

Seeking stability in gene nomenclature

Anne Paxton

April 2021—Human first names are not necessarily known for being meaningful—or unique for that matter. When Shakespeare’s Juliet muses, “What’s in a name?” she’s observing that her lover’s name is more or less an arbitrary label without relevance to the essence of Romeo.

Human gene names, by contrast, cannot be arbitrary. Because the stakes of adopting the right gene name are high, gene nomenclature standards ensure that a lot of essence is conveyed by what is literally “in” each human gene’s name.

Fei Yang, MD, assistant professor of pathology at Oregon Health and Science University, knows a great deal about the need for sound gene nomenclature. In her work as a clinical molecular geneticist with Knight Diagnostic Laboratories, she performs next-generation-sequencing-based mutation screening panels for patients with leukemia and other hematological malignancies to help clinicians choose the best treatments. “I deal with genes every day,” Dr. Yang says. “All the time.”

“Detecting gene mutations plays an increasingly important role in the clinical management of cancer patients,” says Dr. Yang, a member of the CAP/ACMG Biochemical and Molecular Genetics Committee. “In co-management of patients, we have people with different expertise”—oncologists, clinical laboratory personnel, and clinical and basic researchers. When gene mutations are discussed, “It is of the utmost importance that we all know we are referring to the same gene,” she says.

That has made stability a central goal of gene nomenclature. But there’s much more to the nomenclature process than achieving stability.

The HUGO Gene Nomenclature Committee (HGNC), an international, voluntary standard-setting committee of the Human Genome Organization (HUGO), aims to ensure that for research and clinical purposes, the symbol and name of any individual human gene succinctly convey as much as possible about their unique subject and avoid as much confusion as possible when used.

Since its launch, the HGNC has issued a series of updated guidelines for the naming of human genes. The goal has been to reflect the major changes and the increase in knowledge and data in human genomics.

Having named some 42,000 human loci to date, and facing a constant stream of novel human genes ready for naming and existing genes that might be renamed, the HGNC has regular updating to think about. But at the same time, the organization strives, in the interest of consistency and stability, to make minimal changes to gene nomenclature.



Dr. Yang

“Two years ago, I started to design an NGS-based clinical assay to target genes associated with germline cancer predisposition to myeloid malignancies,” Dr. Yang says. “One of the genes, *EFL1*, is associated with an inherited condition called Shwachman-Diamond syndrome, and it has been associated with an increased risk for developing pancytopenia and malignant transformation to MDS or AML.”

"But when I communicated with the company that was helping me design the sequencing primers, this gene could not be taken up into their system." Later, she found out that the company was using an older gene symbol for this gene: *EFTUD1*. The correct name was the one Dr. Yang was using.

"So that caused a problem when I designed the assay. And I had to find another way to be able to communicate with the company."

"In the situation where some of the older gene nomenclature was replaced by a new one, a clinician may get the old nomenclature of some genes based on publications. And when we, as a laboratory, report a gene mutation with a new gene nomenclature, we may seem to be talking about different things, but we're not. So it hampers our communication."

As a result, she supports the aim to standardize gene names—but she notes there is a downside. "I would appreciate it if the incidence of changing the gene nomenclature could be kept at a minimum. Changing gene nomenclature could impact the communication of clinical labs with physicians in reporting the test results, for example."

A good example of such a problem, Dr. Yang says, relates to the *KMT2A* gene. "This is the gene symbol currently approved for a gene previously known as *MLL*. It's a well-known prognostic biomarker for acute myeloid leukemia. Alterations involving *MLL* define a subtype of leukemia with aggressive disease course and suboptimal outcome when treated with conventional chemotherapy."

"This gene has been researched for over 30 years and a lot of data has been published under the gene symbol of *MLL*. It is noticeable that *MLL* is still a preferred gene symbol both in the recent literature as well as among oncologists."

"But since the gene symbol change to *KMT2A* two or three years ago, nowadays when we issue the report we have to use *KMT2A* and then put *MLL* in parentheses afterward in order to avoid confusion and to facilitate communication."

A name change could explain why a pathologist might not understand a particular gene name, she says, making it important to keep current on what is happening with the naming of genes.

As research continues, there will be increasing room for use of the HGNC gene nomenclature, Dr. Yang says. "More and more genes are being discovered by research to be associated with certain conditions and to have clinical utility in dealing with those conditions."

For Karen Tsuchiya, MD, the issue of gene nomenclature first arose when she was a member of the CAP Cytogenetics Committee, and it was about the *KMT2A(MLL)* gene symbol. "One of my colleagues brought it up. It was shortly after the *KMT2A* gene symbol was changed from *MLL*, which it had been for many years," says Dr. Tsuchiya, associate professor of laboratory medicine and pathology, University of Washington School of Medicine.

Her colleague was concerned about the change. "He had had some experiences and found it potentially created patient safety issues if the people you're writing the report for don't know that the gene name and symbol have changed. There was so much literature that used the *MLL* gene symbol, and that was what all the oncologists were familiar with."

Another problem with gene nomenclature is when a gene has one or more aliases. One of the best examples of this, Dr. Tsuchiya says, is *ERBB2*. "Everyone knows it as *HER2*, but the approved gene symbol is *ERBB2*, which is not as well known. As with the *KMT2A (MLL)* gene symbol, there is a lot of literature out there with the alias gene symbol *HER2*." It's to avoid the confusion of this and other aspects of gene naming that she uses the HGNC website (www.genenames.org) frequently in her practice. The online database there contains all approved human gene symbols.

The gene nomenclature process is mostly a cooperative one, she explains. "The HGNC updates the names on a

rolling basis. And the HGNC can't require anyone to go through the organization to name a gene or rename a gene. They can strongly recommend it. But they are relying on journal editors, manuscript reviewers, and professional societies to enforce that recommendation because they can't require it."

What's unique about the most recent gene nomenclature naming guidelines, Dr. Tsuchiya says, is that there are fewer name changes in the offing (Bruford E, et al. *Nat Gen.* 2020;52[8]:754-758). "Over time, the committee has come to recognize it's a problem when there are frequent changes to gene names. It's disconcerting when you realize all of a sudden this has happened."

"You tend to figure that out in different ways," she adds. For example, "Sometimes clinicians will order testing based on specific gene symbols, and there's difficulty if they're searching under a name that the laboratory doesn't have that test designated as. Those are where potential patient safety issues arise," Dr. Tsuchiya says.



Dr. Tsuchiya

Or, "You'll be working on a case and researching a gene, and then you'll realize the name has changed. The committee came to recognize that as technology has evolved and genetic testing has taken on more importance, there was not a good way to disseminate this information. So stabilizing the gene names and symbols should be one of their priorities, and it's one thing the publication emphasizes. Because these names are internationally used."

Having a name that reflects the character and function of the gene product is another priority, Dr. Tsuchiya says. As the HGNC describes its basic approach to gene symbols and names: "Ideally, gene symbols are short, memorable, and pronounceable, and most gene names are long-form descriptions of the symbol. Names should be brief and specific, and should convey something about the character or function of the gene products but not attempt to describe everything known."

Says Dr. Tsuchiya: "There are multiple reasons why gene names have changed over the years," and making the names meaningful is one of the reasons. "Very early on, when the HGNC was initially developed, they decided the gene names should reflect the character and function of the gene product, which is usually, though not always, a protein."

A gene name might get the axe because it could be perceived as pejorative. The genes formerly known as *DOPEY1* and *DOPEY2* are examples. "It just wasn't the greatest thing to sign out a patient report that the patient's parents were going to potentially look at, by using the name *DOPEY1* or *DOPEY2*." Those were renamed to *DOP1A* (DOP1 leucine zipper like protein A) and *DOP1B* (DOP1 leucine zipper like protein B).

Another instance involved a common feature of spreadsheet software that was leading to too many errors in the way a gene name was displayed. That occurred with *SEPT1* and *MARCH1* symbols, which were routinely autocorrected to a date; they were changed to *SEPTIN1* and *MARCHF1* to avoid this autocorrection.

Sometimes the researcher who discovered a gene may have selected the name arbitrarily. "There may have been no rational basis behind what they chose," Dr. Tsuchiya says. "So at some point there was the realization that we need rational and standardized nomenclature. That's when the committee decided it should be primarily based on the function of the gene product. For genes within the same family of genes that have similar function, it makes sense for those to have similar names too."

An added source of stability on the HGNC's website is the display of each gene's HGNC ID as well as its symbol and

name, she says. “That’s very important because the HGNC ID is much more stable than the symbols or names.”

Having an invariant like the ID helps avoid problems, she says. “With a lot of current genetic testing, such as with chromosomal microarrays and NGS, they use a bioinformatics pipeline for data analysis, and if you have a gene name in your pipeline that changes, you can run into problems where the new gene name won’t be recognized by your pipeline. The HGNC ID is one way to mitigate that problem instead of using the actual gene symbol.”

Another important feature of gene nomenclature in practice is its open-access nature, Dr. Tsuchiya notes. “All of us have access to it—basically anybody who has to use gene names or symbols in their practice in the laboratory or on the receiving end, the clinicians. It’s important for all of us to be able to look up the most recent approved gene name and symbol.”

One of the ways professional societies like the CAP have tried to mitigate the problem of lack of recognition of a newly approved gene symbol is with a standardized format, she says. “You’re supposed to use the approved symbol first, then the commonly known or previous symbol in parentheses after it. It doesn’t get rid of all the problems that can occur with lack of recognition, but it can help. The same with *HER2*. You should really use *ERBB2* and then *HER2* in parentheses after it.”

The HGNC advises authors to quote the approved gene symbol (e.g. *BRAF*) at least once in the abstract of any publication, along with the gene’s unique HGNC ID number. The format would be HGNC:number (for example, HGNC:1097).

While it’s a top HGNC priority to stabilize gene names, there are still reasons why the names might change in the future, Dr. Tsuchiya says. One reason is that some gene symbols were never meant to be permanent because nothing was known about their function when they were discovered. “They were meant to be placeholder symbols. For example, genes that start with *KIAA* or that have ‘orf’ in the name in lowercase, many of those will change in the future when their function is discovered. They were never meant to be permanent names to begin with.”

“Some of them are symbols that we do use clinically not infrequently. For example there is a fusion that occurs in low-grade gliomas with one of the *KIAA* genes, and we’ve gotten used to using that name. It’s going to be hard to get used to if that changes.” However, she adds, the HGNC might not change some of those names for that reason. “The plan is to consider those on a case-by-case basis.”

A helpful feature of the HGNC website are “tags” that appear on the symbol report page for a gene. The tags have an icon that looks like a luggage or sales tag. One example is the “stable symbol” tag, which indicates that the gene symbol has been reviewed by HGNC, is considered stable, and should not be subject to future change.

In summary, Dr. Tsuchiya says, “There’s a problem when genes have multiple names, either because different researchers in the past gave the same gene different names, or the name is changed in the present. But it’s problematic if everyone is using something different.” As the guidelines reflect, “The HGNC has developed a standardized and rational approach to designating gene names and symbols that everyone should be using.”□

Anne Paxton is a writer and attorney in Seattle.