RESOURCE REVIEW

Biological Databases for Human Research

Dong Zou #,a, Lina Ma #,b, Jun Yu *,c, Zhang Zhang *.d

CAS Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, China

Received 1 January 2015; revised 16 January 2015; accepted 16 January 2015
Available online 21 February 2015

Handled by Ge Gao

Abstract The completion of the Human Genome Project lays a foundation for systematically studying the human genome from evolutionary history to precision medicine against diseases. With the explosive growth of biological data, there is an increasing number of biological databases that have been developed in aid of human-related research. Here we present a collection of human-related biological databases and provide a mini-review by classifying them into different categories according to their data types. As human-related databases continue to grow not only in count but also in volume, challenges are ahead in big data storage, processing, exchange and curation.

Introduction

As biological data accumulate at larger scales and increase at exponential paces, thanks principally to higher-throughput and lower-cost DNA sequencing technologies, the number of biological databases that have been developed to manage such data deluge is growing at ever-faster rates. The major objectives of biological databases are not only to store, organize and share data in a structured and searchable manner with the aim to facilitate data retrieval and visualization for humans, but also to provide web application programming interfaces (APIs) for computers to exchange and integrate data from various database resources in an automated manner. Therefore, developing databases to deal with gigantic volumes of biological data is a fundamentally essential task in bioinformatics. To be short, biological databases integrate enormous amounts of omics data, serving as crucially important resources and becoming increasingly indispensable for scientists from wet-lab biologists to in silico bioinformaticians.

According to a report of 2014 Molecular Biology Database Collection in the journal Nucleic Acids Research, there are a sum of 1552 databases that are publicly accessible online [1]. It should be noted, however, that such count of publicly accessible databases is conservative. In fact, there are some databases providing online services without publication in peer-reviewed journal (e.g., The RNA Modification Database at http://mods.rna.albany.edu) or being developed by commercial companies (e.g., Ingenuity Pathway Analysis at http://www.ingenuity.com/products/ipa), making them under-represented in the scientific community. Considering the continuously proliferating number of biological databases, it becomes increasingly daunting and time-consuming to navigate in the huge volume of databases of interest. The

* Corresponding authors.
E-mail: junyu@big.ac.cn (Yu J), zhangzhang@big.ac.cn (Zhang Z).
# Equal contribution.
a ORCID: 0000-0002-7169-4965.
b ORCID: 0000-0001-6390-6289.
c ORCID: 0000-0002-7202-055X.
d ORCID: 0000-0001-6603-5060.

Peer review under responsibility of Beijing Institute of Genomics, Chinese Academy of Sciences and Genetics Society of China.

http://dx.doi.org/10.1016/j.gpb.2015.01.006
1672-0229 © 2015 The Authors. Production and hosting by Elsevier B.V. on behalf of Beijing Institute of Genomics, Chinese Academy of Sciences and Genetics Society of China.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
completion of the Human Genome Project in 2003 holds significant benefits for many fields from human evolution to personalized healthcare and precision medicine. In this report, we present a collection of biological databases relevant to human research and provide a mini-review by classifying them into different categories.

**Database classification**

Biological databases are developed for diverse purposes, encompass various types of data at heterogeneous coverage and are curated at different levels with different methods, so that there are accordingly several different criteria applicable to database classification.

**Scope of data coverage**

According to the scope of data coverage, biological databases can be classified as comprehensive and specialized databases. Comprehensive databases cover different types of data from numerous species and typical examples are GenBank [2], European Molecular Biology Laboratory (EMBL) [3], and DNA Data Bank of Japan (DDBJ) [4]. These three databases were established as the International Nucleotide Sequence Database Collaboration in 1988 to collect and disseminate DNA and RNA sequences. On the other hand, specialized databases contain specific types of data or data from specific organisms. For example, WormBase [5] is for nematode biology and genomics and RiceWiki [6] is for community curation of rice genes.

**Level of biocuration**

According to level of data curation, biological databases can roughly fall into primary and secondary or derivative databases. Primary databases contain raw data as archival repository such as the NCBI Sequence Read Archive (SRA) [7], whereas secondary or derivative databases contain curated information as added value, e.g., NCBI RefSeq [8].

**Method of biocuration**

As a consequence of the explosive growth of data, curation increasingly requires collective intelligence for collaborative data integration and annotation. Therefore, biological databases can also be classified as (1) expert-curated databases, e.g., RefSeq [8] and TAIR, [9] and (2) community-curated databases, which are curated in a collective and collaborative manner by a number of researchers, e.g., LncRNAWiki [10] and GeneWiki [11].

**Type of data managed**

According to the types of data managed in different databases, biological databases can roughly fall into the following categories: (1) DNA, (2) RNA, (3) protein, (4) expression, (5) pathway, (6) disease, (7) nomenclature, (8) literature, and (9) standard and ontology.

**Human-related databases**

Decoding the human genome bears great significance in, from a theoretical view, unveiling human evolutionary history, and from an application view, exploring personalized medicine against diverse diseases. Considering the heterogeneity in data type, scope and curation, biological databases can be classified into multiple categories under different criteria as presented above, making it easier for people to effectively characterize databases and identify the database(s) of interest. However, some databases are inaccessible over time or poorly maintained/updated or even never used [12]. In this study, therefore, we assemble a collection of human-related databases that are widely used and currently accessible via the Internet (Table 1). As database classification based on data type is informative and straightforward, we assign one major category to each database, albeit one database may correspond to multiple categories. In what follows, we focus on databases categorized in DNA, RNA, protein, expression, pathway and disease, respectively.

**DNA databases**

A DNA database centers on managing DNA data from many or some specific species. The primary function of human DNA databases includes establishment of the reference genome (e.g., NCBI RefSeq [8]), profiling of human genetic variation (e.g., dbSNP [13]), association of genotype with phenotype (e.g., EGA [14]), and identification of human microbiome metagenomes (e.g., IMG/HMP [15]). A representative example of DNA database is GenBank [2], a collection of all publicly-available DNA sequences (http://www.ncbi.nlm.nih.gov/genbank). Since its inception in 1982, GenBank grows at an extraordinary pace and as of December 2014, contains over 184 billion nucleotide bases in more than 179 million sequences (http://www.ncbi.nlm.nih.gov/genbank/statistics).

**RNA databases**

It is well acknowledged that only a tiny proportion of the human genome is transcribed into mRNAs, whereas the vast majority of the genome is transcribed into “dark matter”—non-coding RNAs (ncRNAs) that do not encode proteins [16], including microRNAs (miRNAs), small nuclear RNAs (snRNAs), piwiRNAs (piRNAs), and long non-coding RNA (lncRNA). Therefore, an increasing number of human RNA databases have been built for deciphering ncRNAs (e.g., GENCODE [17]), in particular lncRNAs that attract the rising interest (e.g., LncRNAWiki [10]), and characterizing their functions and interactions (e.g., RNAcentral [18]). A representative example of RNA database is RNAcentral [18]. It provides unified access to the ncRNA sequence data supplied by multiple databases including Rfam [19], IncRNAdb [20], and miRBase [21] (http://rnacentral.org).

**Protein databases**

The purpose of constructing protein databases includes collection of universal proteins (e.g., UniProt [22]), identification of
| Name               | Link                                                                 | Brief description                                                                 | Refs.        | Category |
|--------------------|----------------------------------------------------------------------|-----------------------------------------------------------------------------------|--------------|----------|
| 1000 Genomes       | http://www.1000genomes.org                                          | A deep catalog of human genetic variation                                         | [17]         | DNA      |
| AFND               | http://www.allelefrequencies.net                                     | Allele Frequency Net Database                                                      | [37]         |          |
| dbSNP              | http://www.ncbi.nlm.nih.gov/snp                                      | Database of single nucleotide polymorphisms                                       | [13]         |          |
| DEG                | http://www.essentialgene.org                                        | Database of Essential Genes                                                       | [38]         |          |
| EGA                | http://www.ebi.ac.uk/ega                                             | European Genome–phenome Archive                                                    | [14]         |          |
| Ensembl            | http://www.ensembl.org                                              | Ensembl genome browser                                                            | [39]         |          |
| enuGenes           | http://eugen.org                                                    | Genomic information for eukaryotic organisms                                        | [40]         |          |
| GeneCards          | http://www.gene_cards.org                                           | Integrated database of human genes                                                | [41]         |          |
| IMG/HMP            | https://img.igi.doc.gov/img_h_mp                                     | Human Microbiome MetaGenomes                                                      | [15]         |          |
| JASPAR             | http://jaspar.dk                                                   | Transcription factor binding profile database                                       | [42]         |          |
| JGA                | http://trace.ddbj.nig.ac.jp/jga                                      | Japanese Genotype–phenotype Archive                                                | [43]         |          |
| KEGG               | http://www.genome.jp/kegg                                           | Kyoto Encyclopedia of Genes and Genomes                                            | [44]         |          |
| MITOMAP            | http://www.mitomap.org                                              | Human mitochondrial genome database                                               | [45]         |          |
| NCBI RefSeq        | http://www.ncbi.nlm.nih.gov/refseq                                  | NCBI Reference Sequence Database                                                   | [8]          |          |
| PolymiRTS          | http://compbio.uthsc.edu/miRNSNP                                   | Polymorphism in miRNAs and their Target Sites                                     | [46]         |          |
| UCSC Genome Browser| http://genome.ucsc.edu                                              | UCSC Genome Browser database                                                      | [47]         |          |
| ChIPBase           | http://deepbase.sysu.edu.cn/chipbase                                | Database of transcriptional regulation of lncRNA and miRNA genes                  | [48]         | RNA      |
| DARNED             | http://darned.ucc.ie                                                | Database of RNA EDiting in humans                                                 | [49]         |          |
| DIANA-LncBase      | http://diana.imis.athens-innovation.gr/                             | miRNA targets on lncRNAs                                                          | [50]         |          |
| GENCODE            | http://www.gencodegenes.org                                         | Encyclopedia of genes and gene variants                                            | [17]         |          |
| H-DBAS             | http://www.h-invitational.jp/h-dbas                                | Human-transcriptome Database for Alternative Splicing                             | [51]         |          |
| HEXEvent           | http://hexevent.mmg.aci.edu                                         | Database of Human EXon splicing Events                                             | [52]         |          |
| LNCipedia          | http://www.lncipedia.org                                           | Annotated human lncRNA sequences                                                   | [53]         |          |
| LncRNA2Target      | http://www.lncrna2target.org                                       | Database of differentially-expressed genes after lncRNA knockdown or overexpression| [54]         |          |
| lncRNAdb           | http://www.lncrnadb.org                                            | lncRNA Database                                                                    | [20]         |          |
| lncRNASNP          | http://bioinfo.life.hust.edu.cn/lncRNASNP                           | Database of SNPs in lncRNAs                                                       | [55]         |          |
| LncRNAWiki         | http://lncrna.big.ac.cn                                            | Human lncRNA Wiki                                                                 | [10]         |          |
| miRBase            | http://www.mirbase.org                                             | miRNA Database                                                                     | [21]         |          |
| miRTarBase         | http://mirtarbase.mbc.nctu.edu.tw                                   | Experimentally-validated miRNA–target interactions                                  | [56]         |          |
| miRWalk            | http://mirwalk.uni-hd.de                                           | Database of miRNA–target interactions                                             | [57]         |          |
| NONCODE            | http://www.noncode.org                                             | Database of ncRNA genes                                                            | [58]         |          |
| NPInter            | http://www.bioinfo.org/NPInter                                     | Database of ncRNA interactions                                                     | [59]         |          |
| RADAR              | http://RNAedit.com                                                 | Rigorously Annotated Database of A-to-I RNA editing                               | [60]         |          |
| piRNABank          | http://piRNAbank.cebriab.cinvestav.mx                              | Database of piwi-interacting RNAs                                                 | [61]         |          |
| RBPDB              | http://rbpdb.ccb.rutgers.edu                                       | Database of RNA-binding specificities                                              | [62]         |          |
| RDB                | http://ndbserver.rutgers.edu                                       | The nucleic acid database                                                         | [63]         |          |
| Rfam               | http://rfam.xfam.org                                               | Database of ncRNA families                                                         | [19]         |          |
| RNAcentral         | http://rnacentral.org                                              | International database of ncRNA sequences                                          | [18]         |          |
| snoRNABase         | https://www-snorna.biostat.ful                        | Database of human H/ACA and C/D box snoRNAs                                       | [64]         |          |
| starBase           | http://starbase.sysu.edu.cn                                        | Database of ncRNA interaction networks                                            | [65]         |          |
| TarBase            | http://diana.imis.athens-innovation.gr/                             | Experimentally-validated miRNA–gene interactions                                   | [66]         |          |
| TargetScan         | http://www.targetscan.org                                          | Predicted miRNA targets in mammals                                                | [67]         |          |
| CATH               | http://cath.biochem.ucl.ac.uk                                      | Protein structure classification                                                   | [68]         | Protein  |
| CPLM               | http://cplm.biocuckoo.org                                          | Compendium of Protein Lysine Modifications                                         | [69]         |          |
| DIP                | http://dip.doe-mbi.ucla.edu                                        | Database of Interacting Proteins                                                  | [70]         |          |
| EKPD               | http://ekpd.biocuckoo.org                                         | Eukaryotic Kinase and Phosphatase Database                                         | [71]         |          |
| HPRD               | http://www.hprd.org                                                | Human Protein Reference Database                                                   | [72]         |          |
| hUbiquitome        | http://hUbiquitome.bioinfo.bjmu.edu                                | Ubiquitination sites and cascades                                                  | [73]         |          |
| InterPro           | http://www.ebi.ac.uk/interpro                                      | Protein sequence analysis and classification                                       | [74]         |          |
| MEROPS             | http://merops.sanger.ac.uk                                         | Database of proteolytic enzymes, their substrates, and inhibitors                 | [75]         |          |
| MINT               | http://mint.bio.uniroma2.it/mint                                    | Molecular INTeraction Database                                                     | [76]         |          |

(continued)
| Name       | Link                                                                 | Brief description                                                                 | Refs. | Category |
|------------|----------------------------------------------------------------------|------------------------------------------------------------------------------------|-------|----------|
| ModBase    | http://salilab.org/modbase                                           | Database of comparative protein structure models                                   | [77]  |          |
| mUbiSiDa   | http://reprod.njmu.edu.cn/mUbiSiDa                                  | Mammalian Ubiquitination Site Database                                              | [78]  |          |
| PANTHER    | http://www.pantherdb.org                                             | Protein ANalysis THrough Evolutionary                                              | [79]  |          |
| PDB        | http://www.rcsb.org/pdb                                              | Protein Data Bank for 3D structures of biological macromolecules                  | [25]  |          |
| PDBe       | http://www.ebi.ac.uk/pdbe                                            | Protein Data Bank in Europe                                                        | [23]  |          |
| Pfam       | http://pfam.xfam.org                                                 | Database of conserved protein families and domains                                | [23]  |          |
| PhosSNP    | http://phossnp.biocuckoo.org                                         | Genetic polymorphisms that influence protein phosphorylation                       | [81]  |          |
| PIR        | http://pir.georgetown.edu                                           | Protein Information Resource                                                       | [82]  |          |
| PROSITE    | http://www.expasy.org/prosite                                        | Database of protein domains, families and functional sites                         | [83]  |          |
| SysPTM     | http://lifecenter.sgst.cn/SysPTM                                    | Post-translational modifications                                                   | [84]  |          |
| TreeFam    | http://www.treefam.org                                               | Database of phylogenetic trees of animal species                                  | [24]  |          |
| UniPROBE   | http://thebrain.bwh.harvard.edu/uniprobe                            | Universal PBM Resource for Oligonucleotide Binding Evaluation                     | [85]  |          |
| UniProt    | http://www.uniprot.org                                               | Universal protein resource                                                        | [22]  |          |
| UUCD       | http://uucd.biocuckoo.org                                            | Ubiquitin and Ubiquitin-like Conjugation Database                                  | [86]  |          |
| ArrayExpress | http://www.ebi.ac.uk/arrayexpress                                    | Database of functional genomics experiments                                         | [88]  | Expression |
| BioGPS     | http://biogps.org                                                    | Portal for querying and organizing gene annotation resources                       | [88]  |          |
| Expression Atlas | http://www.ebi.ac.uk/gxa                                          | Differential and baseline expression                                              | [27]  |          |
| Human Protein | http://www.proteinatlas.org                                         | Tissue-based map of the human proteome                                            | [29]  |          |
| Atlas      | MOPED https://www.proteinspire.org                                   | Multi-Omics Profiling Expression Database                                          | [89]  |          |
| NCBI GEO   | http://www.ncbi.nlm.nih.gov.geo                                      | Gene Expression Omnibus                                                            | [26]  |          |
| NRED       | http://nred.matticklab.com                                           | Database of IncRNA expression                                                     | [90]  |          |
| ONCOMINE   | https://www.oncomine.org                                             | Cancer microarray database                                                         | [91]  |          |
| PrimerBank | http://pga.mgh.harvard.edu/primerbank                                | Public resource for PCR primers                                                    | [92]  |          |
| PRIDE      | http://www.ebi.ac.uk/pride                                          | PRoteomics IDEntifications                                                        | [93]  |          |
| TiGER      | http://bioinfo.wilmer.jhu.edu/tiger                                  | Tissue-specific Gene Expression and Regulation                                      | [28]  |          |
| WikiCell   | http://www.wikicell.org                                              | Unified resource for Human transcriptomics research                                | [94]  |          |
| CPDB       | http://consensuspathdb.org                                           | Database of human interaction networks                                             | [95]  | Pathway   |
| HMDB       | http://www.hmdb.ca                                                   | Human Metabolome Database                                                          | [96]  |          |
| KEGG       | http://www.genome.jp/kegg/pathway.html                              | KEGG pathway maps                                                                  | [30]  |          |
| PATHWAY    | html                                                                 |                                                                                   |       |          |
| MetaCyc    | http://metacyc.org                                                   | Metabolic pathway database                                                         | [97]  |          |
| Pathway    | http://www.pathwaycommons.org                                      | Pathway commons                                                                    | [98]  |          |
| Commons    |                                                                     |                                                                                   |       |          |
| PID        | http://pid.nci.nih.gov                                               | Pathway Interaction Database                                                       | [99]  |          |
| Reactome   | http://www.reactome.org                                              | Curated and peer-reviewed pathway database                                         | [100] |          |
| UniPathway | http://www.grenoble.prabi.fr/obwarehouse/unipathway                 | Universal Pathway                                                                 | [101] |          |
| AlzBase    | http://alz.big.ac.cn/alzBase                                         | Database for gene dysregulation in Alzheimer’s disease                             | [102] | Disease   |
| CADgene    | http://www.bioguo.org/CADgene                                       | Coronary Artery Disease gene database                                              | [103] |          |
| COSMIC     | http://cancer.sanger.ac.uk                                          | Catalog Of Somatic Mutations In Cancer                                             | [104] |          |
| DiseaseMeth | http://bioinfo.hrbmu.edu.cn/diseasemeth                            | Human disease methylation database                                                  | [105] |          |
| DisGeNET   | http://www.disgenet.org/web/DisGeNETv2.1                            | Gene–disease associations                                                          | [106] |          |
| GOBO       | http://co.bmc.lu.se/gobo                                            | Gene expression-based Outcome for Breast cancer Online                             | [107] |          |
| GWAS Central | http://www.gwascentral.org                                              | A comprehensive resource for the comparison and interrogation of genome-wide association studies | [108] |          |
| GWASdb     | http://jjwanglab.org/gwasdb                                         | Human genetic variants identified by genome-wide association studies               | [109] |          |
| HbVar      | http://globin.cse.psu.edu/hbvar                                     | Hemoglobin variants and thalassemias                                                | [110] |          |
| HGMD       | http://www.hgmd.org                                                 | Human Gene Mutation Database                                                       | [111] |          |
protein families and domains (e.g., Pfam [23]), reconstruction of phylogenetic trees (e.g., TreeFam [24]), and profiling of protein structures (e.g., PDB [25]). A representative example of protein database is PDB, the main primary database for 3D structures of biological macromolecules determined by X-ray crystallography and NMR. Established in 1971, PDB contains 105,465 biological macromolecular structures as of 30 December 2014, in which 27,393 entries belong to human (http://www.rcsb.org/pdb).

Expression databases

Expression databases can be used for various purposes, including archiving expression data (e.g., GEO [26]), detecting differential and baseline expression (e.g., Expression Atlas [27]), exploring tissue-specific gene expression and regulation (e.g., TiGER [28]), and profiling expression information based on both RNA and protein data (e.g., Human Protein Atlas [29]).

Pathway databases

Pathway databases contain biological pathways for metabolic, signaling, and regulatory pathway analysis. A representative example is KEGG PATHWAY [30], a curated biological pathway resource on the molecular interaction and reaction networks. As the core of KEGG, KEGG PATHWAY integrates many entities that are stored in KEGG sibling databases, including genes, proteins, RNAs, chemical compounds, and chemical reactions (http://www.genome.jp/kegg/pathway.html).

Disease databases

There are at least 200 forms of cancer in the world, causing 14.6% of all human deaths (http://en.wikipedia.org/wiki/Cancer). Thus, obtaining complete cancer genomes and identifying molecular mutations and abnormal genes can provide new insights for cancer prevention, detection, and...
eventually, personalized treatment [31]. Toward this end, there are two well-known cancer projects, viz., The Cancer Genome Atlas (TCGA) [32] and International Cancer Genome Consortium (ICGC) [33]. TCGA, founded in 2006 by the National Cancer Institute and National Human Genome Research Institute at the National Institutes of Health, aims to collect a wide diversity of omics data (including exome, SNP, mRNA, miRNA, and methylation) for more than 20 different types of human cancer (http://cancergenome.nih.gov). Unlike TCGA, ICGC is a voluntary collaborative organization initiated in 2008 and open to all cancer and genomic researchers in the world. It aims to obtain a comprehensive description of genomic, transcriptomic, and epigenomic changes in 50 different tumor types and/or subtypes, which are of clinical and societal importance across the globe (http://icgc.org).

Perspectives

Here we summarize a collection of biological databases relevant to human research. This collection, however, by no means pictures the whole range of human-related databases that are currently available. As primary databases store raw data, databases in this collection are most derivative databases, which are built from primary databases and contain curated information for different data types, and thus would be of great usefulness for studying the human genome. In the era of big data, human-related biological databases continue to grow not only in count but also in volume, posing unprecedented challenges in data storage, processing, exchange, and curation. From this point, it would be necessary to establish a cloud computing platform to store and process such big data and facilitate construction/update of a secondary or derivative database [34]. As biological databases are physically distributed and heterogeneous in data type and format, it is additionally required to build web open APIs to ease data exchange and sharing among different resources [35]. The last but not the least is curation, which becomes an indispensable part in biological databases, principally because curation involves added value by standardization and quality control and accordingly enhances data interoperability and consistency [36]. Taken together, biological databases hold great utilities for human research and can be regarded as an indicator of our potential to translate big data into big discovery. Considering the current situation in China when compared to other countries, it is our hope that this report may raise the general awareness, albeit better data into big discovery. Considering the current situation in China when compared to other countries, it is our hope that this report may raise the general awareness, albeit better

Competing interests

The authors declared that there are no competing interests.

Acknowledgements

This work was supported by the “100-Talent Program” of Chinese Academy of Sciences, the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDB13040500), and the National High-tech R&D Program (863 Program; Grant No. 2012AA020409) by the Ministry of Science and Technology of China awarded to ZZ.

References

[1] Fernandez-Suarez XM, Rigden DJ, Galperin MY. The 2014 Nucleic Acids Research Database Issue and an updated NAR online Molecular Biology Database Collection. Nucleic Acids Res 2014;42:D1–D6.
[2] Benson DA, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Pruitt KD, Brown GR, Hiatt SM, Thibaud-Nissen F, Astashyn A, Natale LA, Lipman DJ. GenBank. Nucleic Acids Res 2014;42:D18–25.
[3] Brooksbank C, Bergman MT, Apweiler R, Birney E, Thornton JM. The European Bioinformatics Institute’s data resources 2014. Nucleic Acids Res 2014;42:D18–25.
[4] Kosuge T, Mashima J, Kodama Y, Fujisawa T, Kaminuma E, Ogasawara O, et al. DDBJ progress report: a new submission system for leading to a correct annotation. Nucleic Acids Res 2014;42:D44–9.
[5] Harris TW, Baran J, Bieri T, Cabunoc A, Chan J, Chen WJ, et al. WormBase 2014: new views of curated biology. Nucleic Acids Res 2014;42:D789–93.
[6] Zhang Z, Sang J, Ma L, Wu G, Wu H, Huang D, et al. RaceWiki: a wiki-based database for human long non-coding RNA genes. Nucleic Acids Res 2014;42:D1222–8.
[7] Kodama Y, Shumway M, Leinonen R International Nucleotide Sequence Database C. The Sequence Read Archive: explosive growth of sequencing data. Nucleic Acids Res 2012;40:D54–6.
[8] Pruitt KD, Brown GR, Hiatt SM, Thibaud-Nissen F, Astashyn A, Ermolaeva O, et al. RefSeq: an update on mammalian reference sequences. Nucleic Acids Res 2014;42:D756–63.
[9] Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, et al. The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. Nucleic Acids Res 2012;40:D1202–10.
[10] Ma L, Li A, Zou D, Xu X, Xia L, Yu J, et al. LncRNAWiki: harnessing community knowledge in collaborative curation of human long non-coding RNAs. Nucleic Acids Res 2015;43:D187–92.
[11] Good BM, Clarke EL, de Alfaro L, Su AI. The Gene wiki in 2011: community intelligence applied to human gene annotation. Nucleic Acids Res 2012;40:D1255–61.
[12] Wren JD, Bateman A. Databases, data toms and dust in the wind. Bioinformatics 2008;24:2127–8.
[13] Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res 2001;29:308–11.
[14] Leinonen R, Akhtar R, Birney E, Bower L, Cerdeno-Tarraga A, Cheng Y, et al. The European nucleotide archive. Nucleic Acids Res 2011;39:D28–31.
[15] Markowitz VM, Chen IM, Chu K, Szeto E, Palaniappan K, Skiray M, et al. IMG/M 4 version of the integrated metagenome comparative analysis system. Nucleic Acids Res 2014;42:D568–73.
[16] Ma L, Bajic VB, Zhang Z. On the classification of long non-coding RNAs. RNA Biol 2013;10:925–33.
[17] Genomes Project Consortium, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, et al. An integrated map of genetic variation from 1,092 human genomes. Nature 2012;491:56–65.
[18] The RNAcentral Consortium. RNAcentral: an international database of ncRNA sequences. Nucleic Acids Res 2015;43:D123–9.
[19] Burge SW, Daub J, Eberhardt R, Tate J, Barquist L, Nawrocki EP, et al. Rfam 11.0: 10 years of RNA families. Nucleic Acids Res 2013;41:D226–32.
[20] Quek XC, Thomson DW, Maag JL, Bartonicek N, Signal B, Clark MB, et al. LncRNAdb v2.0: expanding the reference
Zou D et al / Human-related Biological Databases

61

database for functional long noncoding RNAs. Nucleic Acids Res 2015;43:D168–73.

[21] Kozomara A, Griffiths-Jones S. MiRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res 2014;42:D68–73.

[22] The UniProt Consortium. Ongoing and future developments at the Universal Protein Resource. Nucleic Acids Res 2011;39:D214–9.

[23] Finn RD, Mistry J, Schuster-Bockler B, Griffiths-Jones S, Hollich V, Lassmann T, et al. Pfam: clans, web tools and services. Nucleic Acids Res 2006;34:D247–51.

[24] Li H, Coghlan A, Ruan J, Coin LJ, Hercig JK, Osmotherly L, et al. TreeFam: a curated database of phylogenetic trees of animal gene families. Nucleic Acids Research 2006;34:D575–80.

[25] Rose PW, Beran B, Bi C, Bluhm WF, Dimitropoulos D, Goodsell DS, et al. The RCSB Protein Data Bank: redesigned web site and web services. Nucleic Acids Res 2011;39:D392–401.

[26] Barrett T, Troup DB, Wilhite SE, Ledoux P, Evangelista C, Kim IF, et al. NCBI GEO: archive for functional genomics data sets – 10 years on. Nucleic Acids Res 2011;39:D100510.

[27] Petryszak R, Burdett T, Fiorelli B, Fonseca NA, Gonzalez-Porta M, Hastie E, et al. Expression Atlas update – a database of gene and transcript expression from microarray- and sequencing-based functional genomics experiments. Nucleic Acids Res 2014;42:D296–32.

[28] Liu X, Yu X, Zack DJ, Zhu H, Qian J. TiGER: a database for tissue-specific gene expression and regulation. BMC Bioinformatics 2008;9:271.

[29] Ponten F, Schwenk JM, Asplund A, Edqvist PH. The Human Protein Atlas as a proteomic resource for biomarker discovery. J Intern Med 2011;270:428–46.

[30] Okuda S, Yamada T, Hamajima M, Itoh M, Katayama T, Bork P, et al. KEGG Atlas mapping for global analysis of metabolic pathways. Nucleic Acids Res 2008;36:W423–6.

[31] Stratton MR, Campbell PJ, Futreal PA. The cancer genome. Nature 2009;458:719–24.

[32] Stratton MR, Campbell PJ, Futreal PA. The cancer genome. Nature 2009;458:719–24.

[33] Cancer Genome Atlas Research Network, Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, et al. The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet 2013;45:1113–20.

[34] International Cancer Genome Consortium, Hudson TJ, Anderson W, Artez A, Barker AD, Bell C, et al. International network of cancer genome projects. Nature 2010;464:993–8.

[35] Dai L, Gao X, Guo Y, Xiao J, Zhang Z. Bioinformatics clouds for big data manipulation. Biol Direct 2012;7:34.[discussion].

[36] Zhang Z, Zhu W, Luo J. Bringing biocuration to China. Genomics Proteomics Bioinformatics 2014;12:153–5.

[37] Gonzalez-Galarza FF, Christmas S, Middleton D, Jones AR. Allele frequency net: a database and online repository for immune gene frequencies in worldwide populations. Nucleic Acids Res 2011;39:D913–9.

[38] Luo H, Lin Y, Gao F, Zhang CT, Zhang R. DEG 10, an update of the database of essential genes that includes both protein-coding genes and noncoding genomic elements. Nucleic Acids Res 2014;42:D574–80.

[39] Flicek P, Amode MR, Barrett D, Beal K, Billis K, Brent S, et al. Ensembl 2014. Nucleic Acids Res 2014;42:D749–55.

[40] Gilbert DG. EuGenes: a eukaryote genome information system. Nucleic Acids Res 2002;30:145–8.

[41] Safran M, Chalifa-Caspi V, Shmueli O, Olender T, Lapidot M, Rosen N, et al. Human Gene-Centric Databases at the Weizmann Institute of Science: GeneCards. UDB, CroW 21 and HORDE. Nucleic Acids Res 2003;31:142–6.

[42] Mathelier A, Zhao X, Zhang AW, Parcy F, Worsley-Hunt R, Arenillas DJ, et al. JASP AR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. Nucleic Acids Res 2014;42:D142–7.

[43] Kodama Y, Mashima J, Kosuge T, Katayama T, Fujisawa T, Kaminuma E, et al. The DDBJ Japanese Genotype-phenotype archive for genetic and phenotypic human data. Nucleic Acids Res 2015;43:D18–22.

[44] Kanehisa M, Goto S, Yato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res 2012;40:D109–14.

[45] Ruiz-Pesini E, Lott MT, Procaccio V, Poole JC, Brandon MC, Mishmar D, et al. An enhanced MITOMAP with a global mtDNA mutational phylogeny. Nucleic Acids Res 2007;35:D823–8.

[46] Bhattacharya A, Ziebarth JD, Cui Y. PolymiRTS Database 3.0: linking polymorphisms in microRNAs and their target sites with human diseases and biological pathways. Nucleic Acids Res 2014;42:D86–91.

[47] Rosenbloom KR, Armstrong J, Barber GP, Casper J, Clawson H, Diekhashs M, et al. The UCSC Genome Browser Database: 2015 update. Nucleic Acids Res 2015;43:D670–81.

[48] Yang JH, Li JH, Jiang S, Zhou H, Qu LH. ChIPBase: a database for decoding the transcriptional regulation of long non-coding RNA and microRNA genes from ChIP-Seq data. Nucleic Acids Res 2013;41:D177–87.

[49] Kiran A, Baranov PV. DARNED: a database of A-to-I RNA editing. Nucleic Acids Res 2015;43:D181–6.

[50] Paraskevopoulou MD, Georgiaklas G, Kostoulas N, Reczek M, Maragkakis M, Dalamagas TM, et al. DIANA-LncBase: experimentally verified and computationally predicted microRNA targets on long non-coding RNAs. Nucleic Acids Res 2013;41:D239–45.

[51] Takeda J, Suzuki Y, Sakate R, Sato Y, Gojobori T, Imanishi T, et al. H-DBAS: human-transcriptome database for alternative splicing: update 2010. Nucleic Acids Res 2010;38:D86–90.

[52] Busch A, Hertel KJ. HEXEvent: a database of Human ExON splicing Events. Nucleic Acids Res 2013;41:D118–24.

[53] Volders PJ, Hulsens K, Wang X, Menten B, Martens L, Gevaert K, et al. NCIPedia: a database for annotated human lncRNA transcript sequences and structures. Nucleic Acids Res 2013;41:D246–51.

[54] Jiang Q, Wang J, Wu X, Ma R, Zhang T, Jin S, et al. LncRNA2Target: a database for differentially expressed genes after lncRNA knockdown or overexpression. Nucleic Acids Res 2015;43:D193–6.

[55] Gong J, Liu W, Zhang J, Miao X, Guo AY. LncRNA-SNP: a database of SNPs in lncRNAs and their potential functions in human and mouse. Nucleic Acids Res 2015;43:D811–6.

[56] Huo Z, Xie A, Shrestha S, Lin YL, Khaleel A, Chou CH, et al. MiRTarBase update 2014: an information resource for experimentally validated miRNA-target interactions. Nucleic Acids Res 2014;42:D78–85.

[57] Deepw H, Gretz N, Sticht C. MiRWalk database for miRNA-target interactions. Methods Mol Biol 2014;1182:289–305.

[58] Bu D, Yu K, Sun S, Xie C, Skogerbo G, Miao R, et al. NONCODE v3.0: integrative annotation of long noncoding RNAs. Nucleic Acids Res 2012;40:D210–5.

[59] Yuan J, Wu W, Xie C, Zhao G, Zhao Y, Chen R. NPInter v2.0: an updated database of ncRNA interactions. Nucleic Acids Res 2014;42:D104–8.

[60] Ramaswami G, Li JB, RADAR: a rigorously annotated database of A-to-I RNA editing. Nucleic Acids Res 2014;42:D109–13.

[61] Sai Lakshmi S, Agrawal S. piRNABank: a web resource on classified and clustered Piwi-interacting RNAs. Nucleic Acids Res 2008;36:D173–7.
Cook KB, Kazan H, Zuberi K, Morris Q, Hughes TR. RBPPDB: a database of RNA-binding specificities. Nucleic Acids Res 2011;39:D301–8.

Coimbatore Narayanan B, Westbrook J, Ghosh S, Petrov AI, Sweeney B, Zirbel CL, et al. The Nucleic Acid Database: new features and capabilities. Nucleic Acids Res 2014;42:D114–22.

Xie J, Zhang M, Zhou T, Hua X, Tang L, Wu W, Snio/serRNAbase: a curated database for small non-coding RNAs and cajal body-specific RNAs. Nucleic Acids Res 2007;35:D183–7.

Li JH, Liu S, Zhou H, Qu LH, Yang JH. starBase v2.0: decoding miRNA-ceRNA, miRNA-mRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. Nucleic Acids Res 2014;42:D92–7.

Vlacho IS, Parasekevopoulou MD, Karagkouni D, Georgakilas G, Vergoulis T, Kanellos I, et al. DIANA-TarBase v7.0: indexing more than half a million experimentally supported miRNA:mRNA interactions. Nucleic Acids Res 2015;43:D153–9.

Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell 2005;120:15–20.

Stilling I, Lewis SE, Cuff A, Das S, Ashford P, Dawson NL, et al. CATH: comprehensive structural and functional annotations for genome sequences. Nucleic Acids Res 2015;43:D376–81.

Liu Z, Wang Y, Gao T, Pan Z, Cheng H, Yang Q, et al. CPLM: a database of protein lysine modifications. Nucleic Acids Res 2014;42:D531–6.

 Xenarios I, Salwinski L, Duan XJ, Higney P, Kim SM, Eisenberg D. DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions. Nucleic Acids Res 2002;30:303–5.

Wang Y, Liu Z, Cheng H, Gao T, Pan Z, Yang Q, et al. EKPD: a hierarchical database of eukaryotic protein kinases and protein phosphatases. Nucleic Acids Res 2014;42:D496–502.

Keshava Prasad TS, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, et al. Human Protein Reference Database – 2009 update. Nucleic Acids Res 2009;37:D767–72.

Du Y, Xu N, Lu M, Li T. hUbiquitome: a database of experimentally verified ubiquitination cascades in humans. Database (Oxford) 2011;2011:bar055.

Mitchell A, Chang HY, Daugherty L, Fraser M, Hunter S, Lopez R, et al. The InterPro protein families database: the classification resource after 15 years. Nucleic Acids Res 2015;43:D213–21.

Rawlings ND, Waller M, Barrett AJ, Bateman A. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. Nucleic Acids Res 2014;42:D503–9.

Ceol A, Chatr-Aryamontri A, Liaca L, Peluso D, Briganti L, Perfetto L, et al. MINT, the molecular interaction database – 2009 update. Nucleic Acids Res 2009;2010(38):d3532–9.

Pieper U, Webb BM, Dong GQ, Schneidman-Duhovny D, Fan H, Kim SJ, et al. ModBase, a database of annotated comparative protein structure models and associated resources. Nucleic Acids Res 2014;42:D336–46.

Chen T, Zhou T, He B, Yu H, Guo X, Song X, et al. MUniSiDa: a comprehensive database for protein ubiquitination sites in mammals. PLoS One 2014;9:e85744.

Mi H, Guo N, Kejariwal A, Thomas PD. PANTHER version 6: protein sequence and function evolution data with expanded representation of biological pathways. Nucleic Acids Res 2007;35:D277–52.

Gutmannas A, Alhroub Y, Battle GM, Berrisford JM, Bochet E, Conroy MJ, et al. PDBe: Protein Data Bank in Europe. Nucleic Acids Res 2014;42:D285–91.

Ren J, Jiang C, Gao X, Liu Z, Yuan Z, Jin C, et al. PheoSNP for systematic analysis of genetic polymorphisms that influence protein phosphorylation. Mol Cell Proteomics 2010;9:623–34.

Wu CH, Yeh LS, Huang H, Armusinski L, Castro-Alveraj J, Chen Y, et al. The Protein Information Resource. Nucleic Acids Res 2003;31:345–7.

Sigrist CJ, Cerutti L, de Castro E, Langendijk-Genevaux PS, Buliard V, Baroche A, et al. PROSITE, a protein domain database for functional characterization and annotation. Nucleic Acids Res 2010;38:D616–6.

Li J, Liu J, Li H, Yu J, Sun H, He Y, et al. SysPTM 2.0: an updated systematic resource for post-translational modification. Database (Oxford) 2014;2014:bau025.

Hume MA, Barrera LA, Gisselbrecht SS, Bulyk ML, UniPROBE, update 2015: new tools and content for the online database of protein-binding microarray data on protein–DNA interactions. Nucleic Acids Res 2015;43:D117–22.

Gao T, Liu Z, Wang Y, Cheng H, Yang Q, Guo A, et al. UUCD: a family-based database of ubiquitin and ubiquitin-like conjugation. Nucleic Acids Res 2013;41:D445–51.

Parkinson H, Sarkans U, Kolesnikov N, Abeygunawardena N, Burdett T, Dylag M, et al. ArrayExpress update – an archive of microarray and high-throughput sequencing-based functional genomics experiments. Nucleic Acids Res 2011;39:D1002–4.

Wu C, Orozco C, Boyer J, Leglise M, Goodale J, Batalov S, et al. BioGPS: an extensible and customizable portal for querying and organizing genome annotation resources. Genome Biol 2009;10:R130.

Montague E, Janko I, Stanberry L, Lee E, Choiniere J, Anderson N, et al. Beyond protein expression, MOPED goes multi-omics. Nucleic Acids Res 2015;43:D1145–51.

Dinger ME, Pang KC, Mercer TR, Crowe ML, Grimmmon SM, Mattick JS. NRED: a database of long noncoding RNA expression. Nucleic Acids Res 2009;37:D122–6.

Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, et al. ONCOMINE: a cancer microarray database and integrated data-mining platform. Neoplasia 2004;6:1–6.

Wang X, Spandidos A, Wang H, Seed B. PrimerBank: a PCR primer database for quantitative gene expression analysis, 2012 update. Nucleic Acids Res 2012;40:D1144–9.

Jones P, Cote RG, Cho SY, Kim SJ, et al. PRIDE: new developments and new datasets. Nucleic Acids Res 2008;36:D878–83.

Zhao D, Wu J, Zhou Y, Gong W, Xiao J, Yu J. WikiCell: a unified resource platform for human transcriptomics research. Omics 2012;16:357–62.

Kamburov A, Stelzl U, Lehrah R, Herwig R, The ConsensusPathDB interaction database: 2013 update. Nucleic Acids Res 2013;41:D793–800.

Wischart DS, Jewison T, Guo AC, Wilson M, Knox C, Liu Y, et al. HMDB 3.0 – The Human Metabolome Database in 2013. Nucleic Acids Res 2013;41:D801–7.

Cospi R, Altman T, Billington R, Dreher K, Foester H, Fulker CA, et al. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. Nucleic Acids Res 2014;42:D459–71.

Cerami EG, Gross BE, Demir E, Rodchenkov I, Babur O, Anwar N, et al. Pathway commons, a web resource for biological pathway data. Nucleic Acids Res 2011;39:D685–90.

Schaefer CF, Anthony K, Krupa S, Buchhoff J, Day M, Hannay T, et al. PID: the pathway interaction database. Nucleic Acids Res 2009;37:D674–9.

Croft D, Mundo AF, Haw R, Milacic M, Weiser J, Wu G, et al. The Reactome pathway knowledgebase. Nucleic Acids Res 2014;42:D472–7.

Morgan A, Coissac E, Coudert E, Axelsen KB, Keller G, Bairoch A, et al. UniPathway: a resource for the exploration and annotation of metabolic pathways. Nucleic Acids Res 2012;40:D761–9.

Bai Z, Han G, Xie B, Wang J, Song F, Peng X, et al. AlzBase: an Integrative Database for gene dysregulation in Alzheimer's
[103] Liu H, Liu W, Liao Y, Cheng L, Liu Q, Ren X, et al. CADgene: a comprehensive database for coronary artery disease genes. Nucleic Acids Res 2011;39:D991–6.

[104] Forbes SA, Beare D, Gunasekaran P, Leung K, Bindal N, Boutsalakis H, et al. COSMIC: exploring the world’s knowledge of somatic mutations in human cancer. Nucleic Acids Res 2015;43:D805–11.

[105] Lv J, Liu H, Su J, Wu X, Liu H, Li B, et al. DiseaseMeth: a human disease methylation database. Nucleic Acids Res 2012;40:D1030–5.

[106] Bauer-Mehren A, Rautschka M, Sanz F, Furlong LI. DisGeNET: a cytoscape plugin to visualize, integrate, search and analyze gene-disease networks. Bioinformatics 2010;26:2924–6.

[107] Ringner M, Fredlund E, Hakkinen J, Borg A, Staaf J. GOBO: gene expression-based outcome for breast cancer online. PLoS One 2011;6:e17911.

[108] Li MJ, Wang P, Liu X, Lim EL, Wang Z, Yeager M, et al. GWASdb: a database for human genetic variants identified by genome-wide association studies. Nucleic Acids Res 2012;40:D1047–54.

[109] Giardine B, Borg J, Viennas E, Pavlidis C, Moradkhani K, Joly P, et al. Updates of the HbVar database of human hemoglobin variants and thalassemia mutations. Nucleic Acids Res 2014;42:D1063–9.

[110] Stenson PD, Ball EV, Mort M, Phillips AD, Shaw K, Cooper DN. The Human Gene Mutation Database (HGMD) and its exploitation in the fields of personalized genomics and molecular evolution. Curr Protoc Bioinformatics 2012 [Chapter 1, Unit1.13].

[111] Piirila H, Valiaho J, Vihinen M. Immunodeficiency mutation databases (IDbases). Hum Mutat 2006;27:1200–8.

[112] Chen G, Wang Z, Wang D, Qiu C, Liu M, Chen X, et al. LncRNADisease: a database for long-non-coding RNA-associated diseases. Nucleic Acids Res 2013;41:D983–6.

[113] Yang Z, Yang J, Liu W, Hao Y, Yuan L, Teng M, Zhang X, et al. MiR2Disease: a manually curated database for microRNA deregulation in human disease. Nucleic Acids Res 2009;37:D98–D104.

[114] Amberger JS, Bocchini CA, Schiettecatte F, Scott AF, Hamosh A. OMIM.org: online mendelian inheritance in man (OMIM(R)), an online catalog of human genes and genetic disorders. Nucleic Acids Res 2015;43:D789–98.

[115] The Gene Ontology Consortium. Gene Ontology Consortium: going forward. Nucleic Acids Res 2015;43:D1049–56.

[116] Gray KA, Yates B, Seal RL, Wright MW, Bruton EA. Genenames.org: the HGNC resources in 2015. Nucleic Acids Res 2015;43:D1079–85.

[117] The Europe PMC Consortium. Europe PMC: a full-text literature database for the life sciences and platform for innovation. Nucleic Acids Res 2015;43:D1042–8.

[118] Lu Z. PubMed and beyond: a survey of web tools for searching biomedical literature. Database (Oxford) 2011;2011:baq036.

[119] Sequeira E, McEntyre J, Lipman D. PubMed central decentralized. Nature 2001;410:740.