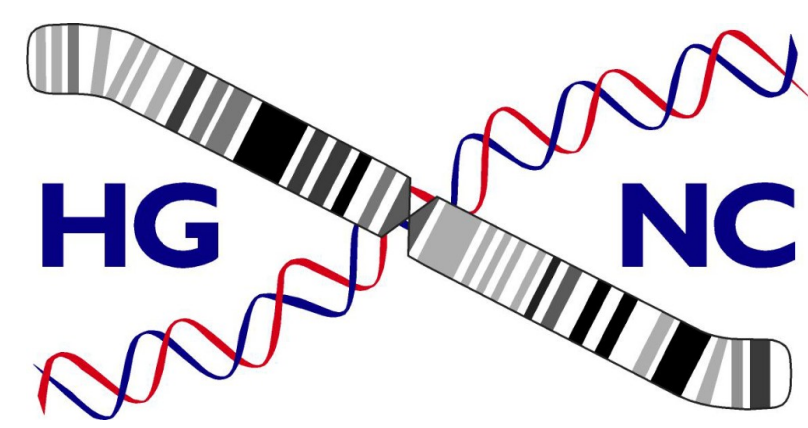


HGNC: the why and how of standardised gene nomenclature



Ruth Seal, Susan Gordon, Michael Lush, Matt Wright, Elspeth Bruford

HUGO Gene Nomenclature Committee (HGNC)

EMBL, Outstation - Hinxton, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SD, UK

Email: hgnc@genenames.org URL: <http://www.genenames.org>

The work of the HGNC is supported by the NHGRI and the Wellcome Trust



Introduction

Established in 1989, the HUGO Gene Nomenclature Committee (HGNC) is the only organisation authorised to assign approved nomenclature to human genes. Standardisation of gene symbols is necessary to allow researchers and curators to refer to the same gene without ambiguity and to find all relevant information for a particular gene. We currently have 28,000+ approved gene symbols in our database (<http://www.genenames.org>). The majority of these symbols represent protein coding genes, however these also include pseudogenes, RNA genes, genomic features and phenotypes.

Gene Symbol Assignment can be based on:

Function

enzyme, complex subunit, receptor, transporter, binding protein

ADA adenosine deaminase

TRAV16 T cell receptor alpha variable 16

GTF2H2C general transcription factor IIH, polypeptide 2C

Sequence comparison

gene family member, homolog, 'like' gene, pseudogene

MAGEA5 melanoma antigen family A, 5

MIOS missing oocyte, meiosis regulator, homolog (Drosophila)

NF1L7 neurofibromin 1-like 7

NCLP1 nucleolin pseudogene 1

Domain structures

intrinsic domains and motifs

AKD1 adenylate kinase domain containing 1

TRIM32 tripartite motif-containing 32

Other relevant information from researchers and databases

associated phenotype, cellular location, chromosomal location

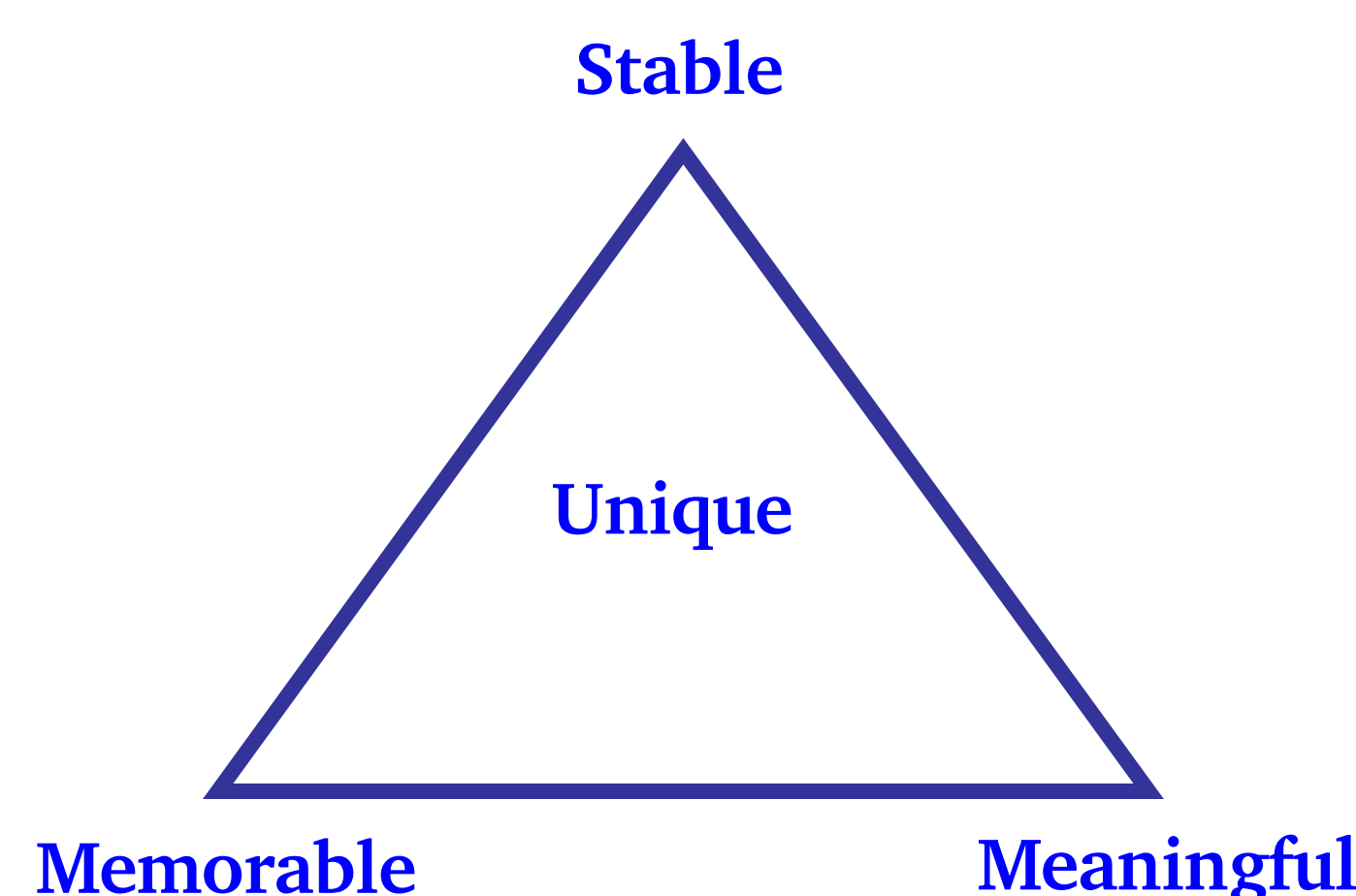
Examples:

DFNB17 deafness, autosomal recessive 17

ERP27 endoplasmic reticulum protein 27

C11orf21 chromosome 11 open reading frame 21

All gene symbols must be **UNIQUE** and should ideally have all of the following properties:



In reality, deciding on a suitable gene symbol often requires compromise.

Here are examples of scenarios that regularly arise:



Scenario #1: The published or proposed symbol or its root symbol is not unique.

Published symbol: Der1 "derlin-1" (in mouse)

Stem symbol "Der" is already in use for "delayed early response" genes in mouse

Action: coordinated with the HGNC to approve unique symbols

Conclusion: approved unique symbol **DERL1**

"Der1-like domain family, member 1"



Scenario #2: The published or proposed symbol matches a word or abbreviation that would cause problems with database searching.

Published symbol: RAIN "RAS-Interacting protein"

Number of Pubmed hits to RAIN: over 7500

Action: Contacted authors to negotiate more appropriate symbol

Conclusion: Approved author-agreed symbol **RASIP1**

"Ras interacting protein 1"



Scenario #3: A single gene has been published using numerous symbols

Published symbols: MAVS, VISA, IPS-1, Cardif

Action: as only one of these symbols, MAVS, does not contravene our guidelines, the authors were contacted to promote use of this symbol

Conclusion: approved unique symbol **MAVS**

"mitochondrial antiviral signaling protein"

CASP8: A gene symbol in HGNC and beyond

The screenshot shows the HUGO Gene Nomenclature Committee website. The main heading is "HUGO Gene Nomenclature Committee" with a tagline "Giving unique and meaningful names to every human gene". Below this, there are navigation links for "Home", "Gene Search", "Guidelines", "Gene Symbols", "Downloads", and "Home". A search bar contains "CASP8". The page displays the "Approved Symbol" as CASP8 and the "Approved Name" as caspase 8, apoptosis-related cysteine peptidase. It also lists various database links and accession numbers.

The screenshot shows the "Symbol Report: CASP8" page. It provides a detailed overview of the gene symbol, including its approved name, HGNC ID (HGNC:1509), status (Approved), chromosome (2q33-q34), and previous symbols. It also lists various database links such as RefSeq, GenBank, EMBL, and UniProt. The page is organized into sections like "Core Data", "Database Links", "Gene Symbol Links", "Specialist Database Links", and "Locus Specific Database Links".

The screenshot shows the "Caspases" page, which lists various caspase genes. The table includes columns for Approved Symbol, Approved Name, Previous Symbols, Aliases, Chromosome, Accession Numbers, and RefSeq IDs. The list includes CASP1 through CASP12, with their respective functions and database identifiers.

The screenshot shows the "HGNC Comparison of Orthology Predictions" page. It allows users to search for orthologs between human and other species. The search results table shows predicted orthologs for CASP8 in various species, including Chimp, Mouse, Rat, Dog, Chicken, Zebrafish, and Human. The table includes columns for Species, Query data, Predicted ortholog data, and Assertion derived from.

The screenshot shows two database entries for the CASP8 gene. On the left is the RGD (Rat Genome Database) entry for "Casp8 (Rattus norvegicus)", which is a member of the cysteine-aspartic acid protease family. On the right is the ZFIN (Zebrafish Information Network) entry for "Chicken : Ros.gga.6332", which is caspase 8, apoptosis-related cysteine peptidase. Both entries provide detailed information about the gene's structure, expression, and function.