



HGNC Newsletter Autumn 2012

[<Previous Issue](#)

There are currently 33600 approved symbols

[Next Issue>](#)



In this newsletter

[Wellcome Trust funding](#)

[Naming genes on alternate loci](#)

[New links to the Genetic Testing Registry](#)

[Update on renaming C\\$orf# symbols](#)

[New Gene Family Resources](#)

[Gene Symbols in the News](#)

[Meeting News](#)

Wellcome Trust funding

We are delighted to announce that we have received funding from the Wellcome Trust Biomedical Resources Grants (099129/Z/12/Z) that will support two team members for the next five years! As well as continuing to name any novel human protein coding and RNA genes and pseudogenes, and reassigning genes with uninformative identifiers, this funding will also help to support our work in expanding our remit to naming genes in other vertebrate species without a dedicated nomenclature committee. We also expect to be able to advertise a new bioinformatics post soon, so keep checking our website for further information.

Naming genes on alternate loci

The HGNC has been busy discussing the format and display of gene nomenclature for genes on alternative loci over the last few months. We have always named genes that are present on the reference assembly, which is a combined reference genome based on only a small number of anonymous donors. Now that the majority of protein coding genes on the reference assembly have been annotated and named, it has become clear that a single tiling path cannot adequately represent regions of complex human variation. The [Genome Reference Consortium \(GRC\)](#) aims to improve the human genome reference assembly by correcting errors, closing sequence gaps and providing alternative assemblies for some of the hypervariable genomic regions. The first alternate loci to be incorporated into the reference assembly were HLA (also known as MHC) haplotypes from chromosome 6p21.3, as described by the [MHC Haplotype Consortium](#). The reference sequence now represents one haplotype from the PGF cell line, instead of a region built from different individuals, and seven additional alternative haplotypes have been incorporated into the reference assembly, designated as ALT_REF_LOCI_1 through to ALT_REF_LOCI_7 by the GRC. The HGNC uses the gene symbols assigned many years ago by the [WHO Nomenclature Committee for Factors of the HLA system](#), and all HLA genes on both the reference sequence and the new alternate loci have already been named according to this system. The genes [HLA-DRB3](#), [HLA-DRB4](#), [HLA-DRB2](#), [HLA-DRB7](#) and [HLA-DRB8](#) are only found on these alternate loci and not on the reference (PGF) sequence.

In between full releases of the reference genome, the GRC generates patches so that the coordinates of the reference remain stable. A patch is a genomic sequence released as an update to the reference assembly, and can either be a fix patch which corrects an error in the assembly or a novel patch that contains alternative loci. Fix patches will be incorporated into the next full assembly release, while novel patches will be represented as alternate reference loci. Since the release of GRCh37, the GRC has incorporated novel patches for eight haplotypes in the Leukocyte Receptor Complex (LRC) region on chromosome 19q13.4 (see the [GRC Genome Update](#) on this region for more details). As for the MHC, the HGNC had previously worked with the community to approve symbols for the genes found only on the alternate LRC haplotypes: [KIR2DL2](#), [KIR2DL5A](#), [KIR2DS1](#), [KIR2DS2](#) and [KIR2DS5](#).

The HGNC is working directly with the GRC and will provide gene symbols for novel genes on patches when requested. The first request we received was to name a gene found on a novel patch that results from a deletion of most of the [APOBEC3B](#) gene, resulting in a fusion of [APOBEC3A](#) with part of the 3' UTR of [APOBEC3B](#). We decided to reserve the use of a specific character that could be used to easily distinguish any new gene symbols approved for genes on alternate loci and we chose the underscore (_) character for this purpose. As a result we have approved the symbol [APOBEC3A_B](#) with the full name "APOBEC3A and APOBEC3B deletion hybrid" for the gene on the novel patch.

Previously there was no easy way to identify all genes in the HGNC database that are only found on alternate loci but we have now provided a separate table labelled "Alternative Loci Statistics" on our [Statistics and Downloads page](#). We have also used the Chromosomal Location field in our Gene Symbol Report to clearly indicate genes on alternate loci and novel patches: after the chromosomal location band we write the name of the alternate locus as named by the GRC, or we include the GRC assembly version and the word "patch" e.g. the Chromosomal Location field for [HLA-DR3B](#) lists "6p21.3 ALT_REF_LOCI_1, 6p21.3 ALT_REF_LOCI_2 and 6p21.3 ALT_REF_LOCI_6", while the Chromosomal Location field for [APOBEC3A_B](#) states "22q13 GRCh37.p9 patch".

New links to the Genetic Testing Registry

We have recently included links from the Clinical Resources section of our Gene Symbol Report pages to the NCBI's [Genetic Testing Registry](#). This resource is a

central location for the submission of genetic test information on a voluntary basis. Genetic Testing Registry reports include a summary of the gene in question, the genomic context, a list of associated conditions and available tests, and relevant references. For an example, please see the [TP53 report](#).

Renaming genes with C\$orf# symbols

Regular readers of our newsletter will know that renaming genes with placeholder "C\$orf#" symbols is a current priority for the HGNC. In 2012 alone we have now renamed 276 C\$orf# symbols with more informative nomenclature. For example, C15orf58 has recently been renamed based on its encoded enzyme function: **GDGPG1** for "GDP-D-glucose phosphorylase 1". Several previously named C\$orf# genes have been identified and named as NADH dehydrogenase (ubiquinone) complex I assembly factors: C20orf7 is now **NDUFAF5**, C8orf38 is now **NDUFAF6** and C2orf56 is now **NDUFAF7**. Other C\$orf# genes have been renamed based on homology e.g. C1orf38 has been renamed as **THEMIS2** after it was identified as the paralog of **THEMIS**, and C3orf26 is now named in concordance with its yeast ortholog as **CMSS1** for "cms1 ribosomal small subunit homolog (yeast)".

If you have information on any genes with a C\$orf# symbol that could be used as the basis for a rename, please either email us at hgnc@genenames.org or fill out our [gene symbol request form](#).

New Gene Family Resources

New G protein-coupled receptor resource

We have a new comprehensive [G protein-coupled receptors page](#) that presents the many classes and subclasses of genes encoding this diverse family of receptors. The GPCR family is subdivided into the following main sections, and each section is further subdivided into receptor types:

[Class A GPCRs, rhodopsin-type](#)

[Class B GPCRs, secretin-type](#)

[Class C GPCRs, glutamate-type](#)

[Class F GPCRs, frizzled-type](#)

[Unclassified GPCRs](#)

New Gene Family pages

[Biogenesis of lysosomal organelles complex-1 subunits](#)

[BRICHOS domain containing](#)

[DENN/MADD domain containing](#)

[Elongator acetyltransferase complex subunits](#)

[Mitochondrial respiratory chain complex assembly factors](#)

Gene Family Tags

Currently each gene family within our database is represented by a combination of letters called a "gene family tag". The gene family tag appears in the gene family URL and can be used to download data for that particular family from our [Custom Downloads page](#). Wherever possible we have used the gene family root symbol for these tags, e.g. the gene family tag for the Acyl CoA thioesterases gene family is ACOT. However where the gene family is based on a functional grouping there is often not a gene family root symbol to use as the tag. In these cases we have had to use combinations of letters for the tag that do not overlap other gene symbols, but this limits the gene symbols that can be used in the future. To avoid this potential problem, we propose replacing these alphabetical tags with unique numerical tags. If you have any comments or questions regarding this proposal please email us at hgnc@genenames.org.

Long non-coding RNAs (lncRNAs)

There are over 1,500 long non-coding RNA entries in our database, with potentially thousands more to follow, so the [lncRNA page](#) was becoming long and unwieldy. To solve this problem the main lncRNA page now provides links to four separate pages that each represent the genomic context of the lncRNA with respect to the nearest protein-coding genes as follows: [intergenic](#), [antisense](#), [intronic](#) and [overlapping](#). The page also includes a curated table of long non-coding RNA genes that encode a transcript with published evidence of function. Wherever possible we include a link to [lncRNAdb](#), the long non-coding RNA database that provides comprehensive annotations of eukaryotic lncRNAs.

Gene Symbols in the News

There have been a number of reports in the international media featuring approved symbols over the last few months. Several of these reports concerned the associations of particular genes with different types of cancer: the **FGF1** gene was found to be [active in aggressive ovarian cancers](#), while the **DCN** gene was found to be at [lower levels in prostate cancer cells](#). A study reported that the **FOXP3** gene is at [lower than normal levels in breast cancer cells](#) and that the encoded protein regulates levels of **SATB1**, a gene that is known to promote metastasis. In other disease-gene association news, variations in or near the following genes have been [linked to an increased risk of Parkinson disease](#): **SNCA**, **MAPT**, **GAK**, **DGKQ**, and **RIT2**; a mutation in the **CIB2** gene has been shown to [cause deafness and hearing loss](#); and a variant of the **PARK2** gene has been linked to lumbar disc degeneration and resulting back pain. A gene that has previously been associated with several different psychiatric disorders has been [linked to happiness levels in women](#) - those carrying a less active **MAOA** variant are reportedly happier than those with a more active version. Finally, a study has recently found that variants of [five genes influence human face shape](#): **PRDM16**, **PAX3**, **TP63**, **C5orf50**, and **COL17A1**.

Meeting News

The HGNC presented its plans to coordinate naming genes across vertebrate species at two separate conferences last September. Elspeth and Ruth attended [Genome Informatics 2012](#) at Robinson College, Cambridge (UK) where they presented a poster on the subject, while Elspeth and Matt attended the [Livestock Genomics meeting](#) which was held on the Genome Campus, Hinxton (UK) where Matt gave a talk. In October Matt attended the RNAcentral consortium meeting at the Moller Centre in Cambridge (UK), where he presented details on how the HGNC name non-protein coding RNA genes. Elspeth will be travelling to San Francisco (USA) in November to attend [ASHG 2012](#) where she will be presenting a poster on naming human genes on alternative loci.

If you would like to be added to our HGNC Newsletter mailing list or if you have questions or comments on any human gene nomenclature issue, please email us at: hgnc@genenames.org

[View the HGNC Newsletter Archive](#)

