

The Risks of Using Unapproved Gene Symbols

Bryony Braschi¹, Ruth L. Seal^{1,2}, Susan Tweedie¹, Tamsin E.M. Jones¹, Elspeth A. Bruford^{1,2*}

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

*contact email elspeth@genenames.org

Summary/Abstract

The use of approved nomenclature in publications is vital to enabling effective scientific communication and is particularly crucial when discussing genes of clinical relevance. Here we discuss several examples of cases where the failure of researchers to use an HGNC approved symbol in publications has led to unrelated human genes being confused with each other in the literature. We also inform authors of the steps they can take to ensure that they use approved nomenclature in their manuscripts, and discuss how referencing HGNC IDs can remove ambiguity when referring to genes that have previously been published using confusing alias symbols.

Researchers can discuss their work with the wider scientific community more easily when they use unique, approved and functionally informative nomenclature in their published papers. As genetics and genomics become increasingly widely used in healthcare, this is especially key when discussing genes of clinical relevance. Many journals already strongly recommend that authors use approved nomenclature, but in reality this is often not enforced at the point of publication. Sadly, the use of unapproved aliases can not only cause confusion and wasted experiments in the laboratory, but even more worryingly, confusion in the clinic has the potential to cause harm to patients. We would like to highlight a few examples of where the use of symbol aliases has caused significant confusion.

TFAZZIN* and *WWTR1

For our first example, until very recently the approved symbol for the gene encoding the protein “tafazzin” was approved as *TAZ* (HGNC:11577; [MIM: 300394]). This gene is of clinical interest, as variants are associated with the X-linked genetic disorder Barth syndrome¹[MIM: 302060]. The tafazzin protein plays a key role in cardiolipin remodeling and so is vital to mitochondrial function. Barth syndrome patients typically have a form of cardiomyopathy associated with an enlarged heart, along with skeletal muscle myopathy and neutropenia, which can result in a weakened immune system².

Unfortunately, the gene approved as *WWTR1* (WW domain containing transcription regulator 1) (HGNC:24042; [MIM: 607392]) has been heavily published using the alias symbol “TAZ”, alongside the name “transcriptional coactivator with PDZ-binding motif”. The

WWTR1 protein interacts with the protein encoded by *YAP1* (Yes1 associated transcriptional regulator) (HGNC:16262; [MIM: 606608]) and the two are often referred to together as “YAP/TAZ”: a pair of transcriptional co-activators of the Hippo signalling pathway³.

This situation not only makes the retrieval of relevant papers about tafazzin and the protein encoded by *WWTR1* more difficult, but can also lead to confusion between the two. For example, Takehara et al. (2018)⁴ had to be retracted by its authors. This group was studying the Hippo pathway and its role in the development of mesothelioma. They intended to use the *WWTR1* encoded protein in their experiments, but due to the common usage of the alias “TAZ”, they confused their gene of interest with tafazzin and used the incorrect protein as part of their studies.

The HGNC discussed whether we could make a change to the gene nomenclature to help avoid this type of situation recurring. Although the researchers working on the approved *TAZ* (tafazzin) gene were correctly using the approved nomenclature, we decided that a pragmatic solution to minimise further confusion could be to update the approved symbol for *TAZ* (HGNC:11577) to *TAFAZZIN* in line with the name of the encoded protein, providing that the community working on this gene were supportive of this change. The vast majority of researchers working on this gene already used the protein name “tafazzin” in their publications.

We wrote to authors who had previously published on *TAZ* (tafazzin) and fortunately, most could see the benefits of this proposal and were in favour of it. In addition, the Barth Syndrome Foundation wrote us a letter of support, reassuring us that the clinical community felt that this would be a justified symbol change to avoid confusion and that they would use the new nomenclature once updated.

A few researchers understandably felt that the “YAP/TAZ” community should just stop using the symbol “TAZ” for the *WWTR1* gene. However, unfortunately we cannot “police” the use of aliases, although we can advocate to journals and authors about how important it is to use standardized, approved nomenclature when discussing genetics.

Our March 2021 symbol update of HGNC:11577 to the approved symbol *TAFAZZIN* (tafazzin, phospholipid-lysophospholipid transacylase) should help minimise future confusion and aid data retrieval by allowing papers about the tafazzin gene to be more easily separated from those discussing the “YAP/TAZ” pathway.

MIB1* and *MKI67

Our second example came to light when we were asked to update the *MIB1* (HGNC:21086; [MIM: 608677]) gene name. *MIB1* was named after its fruit fly ortholog, *mib1* (mind bomb 1), and originally had the approved gene name “mindbomb homolog 1 (Drosophila)”. The gene was first studied in zebrafish where mutations were associated with dramatic disruption to the brain - hence the name⁵. While memorable, such names are less appropriate for human orthologs where they may be discussed in a clinical setting. As we looked at usage of the term ‘mindbomb’ in the literature, we could see that the genes *MIB1* and *MKI67* (marker of proliferation Ki-67) (HGNC:7107; [MIM: 176741]) which has the symbol alias “MIB-1” have been conflated in some publications. The *MKI67* product is associated with proliferating cells

and is used as a marker to distinguish between benign and malignant tissue. The MIB-1 alias was originally used to refer to an antibody that detected the *MKI67*-encoded marker that was named the “Molecular Immunology Borstel” antibody 1 after the institute where a senior author, Johannes Gerdes, was based^{6,7}.

Unfortunately, the “MIB-index” for measuring tumour proliferation has mistakenly been referred to as the ‘mindbomb index’ in over 70 publications; but “mindbomb” comes from the original approved gene name for *MIB1* and should never be associated with *MKI67*. In many of these papers there is no citation or ID for the gene/antibody being used and it is impossible to be sure from phrases such as “Mindbomb E3 ubiquitin protein ligase 1(MIB1) staining was minimal” which gene was being studied. This confusion has also resulted in the NCBI MeSH (Medical Subject Headings) term ‘Ki-67 antigen’ being incorrectly linked to papers about the *MIB1* gene. We have contacted the MeSH staff so that the incorrect ‘Ki-67 antigen’-paper associations are removed. In an effort to make it clear what MIB stands for in the context of the *MKI67* gene, we have now added “Molecular Immunology Borstel antibody 1” as a gene name alias. We have also updated the gene name of *MIB1* to ‘MIB E3 ubiquitin protein ligase 1’ to avoid the contentious term “mindbomb”. We hope this will reduce the usage of mindbomb in any context and reduce confusion. The approved symbol *MIB1* is so well-supported by the community that it is not feasible to change this symbol.

OXR1* and *HCRTR1

A researcher recently wrote to us (April 2021) expressing concern over the confusion in the literature caused by the symbol “*OXR1*” being published as an alias for the gene approved as *HCRTR1* (hypocretin receptor 1) (HGNC:4848; [MIM: 602392]).

The symbol *OXR1* is currently approved for the “oxidation resistance 1” gene (HGNC:15822; [MIM: 605609]), which has been associated with a human phenotype involving cerebellar hypoplasia/atrophy, epilepsy, and global developmental delay⁸[MIM: 213000]. As its encoded protein protects neurons from oxidative stress-induced apoptosis, it is also being investigated in terms of possible therapeutic effects in the treatment of amyotrophic lateral sclerosis⁹[MIM: 105400].

The unrelated genes *HCRTR1* and *HCRTR2* (HGNC:4849; [MIM: 602393]) encode proteins that act as receptors for the neuropeptide hormones known as the hypocretins, also published as “orexins”. A variant of *HCRTR2* has been associated with narcolepsy in dogs¹⁰ and humans^{10,11}, and “orexin” knockout mice also show a phenotype similar to human narcolepsy¹²[MIM: 161400].

The symbol aliases “*OX1R*” and “*OX2R*” have previously been used to refer to *HCRTR1* and *HCRTR2*^{13,14}, but the letter order of the “*OX1R*” alias symbol for *HCRTR1* was unfortunately switched to “*OXR1*” in at least 79 papers in Pubmed.

Unfortunately in this instance it did not seem helpful to change the approved *OXR1* symbol to something different, but in order to try and minimise future confusion, we wrote to all

authors who have published on *HCRTR1* using the “OXR1” alias, reiterating the importance of using approved gene nomenclature in their papers.

SF3B3* and *SAP130

A further example of confusion involving aliases and antibodies concerns the two unrelated genes currently approved as *SF3B3* (splicing factor 3b subunit 3) (HGNC:10770; [MIM 605592]) and *SAP130* (Sin3A associated protein 130) (HGNC:29813; [MIM: 609697]) which both encode proteins that are around 130kDa in weight. Instead of using the HGNC approved name for *SF3B3*, some researchers have used the alias term “spliceosome associated protein 130” in their papers and abbreviated this to “SAP130”^{15,16,17}.

The *SF3B3* gene encodes a splicing protein that activates the immune system by binding the protein product encoded by *CLEC4E* (C-type lectin domain family 4 member E) (HGNC:14555, [MIM: 609962]), which is itself also published on using the alias term “mincle” (macrophage-inducible C-type lectin receptor). This interaction may be of interest to researchers investigating possible therapies for the inflammatory bowel condition Crohn disease¹⁸ [MIM:266600]. In contrast, *SAP130* (HGNC:29813) encodes a protein that is part of the Sin3A corepressor complex, which associates with histone deacetylases¹⁹.

A 2013 publication referred to the “mincle” ligand as “SAP130, a subunit of histone deacetylase” when they were actually studying *SF3B3* and should have called this a splicing factor²⁰. This paper was then added by OMIM as a reference for the gene approved as *SAP130* and antibody companies then attached this article to products for both the *SAP130* and *SF3B3* genes.

As a result, the wrong antibody and/or primers were used by at least one group of researchers: Liu *et al.*²¹ (state that they were using the Abcam Anti-SAP130 antibody ab111739, which detects the Sin3A associated protein rather than the *SF3B3*-encoded protein that they intended to detect in their study. Metselaar *et al.*²² report in their study assessing the confusion in the literature for these two genes that they found 34 papers using “SAP130” as an alias to refer to *SF3B3*. Some of these papers did not quote antibody identification numbers, so it was impossible to trace if the protein that they intended to study was actually used in their experiments.

NRF1* and *NFE2L1

A further example is the confusion between two distinct nuclear transcription factors, *NRF1* (nuclear respiratory factor 1) (HGNC:7996; [MIM:600879]) and *NFE2L1* (nuclear factor, erythroid 2 like 1) (HGNC:7781; [MIM: 163260]), which has the alias symbol “NRF1” for the alias gene name “nuclear factor erythroid 2-related factor 1”. This confusion has been documented in the literature e.g., see “Commentary on Distinct, but Previously Confused, Nrf1 Transcription Factors and Their Functions in Redox Regulation”²³. The use of the *NRF1* symbol to refer to both genes is unfortunate as *NFE2L1* can regulate *NRF1* in coordination with TFAM (HGNC:11741; [MIM: 600438]), and both *NFE2L1* and *NRF1* are upregulated by PITX2 (HGNC:9005; [MIM: 601542]). In cases like these where there has been past confusion, we would recommend inclusion of the HGNC ID to make it absolutely clear which gene is being referred to; for example, *NRF1* (HGNC:7996) and *NFE2L1* (HGNC:7781).

POU5F1

We would also like to give a special mention to a commonly-used symbol alias for the *POU5F1* (POU class 5 homeobox 1) (HGNC:9221; [MIM: 164177]) gene, “OCT4”. We received a substantial amount of media coverage following the mention in our recent guidelines paper²⁴ of how we had changed all approved gene symbols that were being auto-converted to dates by Microsoft Excel, to avoid this problem recurring. However, persistent use of the “OCT4” alias still risks ongoing auto-conversion in spreadsheets, while using the approved symbol, *POU5F1*, will have no such issue.

Placeholder symbols

We are keen to update the nomenclature of genes that currently have C#orf, KIAA or FAM ‘placeholder’ symbols to be more functionally informative. We strongly encourage authors to contact us via our website request form or via email (hgnc@genenames.org) to discuss assigning appropriate new nomenclature before submitting their manuscripts for publication. We aim to work with authors to ensure that any new symbols introduced are unique, informative, good PubMed search terms and do not clash with existing well published aliases.

Minimising gene symbol confusion

One way that we are trying to minimise confusion between approved and alias symbols is by displaying “Curator notes” on certain gene symbol report pages on the HGNC website. We use these notes for example to highlight cases where we are aware that an approved symbol is being used as an alias for another gene, or a published alias for a gene clashes with an approved symbol, or when multiple genes share aliases.

In summary, we urge all researchers to use approved gene nomenclature instead of symbol aliases, and to utilise gene IDs to avoid ambiguity where possible. Using the approved symbol instead of what may be your “favourite” symbol is essential to reduce confusion and misreporting of data. The HGNC multi-symbol checker tool is a fast and easy way to check that you are using HGNC approved gene symbols in your manuscripts. Please do not publish new symbols for genes that are already named, especially when these clash with existing approved symbols; there are already over 450 approved gene symbols that exactly match aliases for different genes. We recommend the use of stable gene IDs, such as HGNC IDs, that are linked to the underlying sequence of the gene, and hence will not change even when changes to the nomenclature are required. Such IDs facilitate efficient data retrieval for both manual and automated searches.

Authors, reviewers and journals can help to minimise future confusion in the literature by ensuring the use of HUGO Gene Nomenclature Committee (HGNC) symbols to refer to human genes in scientific publications. For genes in non-human vertebrates we advise that symbols approved by the relevant species specific nomenclature committees (e.g. MGI for mouse) or the Vertebrate Gene Nomenclature Committee (VGNC) are used. If you are aware of any further examples of confusion caused by gene symbols and aliases please contact us (hgnc@genenames.org).

Acknowledgements

The HGNC is currently funded by Wellcome Trust grant 208349/Z/17/Z and the National Human Genome Research Institute (NHGRI) grant U24HG003345. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Declaration of interest

The authors declare no competing interests.

Web resources

HGNC, <https://www.genenames.org>

HGNC multi-symbol checker tool, <https://www.genenames.org/tools/multi-symbol-checker/>

OMIM, <https://www.omim.org>

NCBI MeSH, <https://www.ncbi.nlm.nih.gov/mesh/>

References

1. Zegallai H.M., Hatch G.M. "Barth Syndrome: Cardiolipin, Cellular Pathophysiology, Management, and Novel Therapeutic Targets." *Mol. Cell. Biochem.* 2021; 476: 1605–29.
2. Sabbah H.N. "Barth Syndrome Cardiomyopathy: Targeting the Mitochondria with Elamipretide." *Heart Fail. Rev.* 2021; 26: 237–53.
3. Sun T., Chi J.-T. "Regulation of Ferroptosis in Cancer Cells by YAP/TAZ and Hippo Pathways: The Therapeutic Implications." *Genes Dis.* 2021; 8: 241–49.
4. Takehara Y., Yamochi T., Nagumo T., Cho T., Urushibara F., Ono K., Fujii T., Okamoto N., Sasaki Y., Tazawa S. et al. "Analysis of YAP1 and TAZ Expression by Immunohistochemical Staining in Malignant Mesothelioma and Reactive Mesothelial Cells." *Oncol. Lett.* 2018; 16: 6209.
5. Schier A. F., Neuhauss S.C., Harvey M., Malicki J., Solnica-Krezel L., Stainier D.Y., Zwartkuis F., Abdelilah S., Stemple D.L., Rangini Z. et al. "Mutations Affecting the Development of the Embryonic Zebrafish Brain." *Development* 1996; 123: 165–78.
6. Cattoretti G., Becker M.H., Key G., Duchrow M., Schlüter C., Galle J., Gerdes J. "Monoclonal Antibodies against Recombinant Parts of the Ki-67 Antigen (MIB 1 and MIB 3) Detect Proliferating Cells in Microwave-Processed Formalin-Fixed Paraffin Sections." *J. Pathol.* 1992; 168: 357–63.
7. Scholzen T., Gerlach C., Cattoretti G.. "An Insider's View on How Ki-67, the Bright Beacon of Cell Proliferation, Became Very Popular. A Tribute to Johannes Gerdes (1950-2016)." *Histopathology* 2018; 73: 191–96.
8. Wang J., Rousseau J., Kim E., Ehresmann S., Cheng Y.-T., Duraine L., Zuo Z., Park Y.-J., Li-Kroeger D., Bi W. et al. "Loss of Oxidation Resistance 1, OXR1, Is Associated with an Autosomal-Recessive Neurological Disease with Cerebellar Atrophy and Lysosomal Dysfunction." *Am. J. Hum. Genet.* 2019; 105: 1237–53.

9. Finelli M.J., Liu K.X., Wu Y., Oliver P.L., Davies K.E. "Oxr1 Improves Pathogenic Cellular Features of ALS-Associated FUS and TDP-43 Mutations." *Hum. Molec. Genet.* 2015; 24: 3529–44.
10. Lin L., Faraco J., Li R., Kadotani H., Rogers W., Lin X., Qiu X., de Jong P.J., Nishino S., Mignot E. "The Sleep Disorder Canine Narcolepsy Is Caused by a Mutation in the Hypocretin (orexin) Receptor 2 Gene." *Cell* 1999; 98: 365–76.
11. Nishino S., Ripley B., Overeem S., Lammers G.J., Mignot E. "Hypocretin (orexin) Deficiency in Human Narcolepsy." *Lancet* 2000; 355: 39–40.
12. Chemelli, R. M., J. T. Willie, C. M. Sinton, J. K. Elmquist, T. Scammell, C. Lee, J. A. Richardson, et al. "Narcolepsy in Orexin Knockout Mice: Molecular Genetics of Sleep Regulation." *Cell* 1999; 98: 437–51.
13. Karhu L., Turku A., Xhaard H. "Modeling of the OX1R-Orexin-A Complex Suggests Two Alternative Binding Modes." *BMC Struct. Biol.* 2015; 15: 9.
14. Callander G.E., Olorunda M., Monna D., Schuepbach E., Langenegger D., Betschart C., Hintermann S., Behnke D., Cotesta S., Fendt M. et al. "Kinetic Properties of 'Dual' Orexin Receptor Antagonists at OX1R and OX2R Orexin Receptors." *Front. Neurosci.* 2013; 7: 230.
15. Cordero-Espinoza L., Hagen T. "Regulation of Cullin-RING Ubiquitin Ligase 1 by Spliceosome-Associated Protein 130 (SAP130)." *Biol. Open* 2013; 2: 838–44.
16. Kim J.-W., Roh Y.-S., Jeong H., Yi H.-K., Lee M.-H., Lim C.-W., Kim B. "Spliceosome-Associated Protein 130 Exacerbates Alcohol-Induced Liver Injury by Inducing NLRP3 Inflammasome-Mediated IL-1 β in Mice." *Am. J. Pathol.* 2018; 188: 967–80.
17. Zhou, H., Yu M., Zhao J., Martin B.N., Roychowdhury S., McMullen M.R., Wang E., Fox P.L., Yamasaki S., Nagy L.E. et al. "IRAKM-Mincle Axis Links Cell Death to Inflammation: Pathophysiological Implications for Chronic Alcoholic Liver Disease." *Hepatology* 2016; 64: 1978–93.
18. Gong W., Guo K., Zheng T., Fang M., Xie H., Li W., Hong Z., Ren H., Gu G., Wang Z. et al. "Preliminary Exploration of the Potential of Spliceosome-Associated Protein 130 for Predicting Disease Severity in Crohn's Disease." *Ann. N.Y. Acad. Sci.* 2020; 1462: 128–38.
19. Fleischer T.C., Yun U.J., Ayer D.E. "Identification and Characterization of Three New Components of the mSin3A Corepressor Complex." *Mol. Cell. Biol.* 2003; 23: 3456–67.
20. Suzuki Y., Nakano Y., Mishiho K., Takagi T., Tsuruma K., Nakamura M., Yoshimura S., Shimazawa M., Hara H. "Involvement of Mincle and Syk in the Changes to Innate Immunity after Ischemic Stroke." *Sci. Rep.* 2020; 3: 3177.
21. Liu K., Liu D., Feng Y., Zhang H., Zeng D., Liu Q., Qu J. 2020. "Spliceosome-Associated Protein 130: A Novel Biomarker for Idiopathic Pulmonary Fibrosis." *Ann. Transl. Med.* 2020; 8: 986.
22. Metselaar P.I., Hos C., Welting O., Bosch J.A., Kraneveld A.D., de Jonge W.J., Te Velde A.A. "Ambiguity about Splicing Factor 3b Subunit 3 (SF3B3) and Sin3A Associated Protein 130 (SAP130)." *Cells* 2021; 10: 590.

23. Zhu Y.P., Xiang Y., L'honoré A., Montarras D., Buckingham M., Zhang Y. "Commentary on Distinct, but Previously Confused, Nrf1 Transcription Factors and Their Functions in Redox Regulation." *Dev. Cell* 2020; 53: 377-378.
24. Bruford E.A., Braschi B., Denny P., Jones T.E.M., Seal R.L., Tweedie S. "Guidelines for Human Gene Nomenclature." *Nat. Genet.* 2020; 52: 754–58.