Nucleic Acids Research, 2018, Vol. 46, Database issue ***D1–D7***
doi: 10.1093/nar/gkx1235

The 2018 *Nucleic Acids Research* database issue and the online molecular biology database collection

Daniel J. Rigden[1,*] and Xosé M. Fernández[2]

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

Received November 28, 2017; Editorial Decision November 29, 2017; Accepted November 29, 2017

ABSTRACT

The 2018 *Nucleic Acids Research* Database Issue contains 181 papers spanning molecular biology. Among them, 82 are new and 84 are updates describing resources that appeared in the Issue previously. The remaining 15 cover databases most recently published elsewhere. Databases in the area of nucleic acids include 3DIV for visualisation of data on genome 3D structure and RNArchitecture, a hierarchical classification of RNA families. Protein databases include the established SMART, ELM and MEROPS while GPCRdb and the newcomer STCRDab cover families of biomedical interest. In the area of metabolism, HMDB and Reactome both report new features while PULDB appears in NAR for the first time. This issue also contains reports on genomics resources including Ensembl, the UCSC Genome Browser and ENCODE. Update papers from the IUPHAR/BPS Guide to Pharmacology and DrugBank are highlights of the drug and drug target section while a number of proteomics databases including proteomicsDB are also covered. The entire Database Issue is freely available online on the *Nucleic Acids Research* website (https://academic.oup.com/nar). The NAR online Molecular Biology Database Collection has been updated, reviewing 138 entries, adding 88 new resources and eliminating 47 discontinued URLs, bringing the current total to 1737 databases. It is available at http://www.oxfordjournals.org/nar/database/c/.

NEW AND UPDATED DATABASES

This 2018 *Nucleic Acids Research* Database Issue is the 25th annual collection of bioinformatic databases. The quarter century arrives with 181 papers which, as ever, span all areas of molecular biology research. The total includes 82 new databases (Table 1) and 84 updates of resources that have previously appeared in the Database Issue. There are also 15 updates on databases previously described elsewhere (Table 2).

As in previous years, databases are grouped into eight broad subject categories. These cover (i) nucleic acid sequence and structure, transcriptional regulation; (ii) protein sequence and structure; (iii) metabolic and signalling 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. In an era of increasingly interdisciplinary research, it is no surprise that the content of many databases spans multiple categories so that resources often do not sit comfortably in a single category. Readers are again urged to browse the whole issue, rather than confining themselves to the most obviously relevant sections. The *Nucleic Acids Research* online Molecular Biology Database Collection, which is available at http://www.oxfordjournals.org/nar/database/c/, retains its more finely grained organisation, encompassing 15 categories and 41 subcategories.

The issue begins with broad surveys of resources at major global centres, including 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. The NCBI Resources paper (1) presents an interesting analysis illustrating the extent of the cross-talk between different databases within the site, exemplifying the value to the user of the extensive data integration implemented at these centres. The EBI paper (2) describes new data types including image data, biobanks and biosamples, as well as charting the continued exponential growth in the volume of many kinds of data. The newest of the three, the BIG Data Center (3), focuses on genomic information, but also hosts facilities for samples, program code, and wikis. Many of the wikis are very active and have previously featured in NAR eg lncRNAwiki (4).

The 'Nucleic acid databases' section begins with updates from the International Nucleotide Sequence Database Collaboration (5) and its three contributors, GenBank, ENA and DDBJ (6–8) which together face the challenge of con-

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

Table 1. Descriptions of new online databases in the 2018 NAR Database issue

Database	URL	Brief description[a]
3DIV	http://kobic.kr/3div	3D-genome Interaction Viewer and database
AAgMarker	http://bioinfo.wilmer.jhu.edu/AAgMarker/index.jsp	Serum autoantigen biomarkers from proteome microarrays
aBiofilm	http://bioinfo.imtech.res.in/manojk/abiofilm/	Anti-biofilm compounds
ActiveDriverDB	https://activedriverdb.org/	Genome variation mapped against post-translational modifications
ADReCS-Target	http://bioinf.xmu.edu.cn/ADReCS-Target	Adverse Drug Reactions linked to proteins, genes and genetic variants
AmyPro	http://amypro.net	Proteins with validated amyloidogenic regions
anti-CRISPRdb	http://cefg.uestc.edu.cn/anti-CRISPRdb/	anti-CRISPR proteins
AraGWAS Catalog	https://aragwas.1001genomes.org	Arabidopsis Genome-Wide Association Studies
ASpedia	http://combio.snu.ac.kr/aspedia	Alternative Splicing Encylopedia
ChannelsDB	http://ncbr.muni.cz/ChannelsDB	Channels, pores and tunnels found in biomacromolecular structures
CirGRDB	http://cirgrdb.biols.ac.cn	Regulation of RNAs in circadian rhythms
ClusterCAD	https://clustercad.jbei.org and http://clustercad.igb.uci.edu	Engineering of type I modular polyketide synthases
CR2Cancer	http://cis.hku.hk/CR2Cancer	Chromatin Regulators and Cancer
CSCD	http://gb.whu.edu.cn/CSCD	Cancer-Specific cRNA Database
dbCAN-seq	http://cys.bios.niu.edu/dbCAN_seq	Genome scale CAZymes and CAZyme gene clusters
dbCoRC	http://dbcorc.cam-su.org/	Core transcriptional Regulatory Circuit models
DifferentialNET	http://netbio.bgu.ac.il/diffnet	Differential protein-protein interactions in human tissues
DiseaseEnhancer	http://biocc.hrbmu.edu.cn/DiseaseEnhancer/	Enhancer-disease associations
DISNOR	http://disnor.uniroma2.it/	Protein interaction networks linking disease genes
DreamBASE	http://rna.sysu.edu.cn/dreamBase	Human expressed pseudogenes: DNA Modification, RNA Regulation and bound proteins
ECOdrug	http://www.ecodrug.org	Evolutionary COnservation of Drug targets
EpiDenovo	http://61.148.58.210:8080/EpiDenovo/	The epigenome in mammalian embryonic development
EPD	https://peptracker.com/epd/	Encyclopedia of Protein Dynamics
EVLncRNAs	http://biophy.dzu.edu.cn/EVLncRNAs/	Experimentally Validated lncRNAs including disease indications
eRAM	http://www.unimd.org/eram/	Annotated rare diseases
ExoRBase	http://www.exoRBase.org	Human blood exosome RNAs
FlavorDB	http://cosylab.iiitd.edu.in/flavordb	Flavour molecules
FusionDB	http://services.bromberglab.org/fusiondb/	Functional-repertoire similarity-based organism network
GVM	http://bigd.big.ac.cn/gvm/	Genome Variation Map
HCMDB	http://hcmdb.i-sanger.com/index	Human Cancer Metastasis DataBase
HEDD	http://zdzlab.einstein.yu.edu/1/hedd.php	Human Enhancer Disease Database
ICG	http://icg.big.ac.cn	Internal Control Genes for RT-qPCR normalization
IMOTA	https://ccb-web.cs.uni-saarland.de/imota/	Interactive Multi-Omics-Tissue Atlas
iPTMnet	http://research.bioinformatics.udel.edu/iptmnet/	Post-Translational Modification networks
ITSoneDB	http://itsonedb.cloud.ba.infn.it/	Eukaryotic ribosomal RNA Internal Transcribed Spacer 1 sequences
jMorp	https://jmorp.megabank.tohoku.ac.jp/	Metabolomics and proteomics of 1000 healthy Japanese people
LINCS Data Portal	http://lincsportal.ccs.miami.edu/dcic-portal/	Cell-based perturbation-response signatures
LinkedOmics	http://www.linkedomics.org	Multi-omics analysis of 32 cancers
Lnc2Meth	http://www.bio-bigdata.com/Lnc2Meth	lncRNAs and DNA methylation
m6AVar	http://m6avar.renlab.org/	Human variants affecting m6A sites
MeDReaders	http://medreader.org/	Transcription factors binding methylated DNA
microbiomeDB	http://microbiomeDB.org	Mining and analysing microbiome data
MINTbase	https://cm.jefferson.edu/MINTbase/	Mitochondrial and nuclear tRNA fragments
miRCarta	https://mircarta.cs.uni-saarland.de/	miRNAs and precursors
mirTrans	http://mcube.nju.edu.cn/jwang/lab/soft/mirtrans/	Cell-specific transcriptional information for human miRNAs
MIST	http://fgrtools.hms.harvard.edu/ProteinSearch/	Model organism molecular interaction data
MGA	http://ccg.vital-it.ch/mga/	Mass Genome Annotation
MMP	https://mmp.sfb.uit.no/databases/	Marine Metagenomics Portal
MSDD	http://www.bio-bigdata.com/msdd/	miRNA SNP Disease Database
mSignatureDB	http://tardis.cgu.edu.tw/msignaturedb	Mutational signatures in human cancers
MVP	http://mvp.medgenius.info	Microbe-phage interactions
NPASS	http://bidd2.nus.edu.sg/NPASS/	Natural Product quantitative Activities
OverGeneDB	http://overgenedb.amu.edu.pl	Overlapping protein-coding genes
PAMDB	http://pseudomonas.umaryland.edu	*Pseudomonas aeruginosa* Metabolome DataBase
PancanQTL	http://bioinfo.life.hust.edu.cn/PancanQTL/	Expression quantitative loci (eQTL) analysis of cancer samples
PedAM	http://www.unimd.org/pedam/	Pediatric Disease Annotation & Medicine
PCSD	http://systemsbiology.cau.edu.cn/chromstates	Plant Chromatin State Database
PGG.Population	https://www.pggpopulation.org	Genomic diversity of diverse human populations
PharmacoDB	http://pharmacodb.pmgenomics.ca	Pharmacogenomics of cancer cell lines
PICKLES	http://pickles.hart-lab.org	Pooled In-vitro Crispr Knockout Library Essentiality Screens
PIT-DB	http://pitdb.org	Proteomics Informed by Transcriptomics
Planteome	http://www.planteome.org	Portal for plant ontologies and annotations
PopHuman	http://pophuman.uab.cat	Population genomics-oriented genome browser

Table 1. Continued

Database	URL	Brief description[a]
qPrimerDB	http://biodb.swu.edu.cn/qprimerdb	qPCR primers for 200 organisms
RISE	http://rise.zhanglab.net	RNA-RNA interactions
RNArchitecture	http://iimcb.genesilico.pl/RNArchitecture/	Structural classification of RNAs
SBCCDB	http://sbcddb.moffitt.org	Sleeping Beauty Cancer Driver DataBase
SCPortalen	http://single-cell.clst.riken.jp/	Human and mouse single-cell centric database
SEECancer	http://biocc.hrbmu.edu.cn/SEECancer	Evolutionary-stage specific somatic events in cancer
StemMapper	http://stemmapper.sysbiolab.eu	Stem cell gene expression
STCRDab	http://opig.stats.ox.ac.uk/webapps/stcrdab	Structural T-Cell Receptor Database
SysteMHC Atlas	https://systemhcatlas.org/	Immunopeptidomics of MHC-bound peptides
Tabloid Proteome	http://iomics.ugent.be/tabloidproteome	Protein associations inferred from Mass Spectrometry
Target-Pathogen	http://target.sbg.qb.fcen.uba.ar/patho	Drug target optimisation in pathogens
TC3A	http://tc3a.org/	3' untranslated regions, alternative polyadenylation and cancer
TissGDB	http://zhaobioinfo.org/TissGDB	Tissue-specific Gene DataBase in cancer
TranslatomeDB	http://www.translatomedb.net/	Translatome data from RNC-Seq and Ribo-Seq
TriForC database	http://bioinformatics.psb.ugent.be/triforc/	Triterpene pathways
TCSBN	http://inetmodels.com	Tissue and Cancer-Specific Biological Networks
VarCards	http://varcards.biols.ac.cn/	Interpretation of coding variants in the human genome
VDJdb	https://vdjdb.cdr3.net/.	T-cell receptor sequences with known antigen specificity
Virus Taxonomy	http://ictv.global	Taxonomy of viruses

[a]For 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]
BioMuta and BioExpress	https://hive.biochemistry.gwu.edu/biomuta and https://hive.biochemistry.gwu.edu/bioxpress	Cancer SNVs and gene expression
BioStudies	https://www.ebi.ac.uk/biostudies/	Data of all kinds relating to a single study
iSyTE	http://research.bioinformatics.udel.edu/iSyTE	integrated Systems Tool for Eye gene discovery
miRandola	http://mirandola.iit.cnr.it/	Extracellular and circulating non-coding RNAs
mirDIP	http://ophid.utoronto.ca/mirDIP/	microRNA Data Integration Portal
MNDR	http://www.rna-society.org/mndr/	Mammal ncRNA-Disease Repository
NLSdb	https://rostlab.org/services/nlsdb/	Nuclear Localization Signals
PAGER 2.0	http://discovery.informatics.uab.edu/PAGER/	Pathway, Annotated-list, and Gene-signature Electronic Repository
ProteomicsDB	https://www.ProteomicsDB.org	Mass spectrometry of the human proteome
PULDB	http://www.cazy.org/PULDB_new/	Polysaccharide Utilization Loci in Bacteroidetes species
ReMap	http://remap.cisreg.eu	Transcription factor ChIP-seq data
RMDB	http://rmdb.stanford.edu	Structure mapping of RNA
SuperDrug2	http://cheminfo.charite.de/superdrug2	Approved drugs
TRRUST v2	http://www.grnpedia.org/trrust/	Transcriptional regulatory interactions in human and mouse.
TumorFusions	http://www.tumorfusions.org	Tumour fusion genes

[a]For full references to the databases featured in this issue, please see the Table of Contents.

tinued exponential growth in nucleic acid sequence data. Transcription factors (TF) and transcriptional regulation are represented by a number of databases. The popular returning database of TF binding profiles, JASPAR (9), is published back to back with the ReMAP database (10) of TF ChIP-seq data: data from ReMAP contributed directly to JASPAR's improved coverage. With recent intense interest in the role of 3D chromatin structure in gene regulation, the 3DIV resource (11) for 3D genome interaction visualisation is timely. The key RNA database Rfam (12) contributes an update describing a move to content based on a set of reference genomes. Mirroring changes made in Pfam (13), this eliminates much unhelpful redundancy and allows for clearer taxonomic comparisons.

miRNA biology is strongly represented by updates from established databases such as DIANA-TarBase (14) and mirDIP (15), as well as new databases such as miRCarta (16). The new MSDD (17) links miRNA SNPs to diseases while EVLncRNAs (18) and MNDR (19) also major on disease links for non-coding RNAs. The well-established MODOMICS database of RNA modifications (20) is the subject of an update paper which, among other developments, reports on the availability of liquid chromatography/mass spectrometry data for modified nucleosides, facilitating profiling of such modifications by these methods. RNA structure is covered by the returning RMDB database (21), containing chemical mapping information that can be used to predict RNA secondary and tertiary structure, and the new RNArchitecture (22) which introduces a hierarchical organisation of RNA families with a focus on 3D structures, in the manner popularised by protein databases like SCOP.

In the section on protein sequence and structure databases, the venerable SMART database celebrates 20 years with an update paper (23). It describes a particularly valuable new visualisation option, whereby domain architecture information can be added to phylogenetic trees with the Interactive Tree of Life (iTOL) tool (24). Another up-

date from PDBe (25) includes mention of a newly developed library of freely available web components for interactive data visualizations. One of these, the LiteMol 3D viewer, notably allows convenient display of electron density in the browser window. An update on the popular ELM database of protein sequence motifs (26) reports, among other developments, on how fascinating examples of bacterial pathogen mimicry of eukaryotic motifs are now included in the database. A new arrival, ChannelsDB (27) contributes our cover image and describes the channels, tunnels and pores in protein structures that allow substrate access to buried catalytic sites, for example, or molecular passage through a transmembrane protein. Certain protein classes or families justify their own bespoke databases through medical or biological importance. T-cell receptors are served in this issue by both VDJdb (28), focussing on receptor sequences of known specificity, and STCRDab (29) which collects and curates structural information, linking to and allowing searches against a wide variety of structural, sequence and functional data. The returning database GPCRdb (30), for G protein-coupled receptors, majors on carefully made homology models and mapping receptors to ligands.

Important updates in the metabolic and signalling section include the human metabolomics database HMDB (31). Release 4.0 brings huge increases in content, an improved interface and new kinds of information—predicted mass spectra and pharmacometabolomics. This issue also reports on a new metabolomics database, PAMDB (32), devoted to the bacterial pathogen *Pseudomonas aeruginosa*, justified not only by the biomedical importance of the organism but also by the novel metabolites that it contains. Metabolic pathways are covered by the well-known returning databases Reactome (33) and WikiPathways (34). The former update is notable for its Enhanced High Level Diagrams which superbly contextualise low-level pathways using images of cells, tissues and organs. Among enzyme-oriented databases MEROPS (35), devoted to proteases and their inhibitors, makes a welcome return with a near-doubling of sequences and cross-references to the PANTHER database (36). PANTHER full-length sequence based clustering is shown to be complementary to MEROP's domain-based structure. Carbohydrate-active enzymes are covered by the arrival in NAR of PULDB (37), covering polysaccharide utilization loci in the prominent gut bacteria of the phylum Bacteriodetes, and dbCAN-seq (38), which usefully extrapolates information from the well-known CAZy database (39) to a genome scale. At the enzyme mechanism level, this issue sees the merger of two databases, MACiE and CSA, each veterans of multiple Database Issues, into a single new resource M-CSA (Mechanism and Catalytic Site Atlas) (40).

In the microbial genomics section, there is an update paper from the yeast-focused SGD (41) which now includes curated lists of yeast genes that can replace the functions of human counterparts or vice versa. The popular TADB, covering toxins and antitoxins, also presents an update (42), as does *Subti*Wiki (43), devoted to the biology of *Bacillus subtilis*. Two new databases address viruses. The Virus Taxonomy (44) appears in NAR for the first time, despite the International Committee behind it dating back to the 1960s. The second, MVP (45) describes the complex interactions between microbes and the phage clusters that can infect one or more of them.

Human and model organism genomics are strongly represented. The core resources Ensembl (46) and the UCSC Genome Browser (47) present their usual updates. The former is supplemented by an Ensembl Genomes paper (48) covering non-vertebrates which reports ∼20 000 new genomes covered. Other well-known returning databases include ENCODE (49), RefSeq (50) and Genomicus (51), the last showcasing new karyotype evolutionary trees. Among new databases, current trends in cell and molecular biology are reflected in StemMapper (52) that focusses specifically on stem cell gene expression, and SCPortalen (53) which stores transcriptomics data, metadata and cell images at the single cell level. Another notable new arrival is PICKLES (54) which collects information on human gene essentiality from the results of genome scale CRISPR knockout and shRNA knockdown experiments in cancer and other cell lines.

As ever, databases devoted to human genomic variation and biomedical research are very well represented. Important returning databases include the IUPHAR/BPS Guide to Pharmacology (55) which covers properties of existing and potential drug targets. The authors of the update also describe a major new sister resource, the Guide to Immunopharmacology. An interesting evolutionary perspective on drug targets is provided by ECOdrug (56) which maps the presence or absence of drug target orthologues across species. This will help in efforts to address ecotoxicology concerns over binding of drugs to non-target wild species and assist with appropriate species choices for ecological risk assessments. The popular DrugBank (57) also returns, now in release 5.0 and bringing huge increases in data volume, new data types such as pharmacotranscriptomics and content reporting on the status of clinical trials. A major new resource is the Genome Variation Map (58) from the BIG Data Center covering 19 species. Its arrival is particularly timely with the announcement that comparable NCBI resources dbSNP and dbVar are to stop accepting non-human submissions (https://ncbiinsights.ncbi.nlm.nih.gov/2017/05/09/phasing-out-support-for-non-human-genome-organism-data-in-dbsnp-and-dbvar/).The well-used ClinVar resource (59) also contributes an update and is joined in interpreting human genome variation and its implications for disease by the newcomer VarCards (60). Two interesting new databases, *PGG*.Population (61) and PopHuman (62) present a population genomics perspective of human genome variation, each containing thousands of human genomes from across the world and allowing interactive exploration of and comparison between populations.

Plant databases represented here include the comparative genomics resources PLAZA (63) and Gramene (64). A major new Arabidopsis resource arrives in the form of the AraGWAS catalog (65) which contains hundreds of thousands of links between SNPs and curated phenotypes. In the last section proteomics databases are well-represented. An update is presented on the major quantitative proteomics resource proteomicsDB (66). Its protein-centric view links to an impressive variety of visualisations and to different

kinds of omics data. Future plans include an extension from its current human focus to model organisms. An intuitive user interface is also a strong point of the new EPD database (67), while PIT-DB (68) explicitly works at the intersection of RNA-seq transcriptomics and proteomics mass spectrometry. After covering such a variety of biological areas, it seems appropriate to finish with mention of the BioStudies database (69) that collates data of any and all kinds relating to a single study.

NAR ONLINE MOLECULAR BIOLOGY DATABASE COLLECTION

We reach this year the 25th update of the NAR online Molecular Biology Database Collection (which is freely available at http://www.oxfordjournals.org/nar/database/c/), featuring 88 new databases (Table 1) and 15 databases not described previously in the NAR Database Issue (Table 2). Within our ongoing verification processes to make sure information is still relevant, we have removed 47 obsolete or discontinued databases. After contacting their authors, 138 database entries have been updated with respect to new URLs, new descriptions, and/or other metadata.

We welcome suggestions for inclusion in the Collection of additional databases that have been published in other journals. 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

1. NCBI Resource Coordinators (2017) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, doi:10.1093/nar/gkx1095.
2. Cook,C.E., Bergman,M.T., Cochrane,G., Apweiler,R. and Birney,E. (2017) The European Bioinformatics Institute in 2017: data coordination and integration. *Nucleic Acids Res.*, doi:10.1093/nar/gkx1154.
3. BIG Data Center Members (2017) Database Resources of the BIG Data Center in 2018. *Nucleic Acids Res.*, doi:10.1093/nar/gkx897.
4. Ma,L., Li,A., Zou,D., Xu,X., Xia,L., Yu,J., Bajic,V.B. and Zhang,Z. (2015) LncRNAWiki: harnessing community knowledge in collaborative curation of human long non-coding RNAs. *Nucleic Acids Res.*, **43**, D187–D192.
5. Karsch-Mizrachi,I., Takagi,T., Cochrane,G. and on behalf of the International Nucleotide Sequence Database Collaboration. (2017) The international nucleotide sequence database collaboration. *Nucleic Acids Res.*, doi:10.1093/nar/gkx1097.
6. Benson,D.A., Cavanaugh,M., Clark,K., Karsch-Mizrachi,I., Ostell,J., Pruitt,K.D. and Sayers,E.W. (2017) GenBank. *Nucleic Acids Res.*, doi:10.1093/nar/gkx1094.
7. Silvester,N., Alako,B., Amid,C., Cerdeño-Tarrága,A., Clarke,L., Cleland,I., Harrison,P.W., Jayathilaka,S., Kay,S., Keane,T. *et al.* (2017) The European Nucleotide Archive in 2017. *Nucleic Acids Res.*, **46**, doi:10.1093/nar/gkx1125.
8. Kodama,Y., Mashima,J., Kosuge,T., Kaminuma,E., Ogasawara,O., Okubo,K., Nakamura,Y. and Takagi,T. (2017) DNA Data Bank of Japan: 30th anniversary. *Nucleic Acids Res.*, doi:10.1093/nar/gkx926.
9. Khan,A., Fornes,O., Stigliani,A., Gheorghe,M., Castro-Mondragon,J.A., van der Lee,R., Bessy,A., Chèneby,J., Kulkarni,S.R., Tan,G. *et al.* (2017) JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res.*, doi:10.1093/nar/gkx1126.
10. Chèneby,J., Gheorghe,M., Artufel,M., Mathelier,A. and Ballester,B. (2017) ReMap 2018: an updated atlas of regulatory regions from an integrative analysis of DNA-binding ChIP-seq experiments. *Nucleic Acids Res.*, doi:10.1093/nar/gkx1092.
11. Yang,D., Jang,I., Choi,J., Kim,M.-S., Lee,A.J., Kim,H., Eom,J., Kim,D., Jung,I. and Lee,B. (2017) 3DIV: a 3D-genome Interaction Viewer and database. *Nucleic Acids Res.*, doi:10.1093/nar/gkx1017.
12. Kalvari,I., Argasinska,J., Quinones-Olvera,N., Nawrocki,E.P., Rivas,E., Eddy,S.R., Bateman,A., Finn,R.D. and Petrov,A.I. (2017) Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. *Nucleic Acids Res.*, doi:10.1093/nar/gkx1038.
13. Finn,R.D., Coggill,P., Eberhardt,R.Y., Eddy,S.R., Mistry,J., Mitchell,A.L., Potter,S.C., Punta,M., Qureshi,M., Sangrador-Vegas,A. *et al.* (2016) The pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.*, **44**, D279–D285.
14. Karagkouni,D., Paraskevopoulou,M.D., Chatzopoulos,S., Vlachos,I.S., Tastsoglou,S., Kanellos,I., Papadimitriou,D., Kavakiotis,I., Maniou,S., Skoufos,G. *et al.* (2017) DIANA-TarBase v8: a decade-long collection of experimentally supported miRNA–gene interactions. *Nucleic Acids Res.*, doi:10.1093/nar/gkx1141.
15. Tokar,T., Pastrello,C., Rossos,A.E.M., Abovsky,M., Hauschild,A.-C., Tsay,M., Lu,R. and Jurisica,I. (2017) mirDIP 4.1—integrative database of human microRNA target predictions. *Nucleic Acids Res.*, doi:10.1093/nar/gkx1144.
16. Backes,C., Fehlmann,T., Kern,F., Kehl,T., Lenhof,H.-P., Meese,E. and Keller,A. (2017) miRCarta: a central repository for collecting miRNA candidates. *Nucleic Acids Res.*, doi:10.1093/nar/gkx851.
17. Yue,M., Zhou,D., Zhi,H., Wang,P., Zhang,Y., Gao,Y., Guo,M., Li,X., Wang,Y., Zhang,Y. *et al.* (2017) MSDD: a manually curated database of experimentally supported associations among miRNAs, SNPs and human diseases. *Nucleic Acids Res.*, doi:10.1093/nar/gkx1035.
18. Zhou,B., Zhao,H., Yu,J., Guo,C., Dou,X., Song,F., Hu,G., Cao,Z., Qu,Y., Yang,Y. *et al.* (2017) EVLncRNAs: a manually curated database for long non-coding RNAs validated by low-throughput experiments. *Nucleic Acids Res.*, doi:10.1093/nar/gkx677.
19. Cui,T., Zhang,L., Huang,Y., Yi,Y., Tan,P., Zhao,Y., Hu,Y., Xu,L., Li,E. and Wang,D. (2017) MNDR v2.0: an updated resource of ncRNA–disease associations in mammals. *Nucleic Acids Res.*, doi:10.1093/nar/gkx1025.
20. Boccaletto,P., Machnicka,M.A., Purta,E., Piatkowski,P., Baginski,B., Wirecki,T.K., de Crécy-Lagard,V., Ross,R., Limbach,P.A., Kotter,A. *et al.* (2017) MODOMICS: a database of RNA modification pathways. 2017 update. *Nucleic Acids Res.*, doi:10.1093/nar/gkx1030.
21. Yesselman,J.D., Tian,S., Liu,X., Shi,L., Li,J.B. and Das,R. (2017) Updates to the RNA mapping database (RMDB), version 2. *Nucleic Acids Res.*, doi:10.1093/nar/gkx873.
22. Boccaletto,P., Magnus,M., Almeida,C., Zyla,A., Astha,A., Pluta,R., Baginski,B., Jankowska,E., Dunin-Horkawicz,S., Wirecki,T.K. *et al.* (2017) RNArchitecture: a database and a classification system of RNA families, with a focus on structural information. *Nucleic Acids Res.*, doi:10.1093/nar/gkx966.
23. Letunic,I. and Bork,P. (2017) 20 years of the SMART protein domain annotation. *Nucleic Acids Res.*, doi:10.1093/nar/gkx922.
24. Letunic,I. and Bork,P. (2016) Interactive tree of life (iTOL) v3: An online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.*, **44**, W242–W245.

25. Mir,S., Alhroub,Y., Anyango,S., Armstrong,D.R., Berrisford,J.M., Clark,A.R., Conroy,M.J., Dana,J.M., Deshpande,M., Gupta,D. *et al.* (2017) PDBe: towards reusable data delivery infrastructure at protein data bank in Europe. *Nucleic Acids Res.*, doi:10.1093/nar/gkx1070.
26. Gouw,M., Michael,S., Sámano-Sánchez,H., Kumar,M., Zeke,A., Lang,B., Bely,B., Chemes,L.B., Davey,N.E., Deng,Z. *et al.* (2017) The eukaryotic linear motif resource – 2018 update. *Nucleic Acids Res.*, doi:10.1093/nar/gkx1077.
27. Pravda,L., Sehnal,D., Vareková,R.S., Navrátilová,V., Toušek,D., Berka,K., Otyepka,M. and Koca,J. (2017) ChannelsDB: database of biomacromolecular tunnels and pores. *Nucleic Acids Res.*, doi:10.1093/nar/gkx868.
28. Shugay,M., Bagaev,D.V., Zvyagin,I.V., Vroomans,R.M., Crawford,J.C., Dolton,G., Komech,E.A., Sycheva,A.L., Koneva,A.E., Egorov,E.S. *et al.* (2017) VDJdb: a curated database of T-cell receptor sequences with known antigen specificity. *Nucleic Acids Res.*, doi:10.1093/nar/gkx760.
29. Leem,J., de Oliveira,S.H.P., Krawczyk,K. and Deane,C.M. (2017) STCRDab: the structural T-cell receptor database. *Nucleic Acids Res.*, doi:10.1093/nar/gkx971.
30. Pándy-Szekeres,G., Munk,C., Tsonkov,T.M., Mordalski,S., Harpsøe,K., Hauser,A.S., Bojarski,A.J. and Gloriam,D.E. (2017) GPCRdb in 2018: adding GPCR structure models and ligands. *Nucleic Acids Res.*, doi:10.1093/nar/gkx1109.
31. Wishart,D.S., Feunang,Y.D., Marcu,A., Guo,A.C., Liang,K., Vázquez-Fresno,R., Sajed,T., Johnson,D., Li,C., Karu,N. *et al.* (2017) HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Res.*, doi:10.1093/nar/gkx1089.
32. Huang,W., Brewer,L.K., Jones,J.W., Nguyen,A.T., Marcu,A., Wishart,D.S., Oglesby-Sherrouse,A.G., Kane,M.A. and Wilks,A. (2017) PAMDB: a comprehensive *Pseudomonas aeruginosa* metabolome database. *Nucleic Acids Res.*, doi:10.1093/nar/gkx1061.
33. Fabregat,A., Jupe,S., Matthews,L., Sidiropoulos,K., Gillespie,M., Garapati,P., Haw,R., Jassal,B., Korninger,F., May,B. *et al.* (2017) The Reactome Pathway Knowledgebase. *Nucleic Acids Res.*, doi:10.1093/nar/gkx1132.
34. Slenter,D.N., Kutmon,M., Hanspers,K., Riutta,A., Windsor,J., Nunes,N., Mélius,J., Cirillo,E., Coort,S.L., Digles,D. *et al.* (2017) WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research. *Nucleic Acids Res.*, doi:10.1093/nar/gkx1064.
35. Rawlings,N.D., Barrett,A.J., Thomas,P.D., Huang,X., Bateman,A. and Finn,R.D. (2017) The *MEROPS* database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res.*, doi:10.1093/nar/gkx1134.
36. Mi,H., Poudel,S., Muruganujan,A., Casagrande,J.T. and Thomas,P.D. (2016) PANTHER version 10: Expanded protein families and functions, and analysis tools. *Nucleic Acids Res.*, **44**, D336–D342.
37. Terrapon,N., Lombard,V., Drula,É., Lapébie,P., Al-Masaudi,S., Gilbert,H.J. and Henrissat,B. (2017) PULDB: the expanded database of Polysaccharide Utilization Loci. *Nucleic Acids Res.*, doi:10.1093/nar/gkx1022.
38. Huang,L., Zhang,H., Wu,P., Entwistle,S., Li,X., Yohe,T., Yi,H., Yang,Z and Yin,Y. (2017) dbCAN-seq: a database of carbohydrate-active enzyme (CAZyme) sequence and annotation. *Nucleic Acids Res.*, doi:10.1093/nar/gkx894.
39. Cantarel,B.L., Coutinho,P.M., Rancurel,C., Bernard,T., Lombard,V. and Henrissat,B. (2009) The carbohydrate-active EnZymes database (CAZy): An expert resource for glycogenomics. *Nucleic Acids Res.*, **37**, D233–D238.
40. Ribeiro,A.J.M., Holliday,G.L., Furnham,N., Tyzack,J.D., Ferris,K and Thornton,J.M. (2017) Mechanism and Catalytic Site Atlas (M-CSA): a database of enzyme reaction mechanisms and active sites. *Nucleic Acids Res.*, doi:10.1093/nar/gkx1012.
41. Skrzypek,M.S., Nash,R.S., Wong,E.D., MacPherson,K.A., Hellerstedt,S.T., Engel,S.R., Karra,K., Weng,S., Sheppard,T.K., Binkley,G. *et al.* (2017) *Saccharomyces* genome database informs human biology. *Nucleic Acids Res.*, doi:10.1093/nar/gkx1112.
42. Xie,Y., Wei,Y., Shen,Y., Li,X., Zhou,H., Tai,C., Deng,Z and Ou,H.-Y. (2017) TADB 2.0: an updated database of bacterial type II toxin–antitoxin loci. *Nucleic Acids Res.*, doi:10.1093/nar/gkx1033.
43. Zhu,B. and Stülke,J. (2017) *Subti* Wiki in 2018: from genes and proteins to functional network annotation of the model organism *Bacillus subtilis*. *Nucleic Acids Res.*, doi:10.1093/nar/gkx908.
44. Lefkowitz,E.J., Dempsey,D.M., Hendrickson,R.C., Orton,R.J., Siddell,S.G. and Smith,D.B. (2017) Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV). *Nucleic Acids Res.*, doi:10.1093/nar/gkx932.
45. Gao,N.L., Zhang,C., Zhang,Z., Hu,S., Lercher,M.J., Zhao,X.-M., Bork,P., Liu,Z and Chen,W.-H. (2017) MVP: a microbe–phage interaction database. *Nucleic Acids Res.*, doi:10.1093/nar/gkx1124.
46. Zerbino,D.R., Achuthan,P., Akanni,W., Amode,M.R., Barrell,D., Bhai,J., Billis,K., Cummins,C., Gall,A., Girón,C.G. *et al.* (2017) Ensembl 2018. *Nucleic Acids Res.*, doi:10.1093/nar/gkx1098.
47. Casper,J., Zweig,A.S., Villarreal,C., Tyner,C., Speir,M.L., Rosenbloom,K.R., Raney,B.J., Lee,C.M., Lee,B.T., Karolchik,D., Hinrichs,A.S. *et al.* (2017) The UCSC Genome Browser database: 2018 update. *Nucleic Acids Res.*, doi:10.1093/nar/gkx1020.
48. Kersey,P.J., Allen,J.E., Allot,A., Barba,M., Boddu,S., Bolt,B.J., Carvalho-Silva,D., Christensen,M., Davis,P., Grabmueller,C. *et al.* (2017) Ensembl Genomes 2018: an integrated omics infrastructure for non-vertebrate species. *Nucleic Acids Res.*, doi:10.1093/nar/gkx1011.
49. Davis,C.A., Hitz,B.C., Sloan,C.A., Chan,E.T., Davidson,J.M., Gabdank,I., Hilton,J.A., Jain,K., Baymuradov,U.K., Narayanan,A.K., Onate,K.C. *et al.* (2017) The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res.*, doi:10.1093/nar/gkx1081.
50. Haft,D.H., DiCuccio,M., Badretdin,A., Brover,V., Chetvernin,V., O'Neill,K., Li,W., Chitsaz,F., Derbyshire,M.K., Gonzales,N.R. *et al.* (2017) RefSeq: an update on prokaryotic genome annotation and curation. *Nucleic Acids Res.*, doi:10.1093/nar/gkx1068.
51. Nguyen,N.T.T., Vincens,P., Crollius,H.R. and Louis,A. (2017) Genomicus 2018: karyotype evolutionary trees and on-the-fly synteny computing. *Nucleic Acids Res.*, doi:10.1093/nar/gkx1003.
52. Pinto,J.P., Machado,R.S.R., Magno,R., Oliveira,D.V., Machado,S., Andrade,R.P., Bragança,J., Duarte,I and Futschik,M.E. (2017) StemMapper: a curated gene expression database for stem cell lineage analysis. *Nucleic Acids Res.*, doi:10.1093/nar/gkx921.
53. Abugessaisa,I., Noguchi,S., Böttcher,M., Hasegawa,A., Kouno,T., Kato,S., Tada,Y., Ura,H., Abe,K., Shin,J.W. *et al.* (2017) SCPortalen: human and mouse single-cell centric database. *Nucleic Acids Res.*, doi:10.1093/nar/gkx949.
54. Lenoir,W.F., Lim,T.L. and Hart,T. (2017) PICKLES: the database of pooled *in-vitro* CRISPR knockout library essentiality screens. *Nucleic Acids Res.*, doi:10.1093/nar/gkx993.
55. Harding,S.D., Sharman,J.L., Faccenda,E., Southan,C., Pawson,A.J., Ireland,S., Gray,A.J.G., Bruce,L., Alexander,S.P.H., Anderton,S. *et al.* (2017) The IUPHAR/BPS Guide to PHARMACOLOGY in 2018: updates and expansion to encompass the new guide to IMMUNOPHARMACOLOGY. *Nucleic Acids Res.*, doi:10.1093/nar/gkx1121.
56. Verbruggen,B., Gunnarsson,L., Kristiansson,E., Österlund,T., Owen,S.F., Snape,J.R. and Tyler,C.R. (2017) ECOdrug: a database connecting drugs and conservation of their targets across species. *Nucleic Acids Res.*, doi:10.1093/nar/gkx1024.
57. Wishart,D.S., Feunang,Y.D., Guo,A.C., Lo,E.J., Marcu,A., Grant,J.R., Sajed,T., Johnson,D., Li,C., Sayeeda,Z. *et al.* (2017) DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.*, doi:10.1093/nar/gkx1037.
58. Song,S., Tian,D., Li,C., Tang,B., Dong,L., Xiao,J., Bao,Y., Zhao,W., He,H. and Zhang,Z. (2017) Genome Variation Map: a data repository of genome variations in BIG Data Center. *Nucleic Acids Res.*, doi:10.1093/nar/gkx986.
59. Landrum,M.J., Lee,J.M., Benson,M., Brown,G.R., Chao,C., Chitipiralla,S., Gu,B., Hart,J., Hoffman,D., Jang,W. *et al.* (2017) ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.*, doi:10.1093/nar/gkx1153.
60. Li,J., Shi,L., Zhang,K., Zhang,Y., Hu,S., Zhao,T., Teng,H., Li,X., Jiang,Y., Ji,L. *et al.* (2017) VarCards: an integrated genetic and clinical database for coding variants in the human genome. *Nucleic Acids Res.*, doi:10.1093/nar/gkx1039.
61. Zhang,C., Gao,Y., Liu,J., Xue,Z., Lu,Y., Deng,L., Tian,L., Feng,Q. and Xu,S. (2017) *PGG*.Population: a database for understanding the genomic diversity and genetic ancestry of human populations. *Nucleic Acids Res.*, doi:10.1093/nar/gkx1032.

62. Casillas,S., Mulet,R., Villegas-Mirón,P., Hervas,S., Sanz,E., Velasco,D., Bertranpetit,J., Laayouni,H. and Barbadilla,A. (2017) PopHuman: the human population genomics browser. *Nucleic Acids Res.*, doi:10.1093/nar/gkx943.
63. Van Bel,M., Diels,T., Vancaester,E., Kreft,L., Botzki,A., Van de Peer,Y., Coppens,F and Vandepoele,K. (2017) PLAZA 4.0: an integrative resource for functional, evolutionary and comparative plant genomics. *Nucleic Acids Res.*, doi:10.1093/nar/gkx1002.
64. Tello-Ruiz,M.K., Naithani,S., Stein,J.C., Gupta,P., Campbell,M., Olson,A., Wei,S., Preece,J., Geniza,M.J., Jiao,Y. *et al.* (2017) Gramene 2018: unifying comparative genomics and pathway resources for plant research. *Nucleic Acids Res.*, doi:10.1093/nar/gkx1111.
65. Togninalli,M., Seren,Ü., Meng,D., Fitz,J., Nordborg,M., Weigel,D., Borgwardt,K., Korte,A and Grimm,D.G. (2017) The AraGWAS Catalog: a curated and standardized *Arabidopsis thaliana* GWAS catalog. *Nucleic Acids Res.*, doi:10.1093/nar/gkx954.
66. Schmidt,T., Samaras,P., Frejno,M., Gessulat,S., Barnert,M., Kienegger,H., Krcmar,H., Schlegl,J., Ehrlich,H.-C., Aiche,S. *et al.* (2017) ProteomicsDB. *Nucleic Acids Res.*, doi:10.1093/nar/gkx1029.
67. Brenes,A., Afzal,V., Kent,R and Lamond,A.I. (2017) The Encyclopedia of Proteome Dynamics: a big data ecosystem for (prote)omics. *Nucleic Acids Res.*, doi:10.1093/nar/gkx807.
68. Saha,S., Chatzimichali,E.A., Matthews,D.A. and Bessant,C. (2017) PITDB: a database of translated genomic elements. *Nucleic Acids Res.*, doi:10.1093/nar/gkx906.
69. Sarkans,U., Gostev,M., Athar,A., Behrangi,E., Melnichuk,O., Ali,A., Minguet,J., Rada,J.C., Snow,C., Tikhonov,A. *et al.* (2017) The BioStudies database—one stop shop for all data supporting a life sciences study. *Nucleic Acids Res.*, doi:10.1093/nar/gkx965.