

The 24th annual *Nucleic Acids Research* database issue: a look back and upcoming changes

Michael Y. Galperin^{1,*}, Xosé M. Fernández-Suárez² and Daniel J. Rigden^{3,*}

¹National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA, ²Thermo Fisher Scientific, Inchinnan Business Park, Paisley, Renfrew PA4 9RF, UK and ³Institute of Integrative Biology, University of Liverpool, Crown Street, Liverpool L69 7ZB, UK

Received November 11, 2016; Editorial Decision November 15, 2016; Accepted November 16, 2016

ABSTRACT

This year's Database Issue of *Nucleic Acids Research* contains 152 papers that include descriptions of 54 new databases and update papers on 98 databases, of which 16 have not been previously featured in *NAR*. As always, these databases cover a broad range of molecular biology subjects, including genome structure, gene expression and its regulation, proteins, protein domains, and protein–protein interactions. Following the recent trend, an increasing number of new and established databases deal with the issues of human health, from cancer-causing mutations to drugs and drug targets. In accordance with this trend, three recently compiled databases that have been selected by *NAR* reviewers and editors as 'breakthrough' contributions, *DENOVO-DB*, the Monarch Initiative, and Open Targets, cover human *de novo* gene variants, disease-related phenotypes in model organisms, and a bioinformatics platform for therapeutic target identification and validation, respectively. We expect these databases to attract the attention of numerous researchers working in various areas of genetics and genomics. Looking back at the past 12 years, we present here the 'golden set' of databases that have consistently served as authoritative, comprehensive, and convenient data resources widely used by the entire community and offer some lessons on what makes a successful database. The Database Issue is freely available online at the <https://academic.oup.com/nar> web site. An updated version of the *NAR* Molecular Biology Database Collection is available at <http://www.oxfordjournals.org/nar/database/a/>.

NEW AND UPDATED DATABASES

The current 2017 *Nucleic Acids Research* Database Issue is the 24th annual collection of bioinformatic databases on various areas of molecular biology. It includes 152 papers, of which 54 describe newly created databases (Table 1), 82 papers provide updates on the databases that have been previously described in *NAR* and 16 contain updates on the databases whose descriptions have previously been published elsewhere (Table 2).

As previously, the issue is organized according to subject categories covering (i) nucleic acid sequence and structure, transcriptional regulation; (ii) protein sequence and structure; (iii) metabolic and signaling pathways, protein–protein interactions; (iv) genomics of viruses, bacteria, protozoa and fungi; (v) genomics of human and model organisms; (vi) human diseases and drugs; (vii) plants and (viii) other topics, such as proteomics databases. Unsurprisingly, many resources straddle multiple categories and defy easy classification so we encourage readers to browse the whole issue, not limiting themselves to a single section. The databases listed in the *Nucleic Acids Research* online Molecular Biology Database Collection, which is available at <http://www.oxfordjournals.org/nar/database/a/>, are split into the same 15 categories and 41 subcategories as before.

In this year's issue, the usual annual survey of the progress in databases held by the U.S. National Center for Biotechnology Information (NCBI), is supplemented by a report from the Beijing Institute of Genomics, Chinese Academy of Sciences, on their BIG Data Center which hosts a variety of genomic databases. [Because of the high number of references to the databases in the *NAR* 'golden set' (Table 3), we could not properly cite most of the papers included in the current Database Issue. Please refer to this issue's Table of Contents.]

In the 'Nucleic acid databases' section, several resources emphasize the complexity of regulatory processes. Examples include SNP2TFBS, a database of SNPs in predicted transcription factor binding sites (TFBSs); LincSNP, a database that links SNPs to long noncoding RNAs and

*To whom correspondence should be addressed. Email: nardatabase@gmail.com
Correspondence may also be addressed to Daniel J. Rigden. Email: drigden@liv.ac.uk

Table 1. Descriptions of new online databases in the 2017 *NAR* Database issue

Database name	URL	Brief description ^a
3DSNP	http://biotech.bmi.ac.cn/3dsnp/	Human noncoding SNPs: interactions with genes and other SNPs
AAtlas	http://aagatlas.ncpsb.org	Human AutoAntigen database
ADPriboDB	http://adpribodb.leunglab.org/	ADP-ribosylated proteins and sites
antiSMASH	http://antismash-db.secondarymetabolites.org	antibiotics and Secondary Metabolite Analysis SHell
AraPheno	https://arapheno.1001genomes.org	Phenotypic data for <i>Arabidopsis thaliana</i>
ccNET	http://structuralbiology.cau.edu.cn/gossypium/	Co-expression networks for diploid and polyploid <i>Gossypium</i>
CeNDR	http://www.elegansvariation.org	<i>C. elegans</i> Natural Diversity Resource
CGDB	http://cgdb.biocuckoo.org/	Circadian Gene database
CistromeDB	http://cistrome.org/db	ChIP-Seq and DNase-Seq data in human and mouse
Coexpedia	http://www.coexpedia.org	Gene co-expression data mapped to medical subject headings (MeSH).
dbSAP	http://www.megabionet.org/dbSAP	Single Amino acid Polymorphisms: SNP-derived variation in human proteins
denovo-db	http://denovo-db.gs.washington.edu	Human <i>de novo</i> gene variants detected by parent-child sequencing
DrugCentral	http://drugcentral.org	Active ingredients of approved pharmaceutical products, indications and mode of action
EURISCO	http://eurisco.ecpgr.org/	European catalogue for plant genetic resources
ExAC browser	http://exac.broadinstitute.org	Exome Aggregation Consortium sequence data
Exposome-Explorer	http://exposome-explorer.iarc.fr	Biomarkers of exposure to disease risk factors
FAIRDOMHub	https://fairdomhub.org/	Findable, Accessible, Interoperable and Reusable Data, Operating procedures and Models
FuzDB	http://protodyn-database.org	Database of fuzzy protein complexes
GenomeCRISPR	http://genomecrispr.org	High-throughput screening using the CRISPR/Cas-9 system
GTRD	http://gtrd.biouml.org	Gene Transcription Regulation Database
HieranoiDB	http://hieranoiDB.sbc.su.se/	Ortholog groups and trees inferred by Hieranoi2 software
IGSR	http://www.1000genomes.org/data-portal	International Genome Sample Resource
IMG/VR	https://img.jgi.doe.gov/vr/	DOE Joint Genome Institute Viral Resource
JET2 Viewer	http://www.lcqb.upmc.fr/jet2-viewer/	Joint Evolutionary Trees: protein-protein interaction patches in known structures
jPOSTrepo	https://repository.jpstodb.org/	Japanese Proteome Standard repository
KERIS	http://igenomed.org/KERIS	Kaleidoscope of gEne Responses to Inflammation among Species
LinkProt	http://linkprot.cent.uw.edu.pl/	Topologically complex protein structures
LNCEditing	http://bioinfo.life.hust.edu.cn/LNCEditing/	RNA editing sites in lncRNAs from human, monkey, mouse and fly
MEGaRes	https://meg.colostate.edu/MEGaRes/	Mechanisms of antimicrobial resistance
Membranome	http://membranome.org/	A database of single-pass membrane proteins
MethSMRT	http://sysbio.susu.edu.cn/methsmrt	DNA methylation data from Single Molecule, Real-Time sequencing
mirDNMR	https://www.wzgenomics.cn/mirdnrm/	Background <i>de novo</i> mutation rates in human genes
Monarch Initiative	http://monarchinitiative.org	Human disease-related genotypes and phenotypes in model organisms
MRPrimerV	http://infolab.dgist.ac.kr/MRPrimerV	PCR primer pairs for detecting RNA virus-mediated infectious diseases
mutLBSgeneDB	http://www.zhaobioinfo.org/mutLBSgeneDB/	Mutations in Ligand Binding Sites gene DataBase
NSDNA	http://www.bio-bigdata.net/nsdna/	Nervous System Disease ncRNA Atlas
Ontobee	http://www.ontobee.org/	Ontology database server of OBO Foundry
Open Targets	https://targetvalidation.org	Target validation platform: links between potential drug targets and diseases
pathDIP	http://ophid.utoronto.ca/pathDIP	Pathway data integration and analysis portal
PathoYeasttract	http://pathoyeasttract.org/index.php	Transcription regulation in pathogenic yeasts
PceRBase	http://bis.zju.edu.cn/pcernadb/index.jsp	Plant competing endogenous RNAs
Pharos	https://pharos.nih.gov/idg/index	Data on unstudied and understudied drug targets
PLaMoM	http://www.byanbioinfo.org/plamom/	Plant Mobile Macromolecules: Extracellular siRNAs, microRNAs, mRNAs and proteins in plants
Plant Reactome	http://plantreactome.gramene.org/	Plant metabolic, regulatory and signaling pathways
PMDBase	http://www.sesame-bioinfo.org/PMDBase	Plant microsatellites and marker development
POSTAR	http://POSTAR.ncrnalab.org	Post-transcriptional regulation by RNA-binding proteins
proGenomes	http://van.embl.de/progene/	Consistently annotated bacterial and archaeal genomes
Proteome-pI	http://isoelectricpointdb.org/	Pre-computed isoelectric points for > 5000 proteomes
REDIportal	http://srv00.recas.ba.infn.it/atlas/	A-to-I RNA editing events in human
RNAlocate	http://www.rna-society.org/rnalocate/	RNA localization in the cell
SNP2TFBS	http://ceg.vital-it.ch/snp2tfbs/	Regulatory SNPs affecting predicted transcription factor binding sites
SoyNet	http://www.inetbio.org/soynet/	Co-functional networks for soy bean <i>Glycine max</i>
TFBSbank	http://tfbsbank.co.uk/	Transcription Factor Binding Site profiles deduced from ChIP-seq or ChIP-chip data
TSTMP	http://tstmp.enzim.ttk.mta.hu	Target Selection database for human TransMembrane Proteins
Uniclust	http://uniclust.mmseqs.com/	Clustered protein sequences and multiple sequence alignments
WERAM	http://weram.biocuckoo.org/	Writers, Erasers and Readers of histone Acetylation and Methylation

^aAt the time of this writing, references to the databases featured in this issue have not yet been finalized; please see the Database Issue Table of Contents.

their TFBSs; LNCEditing, a database of RNA editing in lncRNAs, and POSTAR, a resource on post-transcriptional regulation by RNA-binding proteins.

Major protein sequence databases include updates from UniProt and InterPro, the latter encompassing ever more component databases, most recently the NCBI's Conserved Domain Database (CDD), which is also described in a separate paper in this issue, and the Structure–Function Linkage Database (SFLD), which has been featured in *NAR* previously (1). Accordingly, as described in the UniProt paper, InterPro now serves as a major source of protein functional

annotation for the UniProt entries. Updates on primary protein structure databases include papers on the RCSB Protein Data Bank (PDB) and PDBj. The latter reports on the integration of previously separate visualizations, allowing a single tool to display macromolecular structures not just from the PDB, but also from EMDB and SAS-BDB, containing structural information obtained, respectively, from cryo-EM and small angle solution scattering experiments. PDBj also now allows a search across the same three databases on shape similarity. In the area of modelled structures, the hugely popular Swiss-Model repository re-

Table 2. Updated descriptions of databases most recently published elsewhere

Database	URL	Brief description ^a
CARD	http://arpcard.mcmaster.ca	Comprehensive Antibiotic Research Database
dbDEMOC	http://www.picb.ac.cn/dbDEMOC	Differentially expressed miRNAs in human cancers
DisGeNET	http://www.disgenet.org/	Genetic determinants of human diseases
ECOD	http://proddata.swmed.edu/ecod/	Evolutionary Classification Of protein Domains
GETPrime	http://bbcfertools.epfl.ch/getprime	Gene- or transcript-specific primers for qPCR
HIPPIE	http://cbdm.uni-mainz.de/hippie/	Human Integrated Protein-Protein Interaction rEference
HipSci	http://www.hipsci.org/	Human induced pluripotent Stem cells initiative
IMG-ABC	https://img.jgi.doe.gov/abc-public/	Integrated Microbial Genomes—Atlas of Biosynthetic gene Clusters
Influenza Research Database	http://www.fludb.org	All data on influenza: sequences, strains, alignments, trees, variation, epitopes, classification and surveillance
LincSNP	http://bioinfo.hrbmu.edu.cn/LincSNP	Association of human lncRNAs with disease-related SNPs
MalaCards	http://www.malacards.org/	Human maladies and their annotations, organized into ‘disease cards’
pVOGs	http://research.engineering.uiowa.edu/kristensenlab/VOG	Prokaryotic Virus Orthologous Groups of proteins
Proteome Xchange	http://www.proteomexchange.org/	Proteomics resources portal
RAID	http://www.rna-society.org/raid	Human RNA-RNA and RNA-protein interactions
SZGR	https://bioinfo.uth.edu/SZGR/	SchiZophrenia Gene Resource
WDCM	http://www.wdcm.org	World Data Center of Microorganisms collections
XTalkDB	http://www.xtalkdb.org	Crosstalk among signaling pathways

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

ports new features and policies, including a weekly update of modelled proteomes of 12 ‘core species’ to recognize the possible emergence of better templates in weekly PDB releases. Reacting quickly to the ever-expanding PDB is also a preoccupation of two protein structural domain databases, CATH and ECOD, reporting in update papers here. The CATH paper reports a new, daily-generated automatic supplement CATH-B, as well as developments of its functional families, or FunFams, whose value in sequence annotation has become clear in competitive blind CAFA tests. The ECOD describes a weekly release cycle as well as new ways to search the database and convenient means to superimpose and visualize the search results. Class-specific protein databases include updates from RepeatsDB and DisProt and an interesting new arrival FuzDB, cataloguing protein complexes whose components remain ‘fuzzy’, or locally disordered, even when interacting with other proteins. Another new database, LinkProt, features protein structures with topologically complex shapes.

Metabolic and signaling pathway databases include updates on major resources KEGG and BioGRID. The update from BRENDA database of enzymes, one of the most venerable in the collection, dating as it does from 1987, describes new means of visualization - pathway maps and metabolic overviews. An interesting new arrival, XTalkDB, focuses specifically on cross-talk between signaling pathways. Microbe-related databases in the following section include heavily-used resources for influenza (Influenza Research Database) and *Escherichia coli* (EcoCyc). Other databases focus significantly on pathogens, or on antimicrobial resistance. Eukaryote pathogens are strongly represented by EuPathDB, PHI-base and a newcomer PathoYeast, focusing on transcription regulation in pathogenic yeasts.

In the section covering genomics and comparative genomics, important updates from Ensembl, FlyBase, STRING and the UCSC Genome Browser are included. Easy access to orthologous genes across species is provided by the well-established OrthoDB and the new arrival HieranoiDB which offers beautifully presented trees of orthologues. Another important cross-species analysis is represented by the Monarch Initiative, highlighted by *NAR* re-

viewers and editors as a ‘Breakthrough’ article (2). Working with the Human Phenotype Ontology, also reporting an update in this issue, Monarch Initiative aims to link mutations in orthologous genes to the similar phenotypes often observed in different species. This ambitious objective requires the careful use and integration of ontologies to precisely describe anatomy, diseases and phenotypes, but the pay-off is an ability to link from human diseases to disease models in various model organisms, maximizing the value of data obtained for any given organism (2).

As ever, this issue covers important databases supporting research in the molecular basis of disease and treatment. Cancer is covered not only by the major resource COSMIC, reporting interesting new coverage of the genetics of drug resistance, but also by updates to ChimerDB, recording chimeric transcripts, dbDEMOC, containing information on miRNA expression levels in cancer, and YM500, focusing on small RNA sequences relevant to cancer. Furthermore, the update paper from OGEE, the gene essentiality database, includes an interesting focus on genes that are differentially essential in different cancers. More generally, DisGeNET and Open Targets [another new database designated as a ‘Breakthrough’ paper, (3)], both offer comprehensive resources linking pathogenic gene variants to a variety of other data.

The third database in this issue recognized by the *NAR* reviewers and editors with the ‘Breakthrough’ designation is DENOVO-DB, a database of mutations that have been found in human subjects but which were missing in both of their parents (4). The database lists ~32 000 sites in the genome with data obtained from >16 000 patients carrying some kind of a disease and >17,000 control individuals. The majority of disease variants were from individuals with autism and congenital heart disease with smaller samples coming from schizophrenia, epilepsy, and other neurodevelopmental disorders (4). There is no doubt that this collection will find a variety of uses, from analyzing de novo mutations linked to a particular disease to studying the frequencies of mutations in certain parts of the genome. The mirDNMR database is also a collection of *de novo* mutations with specific focus on the background mutation rates calculated by several statistical approaches (5).

Table 3. The ‘golden set’ of the most popular databases featured in multiple *NAR* issues^a

No. ^b	Database name	Current URL	Brief description	<i>NAR</i> publications, reference ^c
Annual updates				
1	DDBJ	http://www.ddbj.nig.ac.jp	All known nucleotide and protein sequences	2000, 2002–2017 (10)
2	ENA	http://www.ebi.ac.uk/ena	All known nucleotide and protein sequences	1986, 1990, 1997–2017 (11)
3	GenBank	https://www.ncbi.nlm.nih.gov/genbank/	All known nucleotide and protein sequences	1986, 1988, 1990–1994, 1996–2000, 2002–2017 (12)
27	Ensembl	http://www.ensembl.org/	Annotated information on eukaryotic genomes	2002–2017 (13)
87	Mouse Genome Database	http://www.informatics.jax.org	Mouse genome database	1997–2017 (14):
316	UCSC Genome Browser	http://genome.ucsc.edu/	A universal genome viewing and analysis platform	2006–2017 (15)
318	UniProt	http://www.uniprot.org	A universal database of protein sequences (includes Swiss-Prot and TrEMBL)	1991–1994, 1996–2000, 2003, 2004–2010, 2012–2015, 2017 (16)
Regular updates				
338	ArrayExpress	http://www.ebi.ac.uk/arrayexpress	Array-based gene expression data	2003, 2005, 2007, 2009, 2011, 2013, 2015 (17)
420	BioCyc ^d	http://biocyc.org/	Pathway information for sequenced genomes	2005, 2008, 2010, 2012, 2014, 2016 (18)
800	BioGRID	http://www.thebiogrid.org	Genetic and physical interactions in yeast, worm and fly	2006, 2008, 2011, 2013, 2015, 2017 (19)
421	BRENDA	http://www.brenda-enzymes.info	Enzyme names and biochemical properties	2002, 2004, 2007, 2009, 2011, 2013, 2015, 2017 (20)
645	CGD	http://www.candidagenome.org/	Candida Genome Database	2005, 2007, 2010, 2012, 2014, 2016 (21)
1531	CanSAR	http://cansar.icr.ac.uk	Cancer research and drug discovery resource	2012, 2014, 2016 (22)
258	CATH	http://www.cathdb.info	Protein domain structure database	1999, 2000, 2001, 2003, 2005, 2007, 2009, 2011, 2013, 2015, 2017 (23)
1211	CAZy	http://www.cazy.org	Carbohydrate-Active enZymes database	2009, 2014 (24)
204	CDD	http://www.ncbi.nlm.nih.gov/cdd	Conserved Domain Database	2002, 2003, 2005, 2007, 2009, 2011, 2013, 2015, 2017 (25)
646	ChEBI	http://www.ebi.ac.uk/chebi	Chemical Entities of Biological Interest	2008, 2013, 2016 (26)
1548	ChEMBL	https://www.ebi.ac.uk/chembl/	Interaction of drugs and compounds with their targets	2012, 2014, 2017 (27)
803	ChimerDB	http://ercsb.ewha.ac.kr/fusiongene	Chromosome translocations and gene fusions	2006, 2010, 2017 (28)
7	COG	http://www.ncbi.nlm.nih.gov/COG	Clusters of Orthologous Groups of proteins	2000, 2001, 2015 (29)
1188	Comparative Toxicogenomics Database	http://ctdbase.org	A knowledgebase for curated chemical-gene-disease networks	2009, 2011, 2013, 2017 (30)
651	COSMIC	http://cancer.sanger.ac.uk	Catalogue of Somatic Mutations in Cancer	2010, 2011, 2015, 2017 (31)
68	CyanoBase	http://genome.microbedb.jp/cyanobase	Cyanobacterial genomes	1998, 1999, 2000, 2010, 2014, 2017 (32)
885	dbPTM	http://dbPTM.mbc.nctu.edu.tw/	Post-translational modification of proteins	2006, 2013, 2014, 2016 (33)
591	DBTSS	http://dbtss.hgc.jp/	Database of transcriptional start sites	2002, 2004, 2006, 2008, 2010, 2012, 2015 (34)
445	DEG	http://www.essentialgene.org	Database of essential genes	2004, 2009, 2014 (35)
446	DictyBase	http://dictybase.org	Model organism database for <i>Dictyostelium discoideum</i>	2004, 2006, 2009, 2011, 2013 (36)
811	DrugBank	http://www.drugbank.ca/	Drug and drug target database	2006, 2008, 2011, 2014 (37)
108	EcoCyc	http://ecocyc.org/	<i>E. coli</i> K12 genes, metabolic pathways, transporters, and gene regulation	1996, 1997, 1998, 2000, 2002, 2005, 2009, 2011, 2013 (38)
1068	eggNOG	http://eggnog.embl.de/	Evolutionary genealogy of genes: Non-supervised Orthologous Groups	2008, 2010, 2012, 2014, 2016 (39)
1347	ELM	http://elm.eu.org/	Eukaryotic Linear Motif: functional sites in eukaryotic proteins	2003, 2008, 2010, 2011, 2012, 2014, 2016 (40)
812	EMAGE	http://www.emouseatlas.org/emage/	e-Mouse Atlas of Gene Expression	2006, 2008, 2010, 2014 (41)
985	ENCODE project at UCSC	http://genome.ucsc.edu/ENCODE	Encyclopedia of DNA Elements, functional elements in human genome	2007, 2010–2013 (42)
33	EPD	http://epd.vital-it.ch	Eukaryotic Promoter Database	1998, 1999, 2000, 2002, 2004, 2006, 2013, 2015, 2017 (43)
91, 969, 1219	EuPathD	http://eupathdb.org/	Unified genome databases on eukaryotic pathogens (includes PlasmoDB, ToxoDB, ApiDB, TrichDB, TriTrypDB, GiardiaDB, etc.)	2002, 2003, 2007–2013, 2017 (44)
1294	Expression Atlas	http://www.ebi.ac.uk/gxa/	Dene expression patterns deduced from microarray and RNA-seq data	2010, 2012, 2014, 2016 (45)
465	FANTOM	http://fantom.gsc.riken.jp/	Functional annotation of mouse full-length cDNA clones	2002, 2011, 2016, 2017 (46)
1020	FINDBase	http://www.findbase.org	Frequencies of INherited Disorders	2007, 2011, 2014, 2017 (47)
71	FlyBase	http://flybase.org/	Drosophila sequences and genomic information	1994, 1996–1999, 2002, 2003, 2005–2009, 2012–2017 (48)
817	FlyRNAi	http://flyrnai.org/	Genome-wide RNAi analysis in <i>Drosophila</i>	2006, 2012, 2017 (49)

Table 3. Continued

No. ^b	Database name	Current URL	Brief description	NAR publications, reference ^c
472	Gene3D	http://gene3d.biochem.ucl.ac.uk	Structural domain assignments for protein sequences	2003, 2005, 2006, 2008, 2010, 2012, 2014, 2016 (50)
73	Genenames	http://www.genenames.org/	The HGNC human gene nomenclature database	2008, 2011, 2013, 2015, 2017 (51)
989	GenomeRNAi	http://www.genomernai.org	RNA interference data for human and Drosophila	2007, 2010, 2013, 2017 (52)
603	GEO	http://www.ncbi.nlm.nih.gov/geo/	NCBI's Gene Expression Omnibus	2005, 2007, 2009, 2011, 2013 (53)
487	GO	http://www.geneontology.org	Gene Ontology Database	2004, 2006, 2008, 2010, 2012, 2013, 2015, 2017 (54)
389	GOA	http://www.ebi.ac.uk/GOA	Gene Ontology annotations for proteins in UniProt	2004, 2009, 2015 (55)
75	GOLD	https://gold.jgi.doe.gov/	Genomes online database: completed and ongoing genome projects	2001, 2006, 2008, 2010, 2012, 2015, 2017 (56)
166	GPCRdb	http://gpcrdb.org/	Data and tools for studying G protein-coupled receptors	1998, 2001, 2003, 2011, 2014, 2016 (57)
607	Gramene	http://www.gramene.org	Comparative genomics of crops and model plant species	2002, 2006, 2008, 2010, 2013, 2016 (58)
15	GXD	http://www.informatics.jax.org/expression.shtml	Mouse Gene Expression Database	1999-2001, 2004, 2007, 2011, 2014, 2017 (59)
1210	HAMAP	http://hamap.expasy.org/	High-quality Automated and Manual Annotation of Proteins	2009, 2013, 2015 (60)
991	HMDB	http://www.hmdb.ca	Human Metabolome Database	2007, 2009, 2013 (61)
779	IEDB	http://www.iedb.org/	Immune Epitope Database	2008, 2012, 2015 (62)
1089	IMG/M	http://img.jgi.doe.gov/m	JGI's Integrated Microbial Genomics and Metagenomics	2006, 2008, 2012, 2014, 2017 (63)
172	IMGT	http://www.imgt.org	International ImmunoGeneTics database.	1997-2001, 2003-2006, 2008-2010, 2015 (64)
690	InParanoid	http://InParanoid.sbc.su.se	Orthologous relationships between eukaryotic proteomes	2005, 2008, 2010, 2015 (65)
507	IntAct	http://www.ebi.ac.uk/intact/	Protein-Protein INTERACTioN data	2004, 2007, 2010, 2012, 2014 (66)
207	InterPro	http://www.ebi.ac.uk/interpro	Integrated resource of protein families, domains and functional sites	2001, 2003, 2005, 2007, 2009, 2012, 2015, 2017 (67)
367	IPD	http://www.ebi.ac.uk/ipd	Immuno Polymorphism database (includes IMGT/HLA)	2001, 2003, 2005, 2009, 2010, 2011, 2013, 2015 (68)
516	JASPAR	http://jaspar.genereg.net/	PSSMs for transcription factor DNA-binding sites	2004, 2006, 2008, 2010, 2014, 2016 (69)
112	KEGG	http://www.genome.ad.jp/kegg	Kyoto Encyclopedia of Genes and Genomes: genes, proteins, pathways	1999, 2000, 2002, 2004, 2006, 2008, 2010, 2012-2014, 2016, 2017 (70)
177	MEROPS	http://merops.sanger.ac.uk/	Database of proteases (peptidases)	1999, 2000, 2002, 2004, 2006, 2008, 2010, 2012, 2014, 2016 (71)
114	MetaCyc	http://metacyc.org/	Metabolic pathways and enzymes in various organisms	2000, 2002, 2004, 2006, 2008, 2010, 2012, 2014, 2016 (18)
529	miRBase	http://www.mirbase.org/	MicroRNA sequences, names and predicted targets in animals	2006, 2008, 2011, 2014 (72)
1098	miRGator	http://mirgator.kobic.re.kr	MicroRNA expression profiles and mRNA targets	2008, 2011, 2013 (73)
994	miRGen	http://www.microrna.gr/mirgen	MicroRNA promoters and transcription start sites	2007, 2010, 2016 (74)
1423	miRTarBase	http://miRTarBase.mbc.nctu.edu.tw/	Experimentally validated microRNA-target interactions	2011, 2013, 2016 (75)
270	MMDB	http://www.ncbi.nlm.nih.gov/Structure	Molecular Modeling Database of protein structures	1999, 2000, 2002, 2003, 2007, 2012, 2014 (76)
840	MODOMICS	http://genesilico.pl/modomics/	RNA modification pathways	2006, 2009, 2013 (77)
152	Mouse Tumor Biology Database	http://tumor.informatics.jax.org/mtbwi/	Mouse as a model system of human cancers	1999, 2000, 2007, 2015 (78)
1453	neXtProt	https://www.nextprot.org/	A database of human proteins	2012, 2015, 2017 (79)
705	NONCODE	http://noncode.org/	A database of noncoding RNAs	2005, 2008, 2012, 2014, 2016 (80)
143	OMIM	http://www.omim.org	Online Mendelian inheritance in man: A catalog of human genetic and genomic disorders	1994, 2002, 2005, 2009, 2015 (81)
1108	OrthoDB	http://www.orthodb.org	An hierarchical catalog of orthologous proteins	2008, 2011, 2013, 2015, 2017 (82)
552	PANTHER	http://www.pantherdb.org	Protein sequence evolution mapped to functions and pathways	2003, 2005, 2007, 2010, 2013, 2016, 2017 (83)
1000	PATRIC	http://www.patricbrc.org	PathoSystems Resource Integration Center	2007, 2014, 2017 (84)
276	PDB	http://rcsb.org/pdb	Protein DataBank: All biological macromolecular structures	2000-2002, 2004-2006, 2011, 2013, 2015, 2017 (85)
456	PDBe	http://www.ebi.ac.uk/pdbe/	Protein Databank in Europe	2010-2012, 2014, 2016 (86)
278	PDBsum	http://www.ebi.ac.uk/pdbsum	Summaries and analyses of PDB structures	2001, 2005, 2009, 2014 (87)
210	Pfam	http://pfam.xfam.org	Protein families: Multiple sequence alignments and profile hidden Markov models of protein domains	1998-2000, 2002, 2004, 2006, 2008, 2010, 2012, 2014, 2016 (8)
852	PHI-base	http://www4.rothamsted.bbsrc.ac.uk/phibase/	Genes affecting fungal pathogen-host interactions	2006, 2008, 2015, 2017 (88)
194	PIR	http://pir.georgetown.edu/	Protein Information Resource, part of UniProt	1986, 1988, 1991-1994, 1996-2004 (89)
857	PRIDE	http://www.ebi.ac.uk/pride/	Proteomics peptide identification database	2006, 2008, 2013, 2016 (90)
212	PRINTS	http://www.bioinf.man.ac.uk/dbbrowser/PRINTS	Protein fingerprints, conserved motifs used to characterise a protein family	1994, 1996-2000, 2002, 2003 (91)

Table 3. Continued

No. ^b	Database name	Current URL	Brief description	NAR publications, reference ^c
215	Prosite	http://www.expasy.org/prosite	Biologically-significant protein patterns and profiles	1991-1994, 1996, 1997, 1999, 2002, 2004, 2006, 2008, 2010, 2013 (92)
735	PubChem	http://pubchem.ncbi.nlm.nih.gov/	Structures and biological activities of small organic molecules	2009, 2010, 2014, 2016, 2017 (93)
93	RGD	http://rgd.mcw.edu/	Rat Genome Database	2002, 2005, 2007, 2009, 2015 (94)
243	RDP	http://rdp.cme.msu.edu	Ribosomal Database Project: Bacterial and archaeal 16S rRNA and fungal 28S rRNA sequences	19991-1994, 1996, 1997, 1999-2001, 2003, 2005, 2007, 2009, 2014 (95)
612	Reactome	http://www.reactome.org	A database of metabolic and signaling pathways	2005, 2009, 2011, 2014, 2016 (96)
224	REBASE	http://rebase.neb.com/rebase/	Restriction enzyme database	1993, 1994, 1996-2001, 2003, 2005, 2007, 2010, 2015 (97)
391	RefSeq	https://www.ncbi.nlm.nih.gov/refseq/	NCBI Reference Sequence Database	2000, 2001, 2005, 2007, 2009, 2012, 2014-2016 (98)
382	Rfam	http://rfam.xfam.org	RNA families with multiple sequence alignments	2003, 2005, 2009, 2011, 2013, 2015 (99)
282	SCOP	http://scop.mrc-lmb.cam.ac.uk/	Structural Classification Of Proteins	1997, 1999, 2000, 2002, 2004, 2008, 2014 (100)
352	SGD	http://www.yeastgenome.org	Saccharomyces Genome Database	1998, 1999, 2002-2008, 2010, 2012, 2014, 2016 (101)
1183	SILVA	http://www.arb-silva.de/	Aligned small- and large subunit rRNA sequences	2007, 2013, 2014 (102)
867	SIMAP	http://mips.gsf.de/simap/	Similarity Matrix of Proteins	2006, 2008, 2010, 2014 (103)
218	SMART	http://smart.embl-heidelberg.de	Simple Modular Architecture Research Tool: signalling, extracellular and chromatin-associated protein domains	1999, 2000, 2002, 2004, 2006, 2009, 2012, 2015 (104)
1134	STITCH	http://stitch-db.org/	Search Tool for Interactions of Chemicals	2008, 2010, 2012, 2014, 2016 (105)
582	STRING	http://string.embl.de/	Predicted functional associations between proteins	2000, 2003, 2005, 2007, 2009, 2011, 2013, 2015, 2017 (106)
285	SUPERFAMILY	http://supfam.org	Genome-wide identification of protein domains of known structure	2002, 2004, 2007, 2009, 2011, 2015 (107)
585	SWISS-MODEL	http://swissmodel.expasy.org/	3D models for proteins of unknown structure	2003, 2004, 2006, 2009, 2014, 2017 (108)
97	TAIR	http://www.arabidopsis.org/	The Arabidopsis information resource	2001, 2003, 2008, 2012 (109)
1264	TarBase	http://microrna.gr/tarbase	Database of experimentally supported microRNA targets	2006, 2009, 2012, 2015 (110)
790	TCDB	http://www.tcdb.org/	Transporter protein classification database	2006, 2009, 2014, 2016 (111)
1452	UCSC Cancer Genomics Browser	https://genome-cancer.ucsc.edu/	Visualization of cancer genomic datasets	2011, 2013, 2015 (15)
1031	VectorBase	https://www.vectorbase.org/	Invertebrate vectors of human pathogens	2007, 2009, 2012, 2015 (112)
51	WormBase	http://www.wormbase.org	Community portal on all aspects of <i>C. elegans</i> biology	2001, 2003-2008, 2010, 2012, 2014, 2016 (113)
1151	XenBase	http://www.xenbase.org	Xenopus frog database	2008, 2010, 2013, 2015 (114)
792	YEASTRACT	http://www.yeasttract.com	Transcriptional regulation in <i>Saccharomyces cerevisiae</i>	2006, 2008, 2011, 2014 (115)
101	ZFIN	http://zfin.org/	Zebrafish information network	2001, 2003, 2011, 2013 (116)

^aThis list includes databases that have been featured in the NAR Database Issue multiple times as separate papers. This listing omits many NCBI databases whose updated descriptions are published in annual NCBI overview papers.

^bThe database entry in the NAR online Database Collection. For example, the summary for ArrayExpress (no. 338) is available at <http://www.oxfordjournals.org/nar/database/summary/338>.

^cThe reference to the most recent database description that is available in PubMed (excludes the current issue).

^dThis database has switched to subscription-based service and is no longer available without registration.

Finally, in the genomic variation section, a major new arrival is the ExAC browser providing access to exome sequences from over 60 000 human genomes. This unprecedented depth of sampling of human genome data has important implications for attempts to predict observed SNP sequence variants as benign or damaging (6).

Plant databases represented here include an update to the popular PlantTFDB, collecting information on plant transcription factors and SUBA, recording plant subcellular localization data in *Arabidopsis*. Important new databases here include AraPheno, dealing with phenotypic data for the same model plant, and the intriguing PLaMoM which covers macromolecules, nucleic acids and plants, that are mobile over long distances in plants.

Finally, this issue includes descriptions of two important proteomics databases, an update on the widely used ProteomeXchange, dealing with standards and dissemination of proteomics data, and a first paper from one of its members describing the Japanese Proteomics Standards (jPOSTrepo) repository.

UPDATED NAR ONLINE MOLECULAR BIOLOGY DATABASE COLLECTION

This year's update of the NAR online Molecular Biology Database Collection (which is freely available at <http://www.oxfordjournals.org/nar/database/a/>), involved inclusion of 55 new databases (Table 1) and 15 databases that have been previously described elsewhere and were not part of this Collection (Table 2). In the current update, 18 duplicate en-

tries and 30 obsolete databases have been removed from the Collection, and five new databases have been added to the list.

Suggestions for inclusion of additional databases in the *NAR* Collection should be addressed to Xosé M. Fernández-Suárez 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.

LOOKING BACK: WHAT HAS CHANGED, WHAT HAS NOT

The 2006 editorial by MYG (7) included the following paragraph: ‘After 12 years of database issues and 8 years of the accompanying web supplement, it was interesting to check if they are really having an impact. In other words, how many people really care about them and use them? To evaluate the impact of the *NAR* database issues, I have used a tool that, despite all complaints and caveats, is commonly utilized for evaluating research productivity, namely the *Science Citation Index*[®] produced by the Institute for Scientific Information (ISI). If databases are put on the web for the benefit of the research community, the frequency with which people use (and cite) a given database could serve as an indication of whether this database serves a useful purpose. An inspection of the citation figures for the 141 papers published 2 years ago in the 2004 *NAR* Database Issue (all citation data are as of 15 October 2005) revealed a very encouraging trend. Most of the papers were well—or very well—cited. Only five papers have not been cited at all and the same number of database descriptions — five — have been cited >100 times, becoming, in ISI parlance, instant ‘citation classics’. Whatever the caveats, the fact that the paper describing the Pfam domain database [<http://www.sanger.ac.uk/Software/Pfam/>, *NAR* Collection entry no. 210, (6)] has been cited 375 times in <2 years definitely indicates that this database is widely used by the research community. Indeed, comparing a protein sequence against Pfam has become standard practice in sequence analysis, particularly in genome annotation. It is probably no coincidence that the first author of the Pfam paper also serves as the Editor of the *NAR* database issues. In the interest of full disclosure, I have cited this Pfam paper myself eight times since its publication in 2004.’ We hope the readers will excuse this small piece of self-plagiarism, which shows how little has changed in more than a decade. Pfam still remains our citation leader, even though it has been moved from the Wellcome Trust Sanger Institute to the EMBL-European Bioinformatics Institute and its URL has been changed to <http://pfam.xfam.org> (8). The *NAR* Database Issue as a whole is still very well cited and serves as a publication venue for a wide variety of hugely popular databases. In response to repeated requests of various researchers, Table 3 presents a list of such perennial favorites published in *NAR* three or more times.

SOME LESSONS LEARNT

Ten years ago Alex Bateman published an editorial that included a section on ‘What makes a good database?’

(9). This paper remains a must-read for anyone planning to submit a paper to the *NAR* Database Issue. It is available online at http://nar.oxfordjournals.org/content/35/suppl_1/D1.full and is linked from the *NAR* Instructions to authors page https://academic.oup.com/nar/pages/ms_prep_database. Here are some additional recommendations that might be useful for future database authors.

1. The database is expected to be maintained for many years, so it is worth spending some effort on finding it a proper name. Names that use the words from standard vocabulary or well-known commercial terms make it difficult for the potential users to find the database URL on the web. Among the databases published this year, the name of the Japanese Proteomics Standard repository has been changed from jPOST (a name shared with the web site of the Jerusalem Post newspaper) to JPOSTrepo, which does not have such a connotation. However, the previously published HIPPIE (Human Integrated Protein–Protein Interaction rEference) and RAID (RNA-associated interactions database) retained their original names. While it is quite unlikely that anyone would confuse the name of the former with the hippie lifestyle or the latter with the name of the popular pest killer, it is equally unlikely that either database will show up near the top in any web search. Therefore, it is always a good idea to do a Google search on the proposed database name: who knows what it might mean in other languages or even in urban slang. That said, creative mnemonics could be helpfully memorable, with notable examples including HIC-Up (Hetero-compound Information Centre-Uppsala), InParanoid (In-paralogs and orthologs in mammalian genomes), COSMIC (Catalog Of Somatic Mutations In Cancer), DARNED (DATabase of RNA EDiting), FINDbase (Frequencies of INherited Disorders database), and YEASTRACT (YEASt Transcriptional Regulators And Consensus Tracking).
2. To warrant publication, a new database must offer carefully curated data and provide substantial improvement in coverage and convenience over all previously created databases (including previous versions of that same database). Many prospective authors have been utterly surprised by this requirement and disappointed to see that there already existed similar (or better) databases created elsewhere. It was always hard for us to understand why would someone commit time and effort to constructing a database without even checking what is already available on the web.
3. The curation requirement means that mere integration of previously created databases is not going to be welcomed, no matter how complicated and successful that integration might have been. The only exceptions we have considered were consortium projects (such as RNACentral and ProteomeXchange in this issue) where authors of diverse databases committed to jointly maintaining their resources and exchanging the data for the benefit of the community.
4. Scientists, like everybody else, are not immune to fashion. In the past years, we have seen rapidly rising—and then quickly falling—numbers of databases dedicated

to protein–protein interactions, noncoding RNA, microRNA, their targets, long noncoding RNA, disease-related genes, drugs and drug targets, and so on. In cases like that, we used to refrain from choosing the best database among several ones created at the same time. Instead, we accepted two or three similar papers and allowed the respective databases to prove themselves. In the emerging areas of research, we see nothing wrong with a bit of competition, as long as these databases remain functional and regularly updated, and continue offering a useful service to the community.

- While it is hard for any single group to compete with such database juggernauts as NCBI, EMBL-EBI, BGI or Swiss Institute for Bioinformatics, some of the most successful databases, such as CAZy and GPCRdb (Table 3), have been created and are being maintained by relatively small groups.

CHANGING OF THE GUARD

This issue has been jointly edited by Drs. Michael Y. Galperin (Bethesda, MD, USA) and Daniel J. Rigden (Liverpool, UK). At the end of 2016, the former has retired from editing *NAR* and the latter assumed full responsibility for the *NAR* Database Issue. We are going to continue using the same E-mail address nardatabase@gmail.com and will adhere to the same database selection principles that have been introduced by the founding editor Sir Richard J. Roberts and successfully continued by Drs. Andreas D. Baxevasis, Alex Bateman, and, most recently, Michael Y. Galperin.

ACKNOWLEDGEMENTS

We thank *NAR* Editorial Administrator Dr. Martine Bernardes-Silva and the Oxford University Press team led by Jennifer Boyd for their help in assembling this issue.

FUNDING

The NIH Intramural Research Program at the National Library of Medicine (to M.Y.G.). The open access publication charge for this paper has been waived by Oxford University Press.

Conflict of interest statement. The authors' opinions do not necessarily reflect the views of their respective institutions. XMFS is an employee of Thermo Fisher Scientific Inc.

REFERENCES

- Akiva, E., Brown, S., Almonacid, D.E., Barber, A.E. 2nd, Custer, A.F., Hicks, M.A., Huang, C.C., Lauck, F., Mashiyama, S.T., Meng, E.C. *et al.* (2014) The Structure-Function Linkage Database. *Nucleic Acids Res.*, **42**, D521–D530.
- Mungall, C.J., McMurry, J.A., Köhler, S., Balhoff, J.P., Borromeo, C., Brush, M., Carbon, S., Conlin, T., Dunn, N., Engelstad, M. *et al.* (2017) The Monarch Initiative: an integrative data and analytic platform connecting phenotypes to genotypes across species. *Nucleic Acids Res.*, **45**, doi:10.1093/nar/gkw1128.
- Koscielny, G., An, P., Carvalho-Silva, D., Cham, J.A., Muñoz-Pomer Fuentes, A., Fumis, L., Gasparyan, R., Hasan, S., Karamanis, N., Maguire, M. *et al.* (2017) Open Targets: a platform for therapeutic target identification and validation. *Nucleic Acids Res.*, **45**, doi:10.1093/nar/gkw1055.
- Turner, T.N., Yi, Q., Krumm, N., Huddleston, J., Hoekzema, K., Stessman, H.A.F., Debly, A.-L., Bernier, R.A., Nickerson, D.A. and Eichler, E.E. (2017) denovo-db: a compendium of human *de novo* variants. *Nucleic Acids Res.*, **45**, doi:10.1093/nar/gkw1865.
- Jiang, Y., Li, Z., Liu, Z., Chen, D., Wu, W., Du, Y., Ji, L., Jin, Z.B., Li, W. and Wu, J. (2017) mirDNMR: a gene-centered database of background *de novo* mutation rates in human. *Nucleic Acids Res.*, **45**, doi:10.1093/nar/gkw1044.
- Editorial. (2016) Rethink the links between genes and disease. *Nature*, **538**, 140.
- Galperin, M.Y. (2006) The Molecular Biology Database Collection: 2006 update. *Nucleic Acids Res.*, **34**, D3–D5.
- Finn, R.D., Coghill, 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.
- Bateman, A. (2007) Editorial. *Nucleic Acids Res.*, **35**, D1–D2.
- Mashima, J., Kodama, Y., Kosuge, T., Fujisawa, T., Katayama, T., Nagasaki, H., Okuda, Y., Kaminuma, E., Ogasawara, O., Okubo, K. *et al.* (2016) DNA Data Bank of Japan (DDBJ) progress report. *Nucleic Acids Res.*, **44**, D51–D57.
- Gibson, R., Alako, B., Amid, C., Cerdeno-Tarraga, A., Cleland, I., Goodgame, N., Ten Hoopen, P., Jayathilaka, S., Kay, S., Leinonen, R. *et al.* (2016) Biocuration of functional annotation at the European nucleotide archive. *Nucleic Acids Res.*, **44**, D58–D66.
- Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J. and Sayers, E.W. (2016) GenBank. *Nucleic Acids Res.*, **44**, D67–D72.
- Yates, A., Akanni, W., Amode, M.R., Barrell, D., Billis, K., Carvalho-Silva, D., Cummins, C., Clapham, P., Fitzgerald, S., Gil, L. *et al.* (2016) Ensembl 2016. *Nucleic Acids Res.*, **44**, D710–D716.
- Bult, C.J., Eppig, J.T., Blake, J.A., Kadin, J.A. and Richardson, J.E. (2016) Mouse genome database 2016. *Nucleic Acids Res.*, **44**, D840–D847.
- Speir, M.L., Zweig, A.S., Rosenbloom, K.R., Raney, B.J., Paten, B., Nejad, P., Lee, B.T., Learned, K., Karolchik, D., Hinrichs, A.S. *et al.* (2016) The UCSC Genome Browser database: 2016 update. *Nucleic Acids Res.*, **44**, D717–D725.
- UniProt Consortium. (2015) UniProt: a hub for protein information. *Nucleic Acids Res.*, **43**, D204–D212.
- Kolesnikov, N., Hastings, E., Keays, M., Melnichuk, O., Tang, Y.A., Williams, E., Dylag, M., Kurbatova, N., Brandizi, M., Burdett, T. *et al.* (2015) ArrayExpress update—simplifying data submissions. *Nucleic Acids Res.*, **43**, D1113–D1116.
- Caspi, R., Billington, R., Ferrer, L., Foerster, H., Fulcher, C.A., Keseler, I.M., Kothari, A., Krummenacker, M., Latendresse, M., Mueller, L.A. *et al.* (2016) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.*, **44**, D471–D480.
- Chatr-Aryamontri, A., Breitkreutz, B.J., Oughtred, R., Boucher, L., Heinicke, S., Chen, D., Stark, C., Breitkreutz, A., Kolas, N., O'Donnell, L. *et al.* (2015) The BioGRID interaction database: 2015 update. *Nucleic Acids Res.*, **43**, D470–D478.
- Chang, A., Schomburg, I., Placzek, S., Jeske, L., Ulbrich, M., Xiao, M., Sensen, C.W. and Schomburg, D. (2015) BRENDA in 2015: exciting developments in its 25th year of existence. *Nucleic Acids Res.*, **43**, D439–D446.
- Binkley, J., Arnaud, M.B., Inglis, D.O., Skrzypek, M.S., Shah, P., Wymore, F., Binkley, G., Miyasato, S.R., Simison, M. and Sherlock, G. (2014) The Candida Genome Database: the new homology information page highlights protein similarity and phylogeny. *Nucleic Acids Res.*, **42**, D711–D716.
- Tym, J.E., Mitsopoulos, C., Coker, E.A., Razaz, P., Schierz, A.C., Antolin, A.A. and Al-Lazikani, B. (2016) canSAR: an updated cancer research and drug discovery knowledgebase. *Nucleic Acids Res.*, **44**, D938–D943.
- Sillitoe, I., Lewis, T.E., Cuff, A., Das, S., Ashford, P., Dawson, N.L., Furnham, N., Laskowski, R.A., Lee, D., Lees, J.G. *et al.* (2015) CATH: comprehensive structural and functional annotations for genome sequences. *Nucleic Acids Res.*, **43**, D376–D381.
- Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M. and Henrissat, B. (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.*, **42**, D490–D495.

25. Marchler-Bauer, A., Derbyshire, M.K., Gonzales, N.R., Lu, S., Chitsaz, F., Geer, L.Y., Geer, R.C., He, J., Gwadz, M., Hurwitz, D.I. *et al.* (2015) CDD: NCBI's conserved domain database. *Nucleic Acids Res.*, **43**, D222–D226.
26. Hastings, J., Owen, G., Dekker, A., Ennis, M., Kale, N., Muthukrishnan, V., Turner, S., Swainston, N., Mendes, P. and Steinbeck, C. (2016) ChEBI in 2016: Improved services and an expanding collection of metabolites. *Nucleic Acids Res.*, **44**, D1214–D1219.
27. Bento, A.P., Gaulton, A., Hersey, A., Bellis, L.J., Chambers, J., Davies, M., Kruger, F.A., Light, Y., Mak, L., McGlinchey, S. *et al.* (2014) The ChEMBL bioactivity database: an update. *Nucleic Acids Res.*, **42**, D1083–D1090.
28. Kim, P., Yoon, S., Kim, N., Lee, S., Ko, M., Lee, H., Kang, H., Kim, J. and Lee, S. (2010) ChimerDB 2.0—a knowledgebase for fusion genes updated. *Nucleic Acids Res.*, **38**, D81–D85.
29. Galperin, M.Y., Makarova, K.S., Wolf, Y.I. and Koonin, E.V. (2015) Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Res.*, **43**, D261–D269.
30. Davis, A.P., Murphy, C.G., Johnson, R., Lay, J.M., Lennon-Hopkins, K., Saraceni-Richards, C., Sciaky, D., King, B.L., Rosenstein, M.C., Wieggers, T.C. *et al.* (2013) The Comparative Toxicogenomics Database: update 2013. *Nucleic Acids Res.*, **41**, D1104–D1114.
31. Forbes, S.A., Beare, D., Gunasekaran, P., Leung, K., Bindal, N., Boutselakis, H., Ding, M., Bamford, S., Cole, C., Ward, S. *et al.* (2015) COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res.*, **43**, D805–D811.
32. Fujisawa, T., Okamoto, S., Katayama, T., Nakao, M., Yoshimura, H., Kajiya-Kanegae, H., Yamamoto, S., Yano, C., Yanaka, Y., Maita, H. *et al.* (2014) CyanoBase and RhizoBase: databases of manually curated annotations for cyanobacterial and rhizobial genomes. *Nucleic Acids Res.*, **42**, D666–D670.
33. Huang, K.Y., Su, M.G., Kao, H.J., Hsieh, Y.C., Jhong, J.H., Cheng, K.H., Huang, H.D. and Lee, T.Y. (2016) dbPTM 2016: 10-year anniversary of a resource for post-translational modification of proteins. *Nucleic Acids Res.*, **44**, D435–D446.
34. Suzuki, A., Wakaguri, H., Yamashita, R., Kawano, S., Tsuchihara, K., Sugano, S., Suzuki, Y. and Nakai, K. (2015) DBTSS as an integrative platform for transcriptome, epigenome and genome sequence variation data. *Nucleic Acids Res.*, **43**, D87–D91.
35. Luo, H., Lin, Y., Gao, F., Zhang, C.T. and Zhang, R. (2014) DEG 10, an update of the database of essential genes that includes both protein-coding genes and noncoding genomic elements. *Nucleic Acids Res.*, **42**, D574–D580.
36. Basu, S., Fey, P., Pandit, Y., Dodson, R., Kibbe, W.A. and Chisholm, R.L. (2013) DictyBase 2013: integrating multiple Dictyostelid species. *Nucleic Acids Res.*, **41**, D676–D683.
37. Law, V., Knox, C., Djoumbou, Y., Jewison, T., Guo, A.C., Liu, Y., Maciejewski, A., Arndt, D., Wilson, M., Neveu, V. *et al.* (2014) DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res.*, **42**, D1091–D1097.
38. Keseler, I.M., Mackie, A., Peralta-Gil, M., Santos-Zavaleta, A., Gama-Castro, S., Bonavides-Martinez, C., Fulcher, C., Huerta, A.M., Kothari, A., Krummenacker, M. *et al.* (2013) EcoCyc: fusing model organism databases with systems biology. *Nucleic Acids Res.*, **41**, D605–D612.
39. Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M.C., Rattai, T., Mende, D.R., Sunagawa, S., Kuhn, M. *et al.* (2016) eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res.*, **44**, D286–D293.
40. Dinkel, H., Van Roey, K., Michael, S., Kumar, M., Uyar, B., Altenberg, B., Milchevskaya, V., Schneider, M., Kuhn, H., Behrendt, A. *et al.* (2016) ELM 2016—data update and new functionality of the eukaryotic linear motif resource. *Nucleic Acids Res.*, **44**, D294–D300.
41. Richardson, L., Venkataraman, S., Stevenson, P., Yang, Y., Moss, J., Graham, L., Burton, N., Hill, B., Rao, J., Baldock, R.A. *et al.* (2014) EMAGE mouse embryo spatial gene expression database: 2014 update. *Nucleic Acids Res.*, **42**, D835–D844.
42. Rosenbloom, K.R., Sloan, C.A., Malladi, V.S., Dreszer, T.R., Learned, K., Kirkup, V.M., Wong, M.C., Maddren, M., Fang, R., Heitner, S.G. *et al.* (2013) ENCODE data in the UCSC Genome Browser: year 5 update. *Nucleic Acids Res.*, **41**, D56–D63.
43. Dreos, R., Ambrosini, G., Perier, R.C. and Bucher, P. (2015) The Eukaryotic Promoter Database: expansion of EPDnew and new promoter analysis tools. *Nucleic Acids Res.*, **43**, D92–D96.
44. Aurecochea, C., Barreto, A., Brestelli, J., Brunk, B.P., Cade, S., Doherty, R., Fischer, S., Gajria, B., Gao, X., Gingle, A. *et al.* (2013) EuPathDB: the eukaryotic pathogen database. *Nucleic Acids Res.*, **41**, D684–D691.
45. Petryszak, R., Keays, M., Tang, Y.A., Fonseca, N.A., Barrera, E., Burdett, T., Fullgrave, A., Fuentes, A.M., Jupp, S., Koskinen, S. *et al.* (2016) Expression Atlas update—an integrated database of gene and protein expression in humans, animals and plants. *Nucleic Acids Res.*, **44**, D746–D752.
46. Kawaji, H., Severin, J., Lizio, M., Forrest, A.R., van Nimwegen, E., Rehli, M., Schroder, K., Irvine, K., Suzuki, H., Carninci, P. *et al.* (2011) Update of the FANTOM web resource: from mammalian transcriptional landscape to its dynamic regulation. *Nucleic Acids Res.*, **39**, D856–D860.
47. Papadopoulos, P., Viennas, E., Gkantouna, V., Pavlidis, C., Bartsakoulia, M., Ioannou, Z.M., Ratb, I.I., Sefiani, A., Tsaknakis, J., Poulas, K. *et al.* (2014) Developments in FINDBase worldwide database for clinically relevant genomic variation allele frequencies. *Nucleic Acids Res.*, **42**, D1020–D1026.
48. Attrill, H., Falls, K., Goodman, J.L., Millburn, G.H., Antonazzo, G., Rey, A.J. and Marygold, S.J. (2016) FlyBase: establishing a Gene Group resource for *Drosophila melanogaster*. *Nucleic Acids Res.*, **44**, D786–D792.
49. Flockhart, I.T., Booker, M., Hu, Y., McElvany, B., Gilly, Q., Mathy-Prevot, B., Perrimon, N. and Mohr, S.E. (2012) FlyRNAi.org—the database of the *Drosophila* RNAi screening center: 2012 update. *Nucleic Acids Res.*, **40**, D715–D719.
50. Lam, S.D., Dawson, N.L., Das, S., Sillitoe, I., Ashford, P., Lee, D., Lehtinen, S., Orengo, C.A. and Lees, J.G. (2016) Gene3D: expanding the utility of domain assignments. *Nucleic Acids Res.*, **44**, D404–D409.
51. Gray, K.A., Yates, B., Seal, R.L., Wright, M.W. and Bruford, E.A. (2015) Genenames.org: the HGNC resources in 2015. *Nucleic Acids Res.*, **43**, D1079–D1085.
52. Schmidt, E.E., Pelz, O., Buhlmann, S., Kerr, G., Horn, T. and Boutros, M. (2013) GenomeRNAi: a database for cell-based and in vivo RNAi phenotypes, 2013 update. *Nucleic Acids Res.*, **41**, D1021–D1026.
53. Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., Marshall, K.A., Phillippy, K.H., Sherman, P.M., Holko, M. *et al.* (2013) NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.*, **41**, D991–D995.
54. Gene Ontology Consortium. (2015) Gene Ontology Consortium: going forward. *Nucleic Acids Res.*, **43**, D1049–D1056.
55. Huntley, R.P., Sawford, T., Mutowo-Muullenet, P., Shypitsyna, A., Bonilla, C., Martin, M.J. and O'Donovan, C. (2015) The GOA database: Gene Ontology Annotation updates for 2015. *Nucleic Acids Res.*, **43**, D1057–D1063.
56. Reddy, T.B., Thomas, A.D., Stamatidis, D., Bertsch, J., Isbandi, M., Jansson, J., Mallajosyula, J., Pagani, I., Lobos, E.A. and Kyrpides, N.C. (2015) The Genomes OnLine Database (GOLD) v.5: a metadata management system based on a four level (meta)genome project classification. *Nucleic Acids Res.*, **43**, D1099–D1106.
57. Isberg, V., Mordalski, S., Munk, C., Rataj, K., Harpsøe, K., Hauser, A.S., Vroling, B., Bojarski, A.J., Vriend, G. and Gloriam, D.E. (2016) GPCRdb: an information system for G protein-coupled receptors. *Nucleic Acids Res.*, **44**, D356–D364.
58. Tello-Ruiz, M.K., Stein, J., Wei, S., Preece, J., Olson, A., Naithani, S., Amarasinghe, V., Dharmawardhana, P., Jiao, Y., Mulvaney, J. *et al.* (2016) Gramene 2016: comparative plant genomics and pathway resources. *Nucleic Acids Res.*, **44**, D1133–D1140.
59. Smith, C.M., Finger, J.H., Hayamizu, T.F., McCright, I.J., Xu, J., Berghout, J., Campbell, J., Corbani, L.E., Forthofer, K.L., Frost, P.J. *et al.* (2014) The mouse Gene Expression Database (GXD): 2014 update. *Nucleic Acids Res.*, **42**, D818–D824.
60. Pedruzzi, I., Rivoire, C., Auchincloss, A.H., Coudert, E., Keller, G., de Castro, E., Baratin, D., Cuhe, B.A., Bougueleret, L., Poux, S. *et al.* (2015) HAMAP in 2015: updates to the protein family classification and annotation system. *Nucleic Acids Res.*, **43**, D1064–D1070.

61. Wishart, D.S., Jewison, T., Guo, A.C., Wilson, M., Knox, C., Liu, Y., Djoumbou, Y., Mandal, R., Aziat, F., Dong, E. *et al.* (2013) HMDB 3.0—The Human Metabolome Database in 2013. *Nucleic Acids Res.*, **41**, D801–D807.
62. Vita, R., Overton, J.A., Greenbaum, J.A., Ponomarenko, J., Clark, J.D., Cantrell, J.R., Wheeler, D.K., Gabbard, J.L., Hix, D., Sette, A. *et al.* (2015) The immune epitope database (IEDB) 3.0. *Nucleic Acids Res.*, **43**, D405–D412.
63. Markowitz, V.M., Chen, I.M., Palaniappan, K., Chu, K., Szeto, E., Pillay, M., Ratner, A., Huang, J., Woyske, T., Huntemann, M. *et al.* (2014) IMG 4 version of the integrated microbial genomes comparative analysis system. *Nucleic Acids Res.*, **42**, D560–D567.
64. Lefranc, M.P., Giudicelli, V., Duroux, P., Jabado-Michaloud, J., Folch, G., Aouinti, S., Carillon, E., Duvergey, H., Houles, A., Paysan-Lafosse, T. *et al.* (2015) IMGT[®], the international ImMunoGeneTics information system[®] 25 years on. *Nucleic Acids Res.*, **43**, D413–D422.
65. Sonnhammer, E.L.L. and Östlund, G. (2015) InParanoid 8: orthology analysis between 273 proteomes, mostly eukaryotic. *Nucleic Acids Res.*, **43**, D234–D239.
66. Orchard, S., Ammari, M., Aranda, B., Breuza, L., Briganti, L., Broackes-Carter, F., Campbell, N.H., Chavali, G., Chen, C., del-Toro, N. *et al.* (2014) The MIntAct project—IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res.*, **42**, D358–D363.
67. Mitchell, A., Chang, H.Y., Daugherty, L., Fraser, M., Hunter, S., Lopez, R., McAnulla, C., McMenamin, C., Nuka, G., Pesseat, S. *et al.* (2015) The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Res.*, **43**, D213–D221.
68. Robinson, J., Halliwell, J.A., Hayhurst, J.D., Flicek, P., Parham, P. and Marsh, S.G. (2015) The IPD and IMGT/HLA database: allele variant databases. *Nucleic Acids Res.*, **43**, D423–D431.
69. Mathelier, A., Fornes, O., Arenillas, D.J., Chen, C.Y., Denay, G., Lee, J., Shi, W., Shyr, C., Tan, G., Worsley-Hunt, R. *et al.* (2016) JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.*, **44**, D110–D115.
70. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. and Tanabe, M. (2016) KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.*, **44**, D457–D462.
71. Rawlings, N.D., Barrett, A.J. and Finn, R. (2016) Twenty years of the MEROPS database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res.*, **44**, D343–D350.
72. Kozomara, A. and Griffiths-Jones, S. (2014) miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.*, **42**, D68–D73.
73. Cho, S., Jang, I., Jun, Y., Yoon, S., Ko, M., Kwon, Y., Choi, I., Chang, H., Ryu, D., Lee, B. *et al.* (2013) MiRGator v3.0: a microRNA portal for deep sequencing, expression profiling and mRNA targeting. *Nucleic Acids Res.*, **41**, D252–D257.
74. Georgakilas, G., Vlachos, I.S., Zagganas, K., Vergoulis, T., Paraskevopoulou, M.D., Kanellos, I., Tsanakas, P., Dellis, D., Fevgas, A., Dalamagas, T. *et al.* (2016) DIANA-miRGen v3.0: accurate characterization of microRNA promoters and their regulators. *Nucleic Acids Res.*, **44**, D190–D195.
75. Chou, C.H., Chang, N.W., Shrestha, S., Hsu, S.D., Lin, Y.L., Lee, W.H., Yang, C.D., Hong, H.C., Wei, T.Y., Tu, S.J. *et al.* (2016) miRTarBase 2016: updates to the experimentally validated miRNA–target interactions database. *Nucleic Acids Res.*, **44**, D239–D247.
76. Madej, T., Lanczycki, C.J., Zhang, D., Thiessen, P.A., Geer, R.C., Marchler-Bauer, A. and Bryant, S.H. (2014) MMDB and VAST+: tracking structural similarities between macromolecular complexes. *Nucleic Acids Res.*, **42**, D297–D303.
77. Machnicka, M.A., Milanowska, K., Osman Oglou, O., Purta, E., Kurkowska, M., Olchowski, A., Januszewski, W., Kalinowski, S., Dunin-Horkawicz, S., Rother, K.M. *et al.* (2013) MODOMICS: a database of RNA modification pathways—2013 update. *Nucleic Acids Res.*, **41**, D262–D267.
78. Bult, C.J., Krupke, D.M., Begley, D.A., Richardson, J.E., Neuhauser, S.B., Sundberg, J.P. and Eppig, J.T. (2015) Mouse Tumor Biology (MTB): a database of mouse models for human cancer. *Nucleic Acids Res.*, **43**, D818–D824.
79. Gaudet, P., Michel, P.A., Zahn-Zabal, M., Cusin, I., Duek, P.D., Evalet, O., Gateau, A., Gleizes, A., Pereira, M., Teixeira, D. *et al.* (2015) The neXtProt knowledgebase on human proteins: current status. *Nucleic Acids Res.*, **43**, D764–D770.
80. Zhao, Y., Li, H., Fang, S., Kang, Y., Wu, W., Hao, Y., Li, Z., Bu, D., Sun, N., Zhang, M.Q. *et al.* (2016) NONCODE 2016: an informative and valuable data source of long non-coding RNAs. *Nucleic Acids Res.*, **44**, D203–D208.
81. Amberger, J.S., Bocchini, C.A., Schiettecatte, F., Scott, A.F. and Hamosh, A. (2015) OMIM.org: Online Mendelian Inheritance in Man (OMIM[®]), an online catalog of human genes and genetic disorders. *Nucleic Acids Res.*, **43**, D789–D798.
82. Kriventseva, E.V., Tegenfeldt, F., Petty, T.J., Waterhouse, R.M., Simao, F.A., Pozdnyakov, I.A., Ioannidis, P. and Zdobnov, E.M. (2015) OrthoDB v8: update of the hierarchical catalog of orthologs and the underlying free software. *Nucleic Acids Res.*, **43**, D250–D256.
83. 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.
84. Wattam, A.R., Abraham, D., Dalay, O., Disz, T.L., Driscoll, T., Gabbard, J.L., Gillespie, J.J., Gough, R., Hix, D., Kenyon, R. *et al.* (2014) PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Res.*, **42**, D581–D591.
85. Rose, P.W., Prlic, A., Bi, C., Bluhm, W.F., Christie, C.H., Dutta, S., Green, R.K., Goodsell, D.S., Westbrook, J.D., Woo, J. *et al.* (2015) The RCSB Protein Data Bank: views of structural biology for basic and applied research and education. *Nucleic Acids Res.*, **43**, D345–D356.
86. Velankar, S., van Ginkel, G., Alhroub, Y., Battle, G.M., Berrisford, J.M., Conroy, M.J., Dana, J.M., Gore, S.P., Gutmanas, A., Haslam, P. *et al.* (2016) PDBE: improved accessibility of macromolecular structure data from PDB and EMDB. *Nucleic Acids Res.*, **44**, D385–D395.
87. de Beer, T.A., Berka, K., Thornton, J.M. and Laskowski, R.A. (2014) PDBsum additions. *Nucleic Acids Res.*, **42**, D292–D296.
88. Urban, M., Pant, R., Raghunath, A., Irvine, A.G., Pedro, H. and Hammond-Kosack, K.E. (2015) The pathogen–host Interactions database (PHI-base): additions and future developments. *Nucleic Acids Res.*, **43**, D645–D655.
89. Wu, C.H., Nikolskaya, A., Huang, H., Yeh, L.S., Natale, D.A., Vinayaka, C.R., Hu, Z.Z., Mazumder, R., Kumar, S., Kourtesis, P. *et al.* (2004) PIRSF: family classification system at the Protein Information Resource. *Nucleic Acids Res.*, **32**, D112–D114.
90. Vizcaino, J.A., Csordas, A., del-Toro, N., Dianas, J.A., Griss, J., Lavidas, I., Mayer, G., Perez-Riverol, Y., Reisinger, F., Ternent, T. *et al.* (2016) 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res.*, **44**, D447–D456.
91. Attwood, T.K., Bradley, P., Flower, D.R., Gaulton, A., Maudling, N., Mitchell, A.L., Moulton, G., Nordle, A., Paine, K., Taylor, P. *et al.* (2003) PRINTS and its automatic supplement, prePRINTS. *Nucleic Acids Res.*, **31**, D400–D402.
92. Sigrist, C.J., de Castro, E., Cerutti, L., Cuče, B.A., Hulo, N., Bridge, A., Bougueleret, L. and Xenarios, I. (2013) New and continuing developments at PROSITE. *Nucleic Acids Res.*, **41**, D344–D347.
93. Kim, S., Thiessen, P.A., Bolton, E.E., Chen, J., Fu, G., Gindulyte, A., Han, L., He, J., He, S., Shoemaker, B.A. *et al.* (2016) PubChem Substance and Compound databases. *Nucleic Acids Res.*, **44**, D1202–D1213.
94. Shimoyama, M., De Pons, J., Hayman, G.T., Laulederkind, S.J., Liu, W., Nigam, R., Petri, V., Smith, J.R., Tutaj, M., Wang, S.J. *et al.* (2015) The Rat Genome Database 2015: genomic, phenotypic and environmental variations and disease. *Nucleic Acids Res.*, **43**, D743–D750.
95. Cole, J.R., Wang, Q., Fish, J.A., Chai, B., McGarrell, D.M., Sun, Y., Brown, C.T., Porras-Alfaro, A., Kuske, C.R. and Tiedje, J.M. (2014) Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res.*, **42**, D633–D642.
96. Fabregat, A., Sidiropoulos, K., Garapati, P., Gillespie, M., Hausmann, K., Haw, R., Jassal, B., Jupe, S., Korninger, F., McKay, S. *et al.* (2016) The Reactome pathway Knowledgebase. *Nucleic Acids Res.*, **44**, D481–D487.
97. Roberts, R.J., Vincze, T., Posfai, J. and Macelis, D. (2015) REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res.*, **43**, D298–D299.

98. O'Leary, N.A., Wright, M.W., Brister, J.R., Ciuffo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D. *et al.* (2016) Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.*, **44**, D733–D745.
99. Nawrocki, E.P., Burge, S.W., XMFS, A., Daub, J., Eberhardt, R. Y., Eddy, S.R., Floden, E.W., Gardner, P.P., Jones, T.A., Tate, J. *et al.* (2015) Rfam 12.0: updates to the RNA families database. *Nucleic Acids Res.*, **43**, D130–D137.
100. Andreeva, A., Howorth, D., Chothia, C., Kulesha, E. and Murzin, A.G. (2014) SCOP2 prototype: a new approach to protein structure mining. *Nucleic Acids Res.*, **42**, D310–D314.
101. Sheppard, T.K., Hitz, B.C., Engel, S.R., Song, G., Balakrishnan, R., Binkley, G., Costanzo, M.C., Dalusag, K.S., Demeter, J., Hellerstedt, S.T. *et al.* (2016) The Saccharomyces Genome Database Variant Viewer. *Nucleic Acids Res.*, **44**, D698–D702.
102. Yilmaz, P., Parfrey, L.W., Yarza, P., Gerken, J., Poeschl, E., Quast, C., Schweer, T., Peplies, J., Ludwig, W. and Glockner, F.O. (2014) The SILVA and 'All-species Living Tree Project (LTP)' taxonomic frameworks. *Nucleic Acids Res.*, **42**, D643–D648.
103. Arnold, R., Goldenberg, F., Mewes, H.W. and Rattei, T. (2014) SIMAP—the database of all-against-all protein sequence similarities and annotations with new interfaces and increased coverage. *Nucleic Acids Res.*, **42**, D279–D284.
104. Letunic, I., Doerks, T. and Bork, P. (2015) SMART: recent updates, new developments and status in 2015. *Nucleic Acids Res.*, **43**, D257–D260.
105. Szklarczyk, D., Santos, A., von Mering, C., Jensen, L.J., Bork, P. and Kuhn, M. (2016) STITCH 5: augmenting protein-chemical interaction networks with tissue and affinity data. *Nucleic Acids Res.*, **44**, D380–D384.
106. Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P. *et al.* (2015) STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.*, **43**, D447–D452.
107. Oates, M.E., Stahlhake, J., Vavoulis, D.V., Smithers, B., Rackham, O.J., Sardar, A.J., Zauha, J., Thurlby, N., Fang, H. and Gough, J. (2015) The SUPERFAMILY 1.75 database in 2014: a doubling of data. *Nucleic Acids Res.*, **43**, D227–D233.
108. Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Gallo Cassarino, T., Bertoni, M., Bordoli, L. *et al.* (2014) SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.*, **42**, W252–W258.
109. Lamesch, P., Berardini, T.Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., Muller, R., Dreher, K., Alexander, D.L., Garcia-Hernandez, M. *et al.* (2012) The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.*, **40**, D1202–D1210.
110. Vlachos, I.S., Paraskevopoulou, M.D., Karagkouni, D., Georgakilas, G., Vergoulis, T., Kanellos, I., Anastasopoulos, I.L., Maniou, S., Karathanou, K., Kalfakakou, D. *et al.* (2015) DIANA-TarBase v7.0: indexing more than half a million experimentally supported miRNA:mRNA interactions. *Nucleic Acids Res.*, **43**, D153–D159.
111. Saier, M.H. Jr, Reddy, V.S., Tsu, B.V., Ahmed, M.S., Li, C. and Moreno-Hagelsieb, G. (2016) The Transporter Classification Database (TCDB): recent advances. *Nucleic Acids Res.*, **44**, D372–D379.
112. Giraldo-Calderon, G.I., Emrich, S.J., MacCallum, R.M., Maslen, G., Dialynas, E., Topalis, P., Ho, N., Gesing, S., Madey, G., Collins, F.H. *et al.* (2015) VectorBase: an updated bioinformatics resource for invertebrate vectors and other organisms related with human diseases. *Nucleic Acids Res.*, **43**, D707–D713.
113. Howe, K.L., Bolt, B.J., Cain, S., Chan, J., Chen, W.J., Davis, P., Done, J., Down, T., Gao, S., Grove, C. *et al.* (2016) WormBase 2016: expanding to enable helminth genomic research. *Nucleic Acids Res.*, **44**, D774–D780.
114. Karpinka, J.B., Fortriede, J.D., Burns, K.A., James-Zorn, C., Ponferrada, V.G., Lee, J., Karimi, K., Zorn, A.M. and Vize, P.D. (2015) Xenbase, the *Xenopus* model organism database; new virtualized system, data types and genomes. *Nucleic Acids Res.*, **43**, D756–D763.
115. Teixeira, M.C., Monteiro, P.T., Guerreiro, J.F., Goncalves, J.P., Mira, N.P., dos Santos, S.C., Cabrito, T.R., Palma, M., Costa, C., Francisco, A.P. *et al.* (2014) The YEASTRACT database: an upgraded information system for the analysis of gene and genomic transcription regulation in *Saccharomyces cerevisiae*. *Nucleic Acids Res.*, **42**, D161–D166.
116. Howe, D.G., Bradford, Y.M., Conlin, T., Eagle, A.E., Fashena, D., Frazer, K., Knight, J., Mani, P., Martin, R., Moxon, S.A. *et al.* (2013) ZFIN, the Zebrafish Model Organism Database: increased support for mutants and transgenics. *Nucleic Acids Res.*, **41**, D854–D860.