



HGNC Newsletter Spring 2015

[<Previous Issue](#)

There are currently 39,249 approved symbols

[Next Issue>](#)



In this newsletter

[New Gene Families released](#)

[Improved search released](#)

[Updates to download files](#)

[Proposed change to our Custom Downloads tool](#)

[Gene Symbols in the News](#)

[Meeting News](#)

[Publications](#)

New Gene Families released

Thanks to all who tested the beta version of our Gene Families that was described in our [previous newsletter](#). Our improved Gene Families have now been fully released, and all families are browsable using our [Gene Families index](#). The index is sorted alphabetically according to family name and users can alter the number of families listed per page, and filter the results based on either Gene family name or Root symbol. Underneath the index we now display the total number of families and the total number of genes that are included within any family. Note that one gene can appear in more than one family.

One of the greatest improvements is that our Gene Family data are now included in our main [Search tool](#), see below for details on how the Search works. Users can also access gene families from individual Symbol Reports, e.g. see the Report for [CELSR1](#), which is a member of both the "[Major Cadherins](#)" and "[Adhesion G protein-coupled receptors, subfamily C](#)". A file containing the complete HGNC Gene Family dataset can be accessed from our [Statistics and Downloads page](#).

Our new gene family pages appear in a standardised format and all have a unique gene family name and ID; the ID is contained in the URL for the gene family page (e.g. the major cadherin family page is <http://www.genenames.org/cgi-bin/genefamilies/set/18>, the family ID is 18). Where applicable the family root symbol and gene family synonyms are displayed. Many families contain a short text description which is often from Wikipedia, UniProt or written by HGNC curators - the source of the description is always provided. If families are part of larger hierarchies, such as [G protein-coupled receptors](#), a browsable hierarchy map is displayed along with text links to family subsets and larger family groupings. If the family is not part of a hierarchy or is the lowest level of a hierarchy, such as [Melanocortin receptors](#), then a table with links to Symbol Reports for the family members is displayed. If the family is part of a high level hierarchy, the number of genes contained within all subsets is displayed and a table showing all genes within the subsets can now be displayed in HTML by clicking on the plus symbol provided. All gene family pages include the ability to download the member genes in a TSV file. The custom downloads tool now has our "Gene family name" and "Gene family ID" as selectable columns. These columns replace the old gene family columns but it is still possible to access our old gene families by selecting "Old gene families archive" from the dropdown menu under the Gene Families tab. For a full description of our families, please read our [Gene families help](#).

Improved search released

In our previous newsletter we described our plans to release an improved Search facility, which is now fully public. The Search box found in the header of every [genenames.org](#) page now features a dropdown menu that allows users to "Search everything" across our site, or choose to perform a "Symbol search", a "Gene family search" or a "Site search". The default option is "Search everything" which performs a full-text search over all the HGNC indexes and fields. Reports matching the keyword/ID are displayed in order of relevance. The Search is faceted and on the left-hand side of the results page the filter options (where available) are shown with the numbers of reports associated with each facet. If the results include gene symbols then clicking on the facet "Gene" will filter the results by this type and will change the faceting to display the locus groups and types that are relevant to the search results, enabling further filtering by locus type. Users can also change the default number of results displayed per page from 10 up to 200.

The results display specific fields from within the search index which differ depending on the document type. The first line of each result contains the **gene symbol** and the **gene name** if the result is a gene symbol report, the **family name** if it's a gene family report, or the **page title** if it hits any other page within the site. The second row shows the "Document type" (i.e gene, family or site) if searching everything, and will also contain some of the important fields to help identify the hit. The third row reports the field the keyword/ID matches, e.g. when searching with the gene symbol "ZNF3" this field would contain

Matches: Gene symbol: **ZNF3**", see figure below. For a full description of our Search, including how to perform advanced searches, please see our [Search help](#).

Search

Document type

Gene 1
Family 1
Site 1

Items per page

10 20 50 100 200

Results: 1 to 3 of 3

ZNF3: zinc finger protein 3

Document type: Gene HGNC_ID: HGNC:13089 Locus group: protein-coding gene
Matches: Gene symbol: **ZNF3**

REST service help

Document type: Site URL: <http://www.genenames.org/help/rest-web-service-help>
Matches: Page content: ... /fetch/symbol/**ZNF3** Fetch will not accept wildcards or multiple queries and will only accept a field/term ...

Zinc fingers, C2H2-type

Document type: Family HGNC Family ID: 28 Root symbol: ZNF
Matches: Gene symbol: **ZNF3**



Updates to download files

The downloadable files provided within our [Statistics & Downloads page](#) have also been improved. The files now contain all the information displayed in Symbol Reports instead of only the HGNC core data. The files are still available as TSV text but we now also provide our data as JSON (JavaScript Object Notation) files.

The old file format can still be found on our [FTP site](#) and will continue to be updated in the short term. If you use the old file format please be aware that these files will report on the previous version of our gene families data and are not comparable to the new families. See the [statistics & downloads help](#) for more information.

Proposed change to our Custom Downloads tool

We are proposing a simplification of our "Custom Downloads" tool by bringing the data provided in line with that displayed in our Symbol reports. Currently users of our Custom Downloads tool can choose to download two separate fields for some database IDs and accessions, where one contains "HGNC curated" data and the other contains "mapped data" which is provided from external resources. However, we have found that this confuses users, who complain about missing accessions and IDs within the downloaded data when they compare it to the Symbol reports, as it is not clear that both the "HGNC curated" and "mapped data" fields need to be downloaded. In our Symbol reports only one ID is shown, and the HGNC curated data, where available, always takes precedence over the mapped data. Our new [REST service](#) also only provides the coalesced fields, reflecting the data displayed in the Symbol reports. Please use our [feedback form](#) to comment and let us know if this proposed update to our Custom Downloads may cause you any problems.

Gene symbols in the news

There has been good news in the field of gene therapy this month - [researchers have managed to successfully control the disease of six male patients with Wiskott-Aldrich syndrome](#) by modifying the **WAS** gene in CD34⁺ cells with a lentiviral vector. The therapy was a success for six out of seven patients.

In other news, a recent study has explored genetic reasons as to why [some individuals retain better cognitive function than others as they age](#). Researchers found that 45 year olds carrying a specific variant of the **CADM2** gene performed worse than those of the same age who did not have the variant. A separate study found that smokers with a particular variant of the **CHRNA5** gene, which encodes a nicotinic cholinergic receptor, have been shown [to have more difficulty quitting smoking](#) than the general population and are also more likely to develop lung cancer. Luckily these individuals respond well to nicotine replacement therapies.

Meeting News

Elspeth recently attended the NC-IUPHAR Meeting in Edinburgh, Scotland from 17-19 April. She enjoyed meeting the committee members and gave a talk describing the work of the HGNC. Susan is attending the [8th International Biocuration Conference](#) in Beijing, China from 23-26 April where she will be spreading the good news about our new gene families to other biocurators in the form of a poster. Ruth will be advertising the gene families to a completely different audience at the upcoming [ESHG 2015](#) meeting in Glasgow, Scotland in early June. Beth is heading to Barcelona towards the end of May to attend [Quest for Orthologs 4](#).

Publications

Hamann J, Aust G, Arac D, Engel F..Wright MW, Xu L, Langenhan T, Schioth H. **International Union of Basic and Clinical Pharmacology. XCIV. Adhesion G Protein-Coupled Receptors**. Pharmacol Rev 2015 67:338-367. PMID:[25713288](#)

at: hgnc@genenames.org

[View the HGNC Newsletter Archive](#)