

Genenames.org: the HGNC resources in 2023

Ruth L. Seal^{1,2,*}, Bryony Braschi¹, Kristian Gray^{1,2}, Tamsin E.M. Jones¹, Susan Tweedie¹, Liara Haim-Vilmovsky¹ and Elspeth A. Bruford^{1,2}

¹HUGO Gene Nomenclature Committee, European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton CB10 1SD, UK and ²Department of Haematology, University of Cambridge School of Clinical Medicine, Cambridge CB2 0PT, UK

Received September 15, 2022; Revised September 28, 2022; Editorial Decision September 29, 2022; Accepted October 03, 2022

ABSTRACT

The HUGO Gene Nomenclature Committee (HGNC) assigns unique symbols and names to human genes. The HGNC database (www.genenames.org) currently contains over 43 000 approved gene symbols, over 19 200 of which are assigned to protein-coding genes, 14 000 to pseudogenes and nearly 9000 to non-coding RNA genes. The public website, www.genenames.org, displays all approved nomenclature within Symbol Reports that contain data curated by HGNC nomenclature advisors and links to related genomic, clinical, and proteomic information. Here, we describe updates to our resource, including improvements to our search facility and new download features.

INTRODUCTION

The HUGO Gene Nomenclature Committee (HGNC) has been in operation for over 40 years and maintains one of the longest running databases that delivers biological standards for the scientific community. The HGNC is the sole authority for approving human gene symbols and corresponding descriptive gene names. HGNC is an Elixir UK service node (1) and a recommended resource on FAIRsharing (2). Standardised gene nomenclature is essential for effective scientific communication and facilitates retrieval of information about genes. The HGNC works with researchers, gene annotators and representatives from the nomenclature committees for other species to name protein-coding genes, pseudogenes and non-coding RNA genes. We record previous symbols and names, and alias symbols and names. We recently published an article ‘The risks of using unapproved gene symbols’ (3) that presents examples of where the use of alias symbols can cause problems, to raise awareness of this to the research community. We have also published a review of how we name long non-coding RNA (lncRNA) genes (4), which includes a description of how the seven most highly published lncRNA genes were named.

Genes with HGNC-approved nomenclature are all assigned a unique ID in the format HGNC:#, where # is a unique number. HGNC IDs are stable and do not change if the gene symbol changes. Therefore, we encourage all resources discussing/citing human genes on a large scale to use HGNC IDs. These IDs should also ideally be mentioned in the literature, especially when the symbol of one gene matches the alias of another to disambiguate between such genes, e.g. *GRID1* (HGNC:4575) has the symbol alias GluD1 and there is a separate gene with the approved gene symbol *GLUD1* (HGNC:4335).

For each approved gene, the HGNC creates a Symbol Report on www.genenames.org that lists our approved nomenclature and related useful information (see Figure 1 for an example). The top section of the report entitled ‘HGNC data’ includes our approved nomenclature, locus type, HGNC ID, aliases, previous nomenclature where applicable, status (approved or withdrawn), and chromosomal location. In addition, the HGNC data section may include curator notes and/or links to gene groups curated by the HGNC. HGNC Symbol Reports also provide numerous useful external data links to biomedical resources such as NCBI Gene (5), Ensembl (6) UniProt (7), GeneCards (8), RNAcentral (9), UCSC genome browser (10), OMIM (11), ClinGen (12), GeneTests (13) and Monarch (14), which are sorted by type into the categories Gene Resources, Nucleotide Resources, Protein Resources, Clinical Resources and Other Resources. The ‘Orthologs from Selected Species’ section contains links for orthologs, if present, to MGD for mouse (15), RGD for rat (16) and VGNC (the HGNC’s sister project, the Vertebrate Gene Nomenclature Committee) for chimpanzee, rhesus macaque, dog, cat, horse, cattle and pig. For protein coding genes there is a separate ‘HCOP homology predictions’ tab that provides orthology predictions for 19 species, including all major model organisms, aggregated by our HGNC comparison of orthology predictions (HCOP) tool. The most recent version of the HCOP tool is described in full in (17).

In addition to HCOP, genenames.org has a general search function at the top of every page and a Multi-Symbol Checker tool that checks any inputted symbols against all

*To whom correspondence should be addressed. Tel: +44 1223 494444; Fax: +44 1223 494446; Email: rseal@ebi.ac.uk

Symbol report for AADAC Stable symbol ?

Report HCOP homology predictions

HGNC data for AADAC

Approved symbol ? AADAC
Approved name ? arylacetamide deacetylase
Locus type ? gene with protein product
HGNC ID ? HGNC:17
Symbol status ? Approved
Previous names ? " arylacetamide deacetylase (esterase) "
Alias symbols ? DAC; CES5A1
Chromosomal location ? 3q25.1
Gene groups ? Lipases
Arylacetamide deacetylase family

Gene resources for AADAC ?

Ensembl	ENSG00000114771 Curated Ensembl region in detail , Ensembl gene sequence	NCBI Gene	13 Curated
UCSC	uc003eze.4	Alliance of Genome Resources	HGNC:17

Nucleotide resources for AADAC ?

MANE Select	NM_001086.3 ENST00000232892.12	INSDC	L32179 Curated ENA , GenBank , DDBJ
RefSeq	NM_001086 Curated NCBI sequence viewer	CCDS	CCDS33877 Curated

Protein resources for AADAC ?

UniProt/Swiss-Prot	P22760 InterPro , PDBe , Reactome	AlphaFold	AF-P22760-F1
Enzyme Commission	3.1.1.3 Curated ExPASy , BRENDA , KEGG , IntEnz , IUBMB Enzyme Nomenclature		

Orthologs from selected species for AADAC ?

Canis familiaris	AADAC (VGNC:37404) VGNC	Equus caballus	AADAC (VGNC:56711) VGNC
Macaca mulatta	AADAC (VGNC:81172) VGNC	Mus musculus	Aadac (MGI:1915008) Curated
Pan troglodytes	AADAC (VGNC:11317) VGNC	Rattus norvegicus	Aadac (RGD:631440)

Specialist resources for AADAC ?

MEROPS [S09.991](#)

Clinical resources for AADAC ?

OMIM	600338	GenCC	HGNC:17
MedlinePlus	Search via AADAC	DECIPHER	Search via AADAC
ClinGen	Search via HGNC:17	Genetic Testing Registry	Search via NCBI Gene ID 13
ClinVar	Search via NCBI Gene ID 13	dbVar	Search via NCBI Gene ID 13

Other resources for AADAC ?

AmiGO	Search via P22760	QuickGO	Search via P22760
BioGPS	Search via NCBI Gene ID 13	GeneCards	Search via HGNC:17
Monarch	Search via HGNC:17	WikiGenes	Search via NCBI Gene ID 13

References for AADAC ?

Radiation hybrid mapping of human arylacetamide deacetylase (AADAC) locus to chromosome 3.
 Yamazaki K *et al.* *Genomics* 1997 Sep;44(2):248-250
 PMID: 9299245 [Europe PMC](#), [Pubmed](#)

Human liver arylacetamide deacetylase. Molecular cloning of a novel esterase involved in the metabolic activation of arylamine carcinogens with high sequence similarity to hormone-sensitive lipase.
 Probst MR *et al.* *J Biol Chem* 1994 Aug;269(34):21650-21656
 PMID: 8063807 [Europe PMC](#), [Pubmed](#)

Figure 1. An example Symbol Report from www.genenames.org. New links to MANE Select, AlphaFold and GenCC are highlighted in red. The symbol *AADAC* has been marked by HGNC curators with the Stable symbol tag, meaning that the HGNC considers it unlikely that this symbol will ever need to change. The HGNC data box at the top of the Report presents HGNC-curated nomenclature, HGNC ID, chromosomal location and membership of gene groups. The rest of the Symbol Report displays links to biomedical resources, orthology resources and key references.

HGNC approved symbols, previous, withdrawn and alias symbols in our database. We encourage authors to use this tool to check the gene nomenclature within their papers prior to submitting to journals. We provide several different options for downloading HGNC data, including prepared download files, a BioMart server, a Custom Downloads tool and a REST service. We also have a separate blog site (<https://blog.genenames.org/>) where we publish our regular newsletters, discussion points about gene nomenclature, guest posts from our collaborators and guides on how to use different aspects of our site.

The HGNC database is freely available to all via the web without the need to register or login and is accessible and legible on phone and tablet screens. All data is available for download in both TXT and JSON format.

Here we describe the changes made to the HGNC resource since our last report in 2021 (18). Note that the 2021 report included our VGNC resource; we will not describe this resource here but plan to publish a separate report in 2024 that will describe the VGNC site (<https://vertebrate.genenames.org/>) in full.

HGNC DATA

New gene entries in genenames.org

We have assigned approved nomenclature for 1028 previously unnamed genes within the last two years, including 870 new lncRNA gene entries and 136 new pseudogene entries. This brings our total count of gene entries to 43 170 as of 10 August 2022. We have increased the number of protein coding genes slightly from 19 211 to 19 229 since September 2020 because the naming of currently known protein-coding genes is essentially complete. The new additions include those added after discovery by research groups such as *H2BNI* (H2B.N variant histone 1; HGNC:56200) (19) and *LCE7A* (late cornified envelope 7A; HGNC:55921) (20) and those added by gene annotation groups but not yet studied elsewhere such as *AQP7B* (aquaporin 7B; HGNC:53895) and *CPHXL2* (cytoplasmic polyadenylated homeobox like 2; HGNC:55919).

The majority of lncRNA genes were named using our systematic naming protocol (described in full in (21)); 136 lncRNA genes were assigned the LINC (for long intergenic non-protein coding RNA) root symbol, e.g. *LINC02942* (HGNC:55957); 340 were named as antisense to protein-coding genes with the symbol suffix -AS e.g. *BICDI-ASI* (HGNC:55475); 346 were named as being divergent transcripts of (i.e. sharing a bidirectional promoter with) protein-coding genes with the symbol suffix -DT, e.g. *ACAD9-DT* (HGNC:56086); 12 were named as host genes for either microRNA or small nucleolar RNA genes, e.g. *MIR142HG* (HGNC:55980). Like protein-coding genes, lncRNA genes can be approved with unique symbols based on those published or suggested by research groups; examples of lncRNA genes named based on publications within the last 2 years include *CPMER* (cytoplasmic mesoderm regulator; HGNC:55992) (22) and *SCIRT* (stem cell inhibitory RNA transcript; HGNC:55341) (23).

Pseudogenes can be the result of duplicated genome sequence (also known as unprocessed pseudogenes), retrotransposed from mRNA (also known as processed pseu-

dogenes) or can have degraded in situ. Where human pseudogenes have degraded in situ there are often species with functional copies at a conserved genomic location (also known as unitary pseudogenes). One such example that the HGNC named within the last 2 years is *FSIP2LP* (HGNC:55625); this was named relative to the VGNC protein coding orthologs *FSIP2L* (in horse (VGNC:109005), pig (VGNC:109145), cat (VGNC:109211) and dog (VGNC:109285)), and *Fsip2l* (MGI:2685441) in mouse. Examples of duplicated pseudogenes that we named include *GOLGA6EP* (HGNC:55708) and *GOLGA6FP* (HGNC:49206) that are present on a cluster with other *GOLGA6* protein coding genes. Processed pseudogenes are usually named relative to their parent gene; recently named examples include *MTCHIPI* (HGNC:55886), *WEE1PI* (HGNC:56216) and *RACK1P2* (HGNC:55464).

Marking genes as stable

As outlined in our current guidelines (24) ‘the stability of gene symbols, particularly those associated with disease, is now a key priority’. The HGNC has recently joined the Gene Curation Coalition (GenCC) (25), a project that brings together multiple groups that are either directly involved in gene-disease curation or promote standards needed to support this curation. The HGNC is contributing to this project by reviewing and evaluating the symbols of clinically relevant genes in the GenCC database. The review process primarily checks that approved symbols are not misleading and cannot be considered as pejorative or offensive. Secondary to this, curators consider approved symbol versus alias symbol usage, and whether the approved symbol causes problems for literature searching or data processing, although these factors are judged on a case-by-case basis and do not always trigger a symbol change. If changes are necessary, curators contact research groups, clinical groups and patient groups if available to discuss the suitability of proposed changes. Where curators conclude that symbols are extremely unlikely to ever need to be changed, the Symbol Reports are marked with our ‘Stable tag’ (as shown at the top of the *AADAC* Symbol Report in Figure 1). As of August 2022, the HGNC has applied the ‘Stable tag’ to 2964 genes, an increase of 48% since our previous database publication (18).

Renaming pre-existing gene entries

Although the HGNC is committed to keeping gene symbols as stable as possible, we do make changes to symbols if necessary. When we make such changes the HGNC ID stays the same as shown in the examples below. Therefore, it is essential for databases to track current approved symbols using HGNC IDs. One of our current aims is to replace placeholder nomenclature with informative gene nomenclature once suitable information is available. Our largest group of placeholder symbols are of the format ‘C%orf#’ where % represents the chromosome that the gene is located on, orf stands for ‘open reading frame’ and # is an iterative number. We have updated 43 genes with this symbol format since our last report. Two of these updates were based on a collaboration between the HGNC and experts on dyneins and their assembly factors which has been

published (26)—HGNC:25081 was updated from *C16orf71* to *DNAAF8* (dynein axonemal assembly factor 8) and HGNC:17721 was updated from *C20orf194* to *DNAAF9* (dynein axonemal assembly factor 9). Other examples of C%orf# renames based on publications are the renaming of HGNC:21702 from *C7orf26* to *INTS15* (integrator complex subunit 15) (27) and the renaming of HGNC:28628 from *C12orf45* to *NOPCHAP1* (NOP protein chaperone 1) (28). In some cases, C%orf# symbols are renamed following annotation updates that result in a change of the gene locus type away from protein coding, either to long non-coding RNA or to pseudogene, e.g. HGNC:33774 has been renamed from *C8orf86* to *LINC03042* (long intergenic non-protein coding RNA 3042) and HGNC:21620 has been renamed from *C6orf201* to *TEX56P* (testis expressed 56, pseudogene).

The FAM# root (for ‘family with sequence similarity’) is another class of placeholder symbols and is used to group together sets of paralogous genes for which no other information was known at the time of naming. We have renamed 27 genes with this root in the last 2 years, including the FAM189 root symbol where all members have been renamed together with the ENTREP# (endosomal transmembrane epsin interactor) root symbol (HGNC:24820 is now *ENTREP1* instead of *FAM189A2*, HGNC:29075 is now *ENTREP2* instead of *FAM189A1* and HGNC:1233 is now *ENTREP3* instead of *FAM189B*) based on data from a publication about HGNC:24820 (29) and further discussions with the research group to decide on appropriate nomenclature for all family members. Further examples include HGNC:24587 and HGNC:28593, which have been renamed from *FAM126B* and *FAM126B* to *HYCC1* (hyccin PI4KA lipid kinase complex subunit 1) and *HYCC2* (hyccin PI4KA lipid kinase complex subunit 2), and HGNC:33877 and HGNC:30701, previously *FAM155A* and *FAM155B*, which have been renamed as *NALF1* (NALCN channel auxiliary factor 1) and *NALF2* (NALCN channel auxiliary factor 2).

KIAA# symbols are our third placeholder symbol set and were approved for genes identified by the Kazusa cDNA sequencing project. We have updated the symbols for 5 KIAA#s since our last report, including HGNC:26953 and HGNC:28960, that were renamed together from *KIAA1109* and *KIAA0100* to *BLTP1* and *BLTP2* (for bridge-like lipid transfer protein family member 1 and 2).

New gene groups

The HGNC regularly names genes together with a common root symbol based on a shared characteristic, such as function, homology, or encoded protein structure/domains. We host gene group reports that represent both these shared root symbols and many other types of gene groupings such as membership of protein complexes, as well as gene families that are not named with a common root symbol. Since our last report we have added over 100 new manually curated gene groups to genenames.org. Examples of individual gene groups include the Transcription factor AP-2 family, PARN exonuclease family, Adducin family, CREC family and Mitochondrial translation release factor family. Related gene groups are curated into hierarchies that

support browsing between groups. We have recently curated further methyltransferase gene groups, such as Homocysteine methyltransferases and Seven-beta-strand methyltransferase motif containing, and placed these into our Methyltransferases gene families group hierarchy (Figure 2) with help from our specialist advisor. We also worked on our Nuclear hormone receptors gene group hierarchy to make the subgroups consistent with the IUPHAR/BPS Guide to Pharmacology (30) nuclear hormone receptor pages.

HGNC WEBSITE UPDATES

Improved search facility

In April 2021, we launched a new, improved version of the search on genenames.org. This search now works across the entire website to include Symbol Reports, Gene group reports, announcements, articles and the separate blog.genenames.org site. One major improvement is the addition of an auto-suggest feature for several of our search categories, including approved symbols, previous symbols, aliases, gene names and group names. If there are many suggestions for a particular input term, the search auto-suggests five different matches for each category followed by a link that will show all results matching the term; for example, for the input ‘pla’ the search auto-suggests *PLA1A*, *PLA2G10*, *PLA2G12A*, *PLA2G12API* and *PLA2G12AP2* gene symbol suggestions, and then has the text ‘See all 71 gene symbols containing pla’ (Figure 3) which returns all of the results. The search no longer requires the inclusion of wildcards (*) so that the search term ‘cadherin’ now returns genes with this term in the gene name, alias name and previous name, and also returns the gene group page ‘Cadherins’. As many of our gene group names are plural, in the previous search there was a risk that they could be missed. The new search also accepts all major variant spellings between UK and American English, such as ‘signalling’ and ‘signaling’. The results of the search may now be downloaded in either TXT or JSON by clicking on the file icons in the ‘Download all results’ section below the search filters.

New download features

Following several requests by our users, we now provide download files available in our HGNC archive (<http://ftp.ebi.ac.uk/pub/databases/genenames/hgnc/archive/>) in TXT or JSON format that show a list of symbol changes for the span of a month or a quarter of a year. The files contain the HGNC ID, approved symbol, previous symbol, gene name, locus type, and the date of the symbol change. The ‘symbol-changes’ files are available from 1 September 2021 onwards (the first file features changes made in August 2021) and accompany our ‘HGNC complete set’ and ‘withdrawn’ download files in the HGNC archive. Additionally, we have implemented a new REST API that will allow querying both HGNC and VGNC data.

Another new download feature is the provision of a Globus endpoint for HGNC data available from our Statistics and Download files webpage (<https://www.genenames.org/download/statistics-and-files/>). All HGNC download files can be downloaded from Globus. We recommend this

Gene group: Methyltransferase families ?

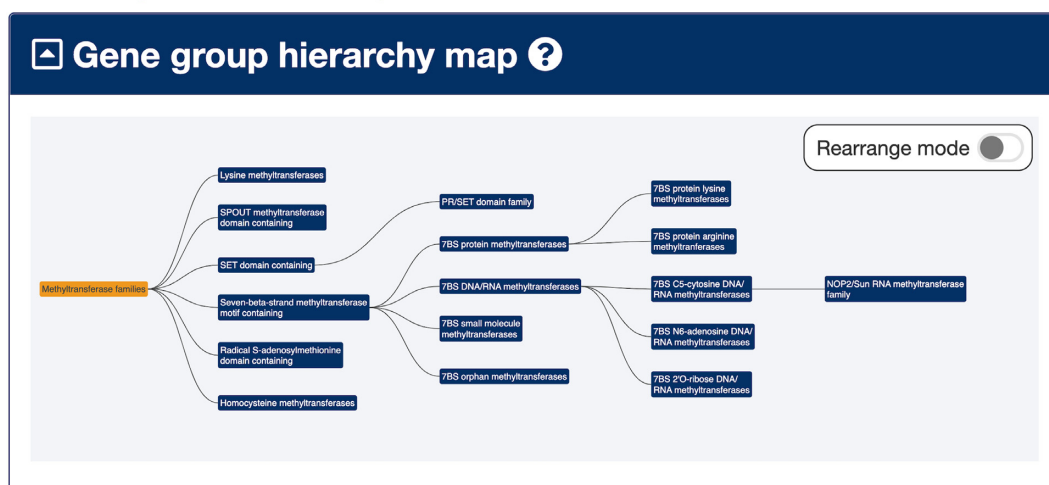


Figure 2. Gene group hierarchy for HGNC curated methyltransferase families. Users can navigate between gene groups by clicking on the gene group boxes.



Figure 3. The new auto-suggest feature on the main search at www.genenames.org. The auto-suggest box is scrollable—this query also provides alias symbol suggestions, gene name suggestions and group name suggestions.

method for users that need to download more than one file at a time, as it is rapid and is not affected by network glitches that may corrupt files during transfer.

New links to other resources

For relevant genes we have added links to the GenCC project, described in the ‘Marking symbols as stable’ section above, to the Clinical Resources section of our Symbol Reports (Figure 1). And for all genes with a UniProtKB/Swiss-Prot (reviewed) protein link, we have also added links to view protein structures predicted by the Artificial Intelligence algorithm AlphaFold2 at the AlphaFold Protein Structure Database (31) from the Pro-

tein Resources section of our Symbol Reports (Figure 1). New links have also been provided from the Nucleotide Resources section of our Symbol Reports to MANE (Matched Annotation from NCBI and EMBL-EBI) Select transcript sequence records. The MANE project (32) aims to provide a set of standard transcripts for human protein-coding genes annotated by both the RefSeq and Ensembl/GENCODE projects, to aid in transcript variant reporting by the clinical and research community. MANE transcripts list versioned RefSeq and Ensembl IDs, i.e. our *AADAC* Symbol Report displays both the RefSeq ID NM_001086.3 and the Ensembl ID ENST00000232892.12 (Figure 1). GenCC, AlphaFold and MANE links are all available to download via our REST service.

FUTURE PLANS

We will continue reviewing the nomenclature of clinically relevant genes and increase the number of Symbol Reports marked with the Stable tag. We will also continue to approve informative symbols and names for published long non-coding RNA genes and for newly identified human protein-coding genes, such as those that may be annotated in future based on Ribo-Seq data (33). We will continue to name annotated pseudogenes and long non-coding RNA genes following our systematic protocols. We will explore automating the naming of a subset of long non-coding RNA genes that follow a set of defined annotation rules. We will continue to update placeholder symbols of protein-coding genes with informative nomenclature whenever possible.

In addition to manually curating new gene groups, we plan to review our gene group resource, including how we connect gene groups within hierarchies, and will explore marking gene groups with types, such as ‘shared homology’ and ‘shared function’. For example, the *AADAC* Symbol Report in Figure 1 shows *AADAC* as a member of two different gene groups: ‘Lipases’, which is a functional grouping based on enzymatic activity, and ‘Arylacetamide deacetylase family’, which represents a group of paralogs. Applying types to these groups would be informative and would allow users to filter or search groups by type.

We will also continue to explore the creation of an InterMine service (34) for querying and downloading HGNC and VGNC data, which would have the advantage of being interoperable with other InterMine tools but would include more HGNC (and VGNC)-specific data than the currently available HumanMine, e.g. our stable human symbol set and gene group data would be incorporated.

DATA AVAILABILITY

HGNC services are freely available at <https://www.genenames.org/>. HGNC code is available at the GitHub repository (<https://github.com/HGNC>).

ACKNOWLEDGEMENTS

The authors would like to thank all HGNC and VGNC alumni for their contributions to this project. We would also like to thank all of the researchers and specialist advisors who have helped the HGNC with gene nomenclature and gene groupings, particularly Pål Falnes for his help with the Methyltransferase gene group membership and hierarchy structure shown in Figure 2. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

FUNDING

Wellcome Trust [208349/Z/17/Z]; National Human Genome Research Institute of the National Institutes of Health [U24HG003345]. Funding for open access charge: Wellcome Trust [208349/Z/17/Z].

Conflict of interest statement. None declared.

REFERENCES

- Hancock, J.M., Game, A., Ponting, C.P. and Goble, C.A. (2016) An open and transparent process to select ELIXIR node services as implemented by ELIXIR-UK. *FI1000Res*, **5**, 2894.
- Sansone, S.A., McQuilton, P., Rocca-Serra, P., Gonzalez-Beltran, A., Izzo, M., Lister, A.L., Thurston, M. and Community, F. (2019) FAIRsharing as a community approach to standards, repositories and policies. *Nat. Biotechnol.*, **37**, 358–367.
- Braschi, B., Seal, R.L., Tweedie, S., Jones, T.E.M. and Bruford, E.A. (2021) The risks of using unapproved gene symbols. *Am. J. Hum. Genet.*, **108**, 1813–1816.
- Seal, R.L., Tweedie, S. and Bruford, E.A. (2022) A standardised nomenclature for long non-coding RNAs. *IUBMB Life*, <https://doi.org/10.1002/iub.2663>.
- Sayers, E.W., Beck, J., Bolton, E.E., Bourexis, D., Brister, J.R., Canese, K., Comeau, D.C., Funk, K., Kim, S., Klimke, W. *et al.* (2021) Database resources of the national center for biotechnology information. *Nucleic Acids Res.*, **49**, D10–D17.
- Cunningham, F., Allen, J.E., Allen, J., Alvarez-Jarreta, J., Amode, M.R., Armean, I.M., Austine-Orimoloye, O., Azov, A.G., Barnes, I., Bennett, R. *et al.* (2022) Ensembl 2022. *Nucleic Acids Res.*, **50**, D988–D995.
- Consortium, UniProt (2021) UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.*, **49**, D480–D489.
- Stelzer, G., Rosen, N., Plaschkes, I., Zimmerman, S., Twik, M., Fishilevich, S., Stein, T.I., Nudel, R., Lieder, L., Mazor, Y. *et al.* (2016) The genecards suite: from gene data mining to disease genome sequence analyses. *Curr. Protoc. Bioinformatics*, **54**, 1.30.1–1.30.33.
- Consortium, RNAcentral (2021) RNAcentral 2021: secondary structure integration, improved sequence search and new member databases. *Nucleic Acids Res.*, **49**, D212–D220.
- Lee, B.T., Barber, G.P., Benet-Pagès, A., Casper, J., Clawson, H., Diekhans, M., Fischer, C., Gonzalez, J.N., Hinrichs, A.S., Lee, C.M. *et al.* (2022) The UCSC genome browser database: 2022 update. *Nucleic Acids Res.*, **50**, D1115–D1122.
- Amberger, J.S., Bocchini, C.A., Scott, A.F. and Hamosh, A. (2019) OMIM.org: leveraging knowledge across phenotype-gene relationships. *Nucleic Acids Res.*, **47**, D1038–D1043.
- Rehm, H.L., Berg, J.S., Brooks, L.D., Bustamante, C.D., Evans, J.P., Landrum, M.J., Ledbetter, D.H., Maglott, D.R., Martin, C.L., Nussbaum, R.L. *et al.* (2015) ClinGen—the clinical genome resource. *N. Engl. J. Med.*, **372**, 2235–2242.
- Pagon, R.A. (2006) GeneTests: an online genetic information resource for health care providers. *J. Med. Libr. Assoc.*, **94**, 343–348.
- Shefchek, K.A., Harris, N.L., Gargano, M., Matentzoglou, N., Unni, D., Brush, M., Keith, D., Conlin, T., Vasilevsky, N., Zhang, X.A. *et al.* (2020) The monarch initiative in 2019: an integrative data and analytic platform connecting phenotypes to genotypes across species. *Nucleic Acids Res.*, **48**, D704–D715.
- Blake, J.A., Baldarelli, R., Kadin, J.A., Richardson, J.E., Smith, C.L., Bult, C.J. and Group, M.G.D. (2021) Mouse genome database (MGD): knowledgebase for mouse-human comparative biology. *Nucleic Acids Res.*, **49**, D981–D987.
- Smith, J.R., Hayman, G.T., Wang, S.J., Lauderkind, S.J.F., Hoffman, M.J., Kaldunski, M.L., Tutaj, M., Thota, J., Nalabolu, H.S., Ellanki, S.L.R. *et al.* (2020) The year of the rat: the rat genome database at 20: a multi-species knowledgebase and analysis platform. *Nucleic Acids Res.*, **48**, D731–D742.
- Yates, B., Gray, K.A., Jones, T.E.M. and Bruford, E.A. (2021) Updates to HCOB: the HGNC comparison of orthology predictions tool. *Brief Bioinform.*, **22**, bbab155.
- Tweedie, S., Braschi, B., Gray, K., Jones, T.E.M., Seal, R.L., Yates, B. and Bruford, E.A. (2021) Genenames.org: the HGNC and VGNC resources in 2021. *Nucleic Acids Res.*, **49**, D939–D946.
- Raman, P., Rominger, M.C., Young, J.M., Molaro, A., Tsukiyama, T. and Malik, H.S. (2022) Novel classes and evolutionary turnover of histone H2B variants in the mammalian germline. *Mol. Biol. Evol.*, **39**, msac019.
- Holthaus, K.B., Lachner, J., Ebner, B., Tschachler, E. and Eckhart, L. (2021) Gene duplications and gene loss in the epidermal differentiation complex during the evolutionary land-to-water transition of cetaceans. *Sci. Rep.*, **11**, 12334.

21. Seal,R.L., Chen,L.L., Griffiths-Jones,S., Lowe,T.M., Mathews,M.B., O'Reilly,D., Pierce,A.J., Stadler,P.F., Ulitsky,I., Wolin,S.L. *et al.* (2020) A guide to naming human non-coding RNA genes. *EMBO J.*, **39**, e103777.
22. Lyu,Y., Jia,W., Wu,Y., Zhao,X., Xia,Y., Guo,X. and Kang,J. (2022) Cpmr: a new conserved eEF1A2-binding partner that regulates eomes translation and cardiomyocyte differentiation. *Stem Cell Rep.*, **17**, 1154–1169.
23. Zagorac,S., de Giorgio,A., Dabrowska,A., Kalisz,M., Casas-Vila,N., Cathcart,P., Yiu,A., Ottaviani,S., Degani,N., Lombardo,Y. *et al.* (2021) SCIRT lncRNA restrains tumorigenesis by opposing transcriptional programs of tumor-initiating cells. *Cancer Res.*, **81**, 580–593.
24. Bruford,E.A., Braschi,B., Denny,P., Jones,T.E.M., Seal,R.L. and Tweedie,S. (2020) Guidelines for human gene nomenclature. *Nat. Genet.*, **52**, 754–758.
25. DiStefano,M.T., Goehringer,S., Babb,L., Alkuraya,F.S., Amberger,J., Amin,M., Austin-Tse,C., Balzotti,M., Berg,J.S., Birney,E. *et al.* (2022) The gene curation coalition: a global effort to harmonize gene-disease evidence resources. *Genet. Med.*, **24**, 1732–1742.
26. Braschi,B., Omran,H., Witman,G.B., Pazour,G.J., Pfister,K.K., Bruford,E.A. and King,S.M. (2022) Consensus nomenclature for dyneins and associated assembly factors. *J. Cell Biol.*, **221**, e202109014.
27. Drew,K., Wallingford,J.B. and Marcotte,E.M. (2021) hu.MAP 2.0: integration of over 15,000 proteomic experiments builds a global compendium of human multiprotein assemblies. *Mol. Syst. Biol.*, **17**, e10016.
28. Abel,Y., Paiva,A.C.F., Bizarro,J., Chagot,M.E., Santo,P.E., Robert,M.C., Quinternet,M., Vandermoere,F., Sousa,P.M.F., Fort,P. *et al.* (2021) NOPCHAP1 is a PAQosome cofactor that helps loading NOP58 on RUVBL1/2 during box C/D snoRNP biogenesis. *Nucleic Acids Res.*, **49**, 1094–1113.
29. Tsunoda,T., Riku,M., Yamada,N., Tsuchiya,H., Tomita,T., Suzuki,M., Kizuki,M., Inoko,A., Ito,H., Murotani,K. *et al.* (2022) ENTREP/FAM189A2 encodes a new ITCH ubiquitin ligase activator that is downregulated in breast cancer. *EMBO Rep.*, **23**, e51182.
30. Harding,S.D., Armstrong,J.F., Faccenda,E., Southan,C., Alexander,S.P.H., Davenport,A.P., Pawson,A.J., Spedding,M., Davies,J.A. and NC-IUPHAR (2022) The IUPHAR/BPS guide to PHARMACOLOGY in 2022: curating pharmacology for COVID-19, malaria and antibacterials. *Nucleic Acids Res.*, **50**, D1282–D1294.
31. Varadi,M., Anyango,S., Deshpande,M., Nair,S., Natassia,C., Yordanova,G., Yuan,D., Stroe,O., Wood,G., Laydon,A. *et al.* (2022) AlphaFold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.*, **50**, D439–D444.
32. Morales,J., Pujar,S., Loveland,J.E., Astashyn,A., Bennett,R., Berry,A., Cox,E., Davidson,C., Ermolaeva,O., Farrell,C.M. *et al.* (2022) A joint NCBI and EMBL-EBI transcript set for clinical genomics and research. *Nature*, **604**, 310–315.
33. Mudge,J.M., Ruiz-Orera,J., Prensner,J.R., Brunet,M.A., Calvet,F., Jungreis,I., Gonzalez,J.M., Magrane,M., Martinez,T.F., Schulz,J.F. *et al.* (2022) Standardized annotation of translated open reading frames. *Nat. Biotechnol.*, **40**, 994–999.
34. Lyne,R., Sullivan,J., Butano,D., Contrino,S., Heimbach,J., Hu,F., Kalderimis,A., Lyne,M., Smith,R.N., Štěpán,R. *et al.* (2015) Cross-organism analysis using intermine. *Genesis*, **53**, 547–560.