fonseca, Petryszak, R., Papatheodorou, I. *et al.* (2018) ArrayExpress update—from bulk to single-cell expression data. *Nucleic Acids Res.*, doi: 10.1093/nar/gky964.
 49. Buniello, A., MacArthur, J.A.L., Cerezo, M., Harris, L.W., Hayhurst, J., Malancane, C., McMahon, A., Morales, J., Mountjoy, E., Sollis, E. *et al.* (2018) The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.*, doi: 10.1093/nar/gky1120.
 50. Li, M., Zou, D., Li, Z., Gao, R., Sang, J., Zhang, Y., Li, R., Xia, L., Zhang, T., Niu, G. *et al.* (2018) EWAS Atlas: a curated knowledgebase of epigenome-wide association studies. *Nucleic Acids Res.*, doi: 10.1093/nar/gky1027.
 51. Liu, D., Zhao, L., Wang, Z., Zhou, X., Fan, X., Li, Y., Xu, J., Hu, S., Niu, M., Song, X. *et al.* (2018) EWASdb: epigenome-wide association study database. *Nucleic Acids Res.*, doi: 10.1093/nar/gky942.
 52. Murga-Moreno, J., Coronado-Zamora, M., Bodelón, A., Barbadilla, A. and Casillas, S. (2018) PopHumanScan: the online catalog of human genome adaptation. *Nucleic Acids Res.*, doi: 10.1093/nar/gky959.
 53. Casillas, S., Mulet, R., Villegas-Miron, P., Hervas, S., Sanz, E., Velasco, D., Bertranpetit, J., Laayouni, H. and Barbadilla, A. (2018) PopHuman: The human population genomics browser. *Nucleic Acids Res.*, **46**, D1003–D1010.

54. Rentschsch,P., Witten,D., Cooper,G.M., Shendure,J. and Kircher,M. (2018) CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.*, doi: 10.1093/nar/gky1016.
55. Yang,Y., Peng,X., Ying,P., Tian,J., Li,J., Ke,J., Zhu,Y., Gong,Y., Zou,D. and Yang,N. (2018) AWESOME: a database of SNPs that affect protein post-translational modifications. *Nucleic Acids Res.*, doi: 10.1093/nar/gky821.
56. Sherry,S.T., Ward,M.H., Kholodov,M., Baker,J., Phan,L., Smigielski,E.M. and Sirotkin,K. (2001) dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.*, **29**, 308–311.
57. Amberger,J.S., Bocchini,C.A., Scott,A.F. and Hamosh,A. (2018) OMIM.org: leveraging knowledge across phenotype–gene relationships. *Nucleic Acids Res.*, doi: 10.1093/nar/gky1151.
58. Tate,J.G., Bamford,S., Jubb,H.C., Sondka,Z., Beare,D.M., Bindal,N., Boutselakis,H., Cole,C.G., Creatore,C., Dawson,E. *et al.* (2018) COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res.*, doi: 10.1093/nar/gky1015.
59. Sedova,M., Iyer,M., Li,Z., Jaroszewski,L., Post,K.W., Hrabe,T., Porta-Pardo,E. and Godzik,A. (2018) Cancer3D 2.0: interactive analysis of 3D patterns of cancer mutations in cancer subsets. *Nucleic Acids Res.*, doi: 10.1093/nar/gky1098.
60. Gao,Y., Wang,P., Wang,Y., Ma,X., Zhi,H., Zhou,D., Li,X., Fang,Y., Shen,W., Xu,Y. *et al.* (2018) Lnc2Cancer v2.0: updated database of experimentally supported long non-coding RNAs in human cancers. *Nucleic Acids Res.*, doi: 10.1093/nar/gky1096.
61. Bao,Z., Yang,Z., Huang,Z., Zhou,Y., Cui,Q. and Dong,D. (2018) LncRNADisease 2.0: an updated database of long non-coding RNA-associated diseases. *Nucleic Acids Res.*, doi: 10.1093/nar/gky905.
62. van der Meer,D., Barthorpe,S., Yang,W., Lightfoot,H., Hall,C., Gilbert,J., Francies,H.E. and Garnett,M.J. (2018) Cell Model Passports—a hub for clinical, genetic and functional datasets of preclinical cancer models. *Nucleic Acids Res.*, doi: 10.1093/nar/gky872.
63. Jung,S., Lee,T., Cheng,C.-H., Buble,K., Zheng,P., Yu,J., Humann,J., Ficklin,S.P., Gasic,K., Scott,K. *et al.* (2018) 15 years of GDR: New data and functionality in the Genome Database for Rosaceae. *Nucleic Acids Res.*, doi: 10.1093/nar/gky1000.
64. Zheng,Y., Wu,S., Bai,Y., Sun,H., Jiao,C., Guo,S., Zhao,K., Blanca,J., Zhang,Z., Huang,S. *et al.* (2018) Cucurbit Genomics Database (CuGenDB): a central portal for comparative and functional genomics of cucurbit crops. *Nucleic Acids Res.*, doi: 10.1093/nar/gky944.
65. Zeng,X., Zhang,P., Wang,Y., Qin,C., Chen,S., He,W., Tao,L., Tan,Y., Gao,D., Wang,B. *et al.* (2018) CMAUP: a database of collective molecular activities of useful plants. *Nucleic Acids Res.*, doi: 10.1093/nar/gky965.
66. Böhm,M., Bohne-Lang,A., Frank,M., Loss,A., Rojas-Macias,M.A. and Lütteke,T. (2018) Glycosciences.DB: an annotated data collection linking glycomics and proteomics data (2018 update). *Nucleic Acids Res.*, doi: 10.1093/nar/gky994.
67. Bonnardel,F., Mariethoz,J., Salentin,S., Robin,X., Schroeder,M., Perez,S., Lisacek,F. and Imberty,A. (2018) UniLectin3D, a database of carbohydrate binding proteins with curated information on 3D structures and interacting ligands. *Nucleic Acids Res.*, doi: 10.1093/nar/gky832.
68. Ma,J., Chen,T., Wu,S., Yang,C., Bai,M., Shu,K., Li,K., Zhang,G., Jin,Z., He,F. *et al.* (2018) iProX: an integrated proteome resource. *Nucleic Acids Res.*, doi: 10.1093/nar/gky869.
69. Moriya,Y., Kawano,S., Okuda,S., Watanabe,Y., Matsumoto,M., Takami,T., Kobayashi,D., Yamanouchi,Y., Araki,N., Yoshizawa,A.C. *et al.* (2018) The jPOST environment: an integrated proteomics data repository and database. *Nucleic Acids Res.*, doi: 10.1093/nar/gky899.
70. Okuda,S., Watanabe,Y., Moriya,Y., Kawano,S., Yamamoto,T., Matsumoto,M., Takami,T., Kobayashi,D., Araki,N., Yoshizawa,A.C. *et al.* (2017) jPOSTrepo: an international standard data repository for proteomes. *Nucleic Acids Res.*, **45**, D1107–D1111.
71. Phaneuf,P.V., Gosting,D., Palsson,B.O. and Feist,A.M. (2018) ALEdb 1.0: a database of mutations from adaptive laboratory evolution experimentation. *Nucleic Acids Res.*, doi: 10.1093/nar/gky983.