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

Daniel J. Rigden1,* 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 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 EBI 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 IneRNAwiki (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-
| Database                | URL                                         | Brief description                                                                 |
|------------------------|---------------------------------------------|-----------------------------------------------------------------------------------|
| 3DIV                   | http://kobic.kr/3div                        | 3D-genome Interaction Viewer and database                                         |
| AAgMarker              | http://bioinf.wilmer.jhu.edu/AAgMarker/index.jsp | Serum autoantigen biomarkers from proteome microarrays                            |
| aBiofilm               | http://bioinfo.imtech.res.in/manojk/abiofilm/ | Anti-biofilm compounds                                                            |
| ActiveReaderDB         | https://activereaderdb.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://cegf.ueste.ac.cn/anti-CRISPRdb/      | anti-CRISPR proteins                                                              |
| AraGWASt Catalog       | https://aragwas.1001genomes.org             | Arabidopsis Genome-Wide Association Studies                                       |
| ASpedia                | http://combio.snu.ac.kr/aspedia             | Alternative Splicing Encyclopedia                                                  |
| ChannelsDB             | http://ncbr.muni.cz/ChannelsDB              | Channels, pores and tunnels found in biomacromolecular structures                 |
| CirGRDB                | http://cigrdb.biols.ac.cn                  | Regulation of RNAs in circadian rhythms                                           |
| ClusterCAD             | https://clustercad.jbei.org and http://clustercad.igb.uc.edu | Engineering of type I modular polyketide synthases                                |
| CR2Cancer              | http://cis.hku.hk/CR2Cancer                 | Chromatin Regulators and Cancer                                                    |
| CSD                    | http://gb.dtu.edu.cn/CSD                    | Cancer-Specific cRNA Database                                                     |
| dbCAN-seq              | http://cys.biox.nwu.edu/dbCAN_seq           | Genome-wide CAZyme and CAZyme gene clusters                                       |
| dbCoRC                 | http://dbcorc.cam-su.org/                  | Core transcriptional Regulatory Circuit models                                     |
| DifferentialNET        | http://netbio.bgu.ac.il/difnet              | Differential protein-protein interactions in human tissues                        |
| DiseaseEnhancer        | http://biocb.hrbmu.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 CoSurvival of Drug targets                                           |
| EpiDenovo              | http://46.148.58.210:8080/EpiDenovo/        | The epigenome in mammalian embryonic development                                  |
| EPD                    | https://peptracker.com/epd/                | Encyclopedia of Protein Dynamics                                                    |
| EVLncRNAs              | http://evl.lncrnas.org                      | Experimentally Validated lncRNAs including disease indications                    |
| eRAM                   | http://www.unimid.org/eram/                | Annotated rare diseases                                                            |
| ExorBase               | http://www.exorb.org                       | Human blood exosome RNAs                                                          |
| FlavorDB               | http://cosylab.iitd.iitd.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://zdlzlab.einstein.yu.edu/1/hedd.php   | Human Enhancer Disease Database                                                    |
| ICG                    | http://icg_big.ac.cn                       | Internal Control Genes for RT-qPCR normalization                                   |
| iMOTA                  | http://ccb-web.cs.uni-saarland.de/imota/    | Interactive Multi-Omics-Tissue Atlas                                              |
| iPTMNet                | http://research.bioinformatics.udeu.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://linportal.ccs.miami.edu/deic-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               | http://mcm.jefferson.edu/MINTbase/          | Mitochondrial and nuclear tRNA fragments                                           |
| mirCarta               | https://mircarta.cs.uni-saarland.de/        | miRNAs and precursors                                                             |
| mirTrans               | http://mctue.nju.edu.cn/wang/lab/mirtans/   | Cell-specific transcriptional information for human miRNAs                        |
| MIST                   | http://fgtoolkit.hms.harvard.edu/ProteinSearch/ | Model organism molecular interaction data                                         |
| MGA                    | http://ecg.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 for human miRNAs                                              |
| MVP                    | http://mvp.medgenius.info/                 | Microbe-phage interactions                                                         |
| NPASS                  | http://bibd2.nus.edu.sg/NPASS/             | Natural Product Quantitative Activities                                            |
| OverGeneDB             | http://overgeneb.amu.edu.pl                 | Overlapping protein-coding genes                                                  |
| PAMDB                  | http://pseudomonas.umaryland.edu           | *Pseudomonas aeruginosa* Metabolve Database                                        |
| PancanQTL              | http://bioinfo.life.hust.edu.cn/PancanQTL/ | Expression quantitative loci (eQTL) analysis of cancer samples                     |
| PeLAM                  | http://www.unimdb.org/PeLAM/              | Pediatric Disease Annotation & Medicine                                            |
| PCSD                   | http://systembiology.cau.edu.cn/chromstates | Plant Chromatin State Database                                                    |
| PGG.Population          | https://www.pggpopulation.org              | Genomic diversity of diverse human populations                                     |
| PharmacodB             | http://pharmacodb.pgenomics.ca             | Pharmacogenomics of cancer cell lines                                              |
| PICKLES                | http://pickles.hart-lab.org/               | Pooled In-vitro Crispr Knockout Library Essentiality Screens                      |
| PIT-DB                 | http://piddb.org                           | Proteomics Informed by Transcriptomics                                             |
| Plantome               | http://www.plantome.org/                   | Portal for plant ontologies and annotations                                       |
| PopHuman               | http://pophuman.uab.cat/                   | Population genomics-oriented genome browser                                       |
Table 1. Continued

| Database             | URL                        | Brief description |
|----------------------|----------------------------|-------------------|
| q PrimerDB           | http://biodb.swu.edu.cn/qprimerdb | qPCR primers for 200 organisms |
| RISE                 | http://rise.zhanglab.net   | RNA-RNA interactions |
| RNA Architecture     | http://lumich.genesilico.pl/RNAArchitecture/ | Structural classification of RNAs |
| SBC/CDBDB            | http://sbcdbd.moliflt.org  | Sleeping Beauty Cancer Driver Database |
| SC Portalen          | http://single-cell.cst.riken.jp/ | Human and mouse single-cell centric database |
| SEE Cancer           | http://bioce.hrbmu.edu.cn/ | Evolutionary-stage specific somatic events in cancer |
| Stem Mapper          | http://stemmapper sybioolah.eu | Stem cell gene expression |
| STCR Dab             | http://opig.stats.os.ac.uk/webapp/stcrdab | Structural T-Cell Receptor Database |
| SysteMHC Atlas       | https://systemmhc atlas.org/ | Immunoproteomics of MHC-bound peptides |
| Tabloid Proteome     | http://ioformatics.ugent.be/tabloidproteome | Protein associations inferred from Mass Spectrometry |
| Target Pathogen      | http://target.sbg.qb.tueen.uba.ar/patho | Drug target optimisation in pathogens |
| TC3A                 | http://tc3a.org/           | 3'untranslated regions, alternative polyadenylation and cancer |
| Tiss GDB             | http://zhao bioinfo.org/TissGDB | Tissue-specific Gene Database in cancer |
| Translatome DB       | http://www.translatomdb.net/ | Translome data from RNC-Seq and Ribo-Seq |
| Tri ForC database    | http://bioinformatics.ugent.be/triforc/ | Tripterpe pathways |
| TC SNB               | http://inet models.com     | Tissue and Cancer-Specific Biological Networks |
| Var Cards            | http://var cards biols ac cn/ | Interpretation of coding variants in the human genome |
| VDJ db               | http://vdj db dr3 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 |
|-------------------|----------------------------|-------------------|
| BioMuta and BioExpress | http://hive.biochemistry.gwu.edu/biomuta and bioexpress | 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.uelu.de/iSyTE | integrated Systems Tool for Eye gene discovery |
| mi Randola        | http://mirandola.iit.cnri.it/ | Extracellular and circulating non-coding RNAs |
| mir DIP           | http://ophid.utoronto.ca/mirDIP/ | microRNA Data Integration Portal |
| MNDR              | http://www.rna-society.org/mndr/ | Mammal ncRNA-Disease Repository |
| NL Sdb            | https://rostlab.org/services/nlsdb/ | Nuclear Localization Signals |
| PAGER 2.0         | http://discovery.informatics.ub.edu/PAGER/ | Pathway, Annotated-list, and Gene-signature Electronic Repository |
| Pro teomics DB    | https://www.ProteomicsDB.org | Mass spectrometry of the human proteome |
| PUL DB            | 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 |
| TRUST v2          | http://www.grnpedia.org/trust/ | Transcriptional regulatory interactions in human and mouse. |
| Tumor Fusions     | http://www.tumorfusions.org | Tumour fusion genes |

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

continued 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 Pfam (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 miR-Carta (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 RNA Architecture (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 STCRDb (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 SubtiWiki (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 Waellken 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., Apevi,G., Apevi,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/gkx1079.
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., Kubo,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., Fores,O., Stigliani,A., Gheorghe,M., Castro-Mondragon,J.A., van der Lee,R., Bessy,A., Chénèby,J., Kulkarni,S.R. and 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. Chénèby,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/gkx1112.
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) Pfam 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., Tatsoglu,S., Papadimitriou,D., Lagarde,V., de Crécy-Lagard,V., Ross,R., Limbach,P.A., Kotter,A. et al. (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/gkx1112.
15. Tokar,T., Pastrello,C., Rossos,A.E.M., Abosky,M., Haushild,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/gkx1038.
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., Qi,Y., Yang,Y. et al. (2017) EVLncRNA: a manually curated database for long non-coding RNAs validated by low-throughput experiments. Nucleic Acids Res., doi:10.1093/nar/gkx677.
19. Cui,Y., 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., Purna,E., Piatkowski,P., Baginski,B., Wiericki,T.K., de Crécy-Lagard,V., Ross,R., Limbach,P.A., Kettler,A. et al. (2017) MODOMICS: a database of RNA modification pathways. 2017 update. Nucleic Acids Res., doi:10.1093/nar/gkx1030.
21. Yesselmant,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,N., Almeida,C., Zyla,A., Astha,A., Pluta,R., Baginski,B., Jankowska,E., Dunin-Horkawicz,S., Wiericki,T.K. et al. (2017) RNAArchitecture: a database and a classification system of RNA families, with a focus on structural information. Nucleic Acids Res., doi:10.1093/nar/gkx926.
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. Xie, Y., Alhroub, Y., Anyango, S., Armstrong, D.R., Berrisford, J.M., Clark, A.R., Conroy, M.J., Dana, D.M., Deshpande, M., Gupta, D. et al. (2017) PDBc: towards reusable data delivery infrastructure at protein data bank in Europe. *Nucleic Acids Res.*, doi:10.1093/nar/gks1070.

26. Gouw, M., Michael, S., Sámano-Sánchez, H., Kumar, M., Zeka, 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., Snehá, D., Varekó, R.S., Navrátilová, T., Vouše, D., Berka, K., Otýpek, 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) VDldb: 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) STCRDb: the structural T-cell receptor database. *Nucleic Acids Res.*, doi:10.1093/nar/gkw397.

30. Pándy-Szekeres, G., Munk, C., Tsonkov, T.M., Mordalski, S., Harpoe, K., Hauser, A.S., Bojariski, A.J. and Glorian, 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., Huang, X., Bateman, A., Wishart, D.S., Barrett, A.J., Thomas, P.D., Krawczyk, K. and Deane, C.M. (2017) TADB 2.0: an updated database of bacterial type II biomacromolecular tunnels and pores. *Nucleic Acids Res.*, doi:10.1093/nar/gkx1093.

33. Slenter, D.N., Kutmon, M., Hanspers, K., Riutta, A., Windsor, J., Munk, C., Tsonkov, T.M., Mordalski, S., Singh, E., Andrade, R.P., Braganc¸a, 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/gkw921.

34. Zhu, B. and Stülke, J. (2017) *Subi* 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.

35. 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/gkx1039.

36. 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.

37. Zerbo, D.R., Achutan, P., Akkani, W., Amode, M.R., Barrett, D., Bhai, J., Bills, K., Cummins, C., Gall, A., Giron, C.G. et al. (2017) PANTHER version 10: Expanded protein evidence. *Nucleic Acids Res.*, doi:10.1093/nar/gkx1099.

38. Casper, J., Zweig, A.S., Villarreal, C., Tyner, C., Speir, M.L., Rosenblom, K.R., Raney, B.J., Lee, C.M., Lee, B.T., Karolichik, D., Hinrichs, A.S. et al. (2017) The UCSC Genome Browser database: 2018 update. *Nucleic Acids Res.*, doi:10.1093/nar/gkx1020.

39. Kersey, P.J., Allen, E., Allot, A., Barba, M., Boddu, S., Bolt, B.J., Carvalho-Silva, D., Christensen, M., Davis, P., Grabmuller, C. et al. (2017) Ensembl Genomes 2018: an integrated omics infrastructure for non-vertebrate species. *Nucleic Acids Res.*, doi:10.1093/nar/gkx1011.

40. Davis, C.A., Hitz, B.C., Sloan, C.A., Chan, E.T., Davidson, J.M., Estades, D., P´andy-Szekeres, G., Munk, C., Tsonkov, T.M., Mordalski, S., Singh, E., Andrade, R.P., Braganc¸a, J., Duarte, I. and Futschik, M.E. (2017) WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research. *Nucleic Acids Res.*, doi:10.1093/nar/gkw1064.

41. 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 peptides in the PANTHER database. *Nucleic Acids Res.*, doi:10.1093/nar/gkw1134.

42. Slentor, D.N., Kutmon, M., Hanspers, K., Riutta, A., Windsor, J., Nunes, N., Melius, 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.

43. Ribeiro, A.J.M., Holliday, G.L., Furnham, N., Tyack, J.D., Ferriss, 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/gkx984.

44. Skrzypek, M.S., Nash, R.S., Wong, E.D., MacPherson, K.A., Hellerstedt, S.T., Engel, S.R., Kuznetsov, V., Sengupta, S., Steppuhn, T.K., Binkley, G. et al. (2017) Saccharomyces genome database informs human biology. *Nucleic Acids Res.*, doi:10.1093/nar/gkx1112.

45. 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.
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., Krmar, 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 (pro)teomics. *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., Minguez, 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.