Mastering Genomic Terminology

Gail P. Jarvik[1] and **James Evans**[2]

[1]Departments of Medicine (Division of Medial Genetics) and Genome Sciences, University of Washington, Seattle

[2]Departments of Genetics and Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC

> "When I use a word," Humpty Dumpty said in rather a scornful tone. "It means just what I choose it to mean - neither more or less."
>
> "The question is," said Alice, "whether you can make words mean so many different things."
>
> "The question is," said Humpty Dumpty, "which is to be master - that's all."
>
> - Lewis Carroll

Imprecise language leads to imprecise thinking and subverts meaningful communication.

As does any highly technical field, Medical Genetics uses a specialty language. However, we fail to communicate with each other and with patients if we do not share common meaning with our common vocabulary.

The goal of this commentary is two-fold: To clarify a few specific examples of terminology in Medical Genetics that we find particularly problematic and to stimulate further efforts to codify the language of our field. By doing so, our intent is to foster better communication with one another, with other medical practitioners and ultimately with patients.

Some examples of commonly misused genomic terminology

carrier   heterozygote
polymorphism   benign
mutation   pathogenic
truncation   premature stop codon
penetrance   expressivity
exome/genome sequencing   whole

The newest challenge to clear language in our field may be the use of the word *carrier*. Historically, many of us were taught to use the word carrier to designate a person who "carries" a single pathogenic allele for an autosomal recessive condition, with the implication that the other allele is wild-type or non-pathogenic and that the "carrier" was thus medically unaffected. However, in this age of accelerating genetic testing, an increasing number of patients are found to have a pathogenic variant for a dominant condition (autosomal or X linked) who do not (yet) manifest the associated condition, leading to increasing (and confusing) use of "carrier" in that very different context. We suspect that augmenting this confusing usage is the lack of an obvious alternative. Therefore, we suggest resolving this ambiguity by retaining the long used term "carrier" for those with a single variant in a recessive disorder and referring to individuals who possess a single pathogenic (or disease-causing or disease risk) allele in a dominant condition as *harboring* such an allele. Alternatively, one could refer to them as "heterozygous for a pathogenic (or disease-causing or disease risk) allele" with designation of the condition as dominant.

The American College of Medical Genetics and Genomics (ACMG) and Association of Molecular Pathology (AMP) variant classification guideline paper address two important points of terminology[1], which may have been lost in the dense paper.

The first clarification ACMG/AMP addressed was the use of the word *mutation*. Perhaps influenced by both Hollywood and the lay press, even the medical community often uses "mutation" to imply that a variant is disease causing. However originally, "mutation" simply meant any deviation from a standard sequence, regardless of the phenotypic impact. The increasing elucidation of variants in patients that range from having no phenotypic effect to pathogenic has made the casual use of "mutation" problematic, undermining clarity in communication with both patients and other providers. Complicating the implications of misusing this particular term is the above alluded to popular conflation of "mutation" with the grotesque and disturbing. These are hardly the messages we usually wish to communicate to patients in the clinical setting. For all the reasons cited above, we suggest that the term mutation has outlived its usefulness in our field and be abandoned. Rather, we should use the far more accurate terms *pathogenic variant*, *risk variant*, *disease-causing variant or de novo variant.*

ACMG/AMP also addressed the use of the word *polymorphism*. Polymorphism, contrary to much casual usage (even in the medical literature), does not mean "benign". Nor is it a synonym of "variant." A polymorphism is a variant with a population frequency of greater than or equal to 1%, and may be pathogenic or not. For example, the *HFE* c.845G>A variant, which causes the amino acid substitution, p.Cys282Tyr, is a polymorphism in the European Ancestry population, since it has an allele frequency of approximately 4% and is pathogenic for hemochromatosis. The terms *benign* or non-pathogenic are preferred for a variant that is shown not to be associated with disease or disease risk, with the later preferred due to its clarity, specificity and lesser chance of confusion given other medical uses of the term "benign". It follows that the term "single nucleotide polymorphism" (SNP), should be reserved for variants with allele frequency   1%; the term single nucleotide variant, SNV, is correct for all variant frequencies.

The term *truncating* when applied to a variant can also lead to misunderstanding. A variant that leads to an early stop codon, whether through a frameshift or a nonsense mechanism, can have two outcomes. First, nonsense mediated decay of an mRNA due to a stop in the first ~90% of the coding region, leads to no protein being translated at all and *haploinsufficiency*. Nonetheless, clinical labs often erroneously report these as "truncating" or even "leading to a truncated protein". Second, variants that leads to a stop in the last ~10% of the coding region are expected to typically escape nonsense-mediated decay of mRNA[2,3] and result in truly truncated protein products. These proteins are often functional, and may or may not be associated with disease[4]. When referring to either case, one should use a phrase indicating that the variant in question *encodes a premature stop of translation*. Ideally, this would be followed with whether this stop is expected to lead to haploinsufficiency or an actually truncated protein, if known.

It is the phenotype or trait that exhibits a given inheritance pattern, not the pathogenic variant. Thus, it is incorrect to say that a variant "is" autosomal recessive (or dominant, etc.), but correct to say it is associated with an autosomal recessive (or dominant) pattern of inheritance. Likewise, a variant can be accurately said to "be on" the X chromosome, but the related trait or disease is X-linked (or dominant or recessive).

Also common is confusion of the terms penetrance and expressivity. The phrase "variable penetrance" will rarely be correct, but may hint that the writer is actually intending to discuss expressivity. *Penetrance* is a binary function, with a trait or disease either manifesting or not (penetrant or non-penetrant, respectively). At the individual level there never exists variable, mild, or severe penetrance; the trait or disease is either penetrant (any manifestation of the disorder or phenotype found), or not (with no manifestation exhibited in the individual). *Expressivity*, on the other hand, refers to the range of phenotypes that may manifest in the context of a given disorder, such as mild vs. severe learning disability associated with a given pathogenic variant. Moreover, when penetrance is discussed, the context must be specified. For example, the penetrance of a given pathogenic *BRCA1* variant is different whether one is discussing breast cancer or ovarian cancer. Finally, many genetic disorders manifest age-related penetrance and this should be specified when this characteristic is discussed.

A commonly ambiguous term is "deleterious". This is often misleadingly used to imply pathogenicity. Yet what is typically meant by the term is that it disrupts, or is predicted to disrupt, the function of the encoded protein. Yet a variant may be deleterious and not pathogenic for a given disease because the gene in which it resides is not rigorously linked to the phenotype in question. Similarly, the term pathogenic should not be used when disease association in not established, even when there is evidence of altered *function* of the resulting protein, unless it is shown that that change will result in disease risk. We suggest that "deleterious" be avoided in the context of medical genetics and, rather, the impact of the variant on the protein in question be articulated (e.g. resulting in haploinsufficiency, truncation, reduced enzyme activity *in vitro*, or pathogenic if known). "Deleterious" is properly used in the context of population biology for an allele shown to reduce genetic fitness.

Others have called for removal of the word "whole" from terms such as "whole genome sequencing" and whole exome sequencing". At the risk of descending to the quixotic, we agree that these terms are neither correct nor precise. Neither WGS nor WES are "whole", with non-trivial parts of the genome eluding even high quality sequencing at present. Dropping the "whole" would also remind us of the limits of our technology. The term "Next Generation Sequencing" is similarly problematic. When does a new platform become the *next* next generation? Moreover, other than an (admittedly appealing) invocation of *Star Trek*, the term tells one nothing of the underlying technology. Rather, "massively parallel sequencing" is preferred, with the virtue that it actually communicates the underlying concept of the technology.

We understand that nomenclature and "rules" about terminology can be tedious and readily descend to the pedantic. However, a lack of attention to proper language can impede and overtly derail communication. As has been said (most commonly attributed to George Bernard Shaw), "The single biggest problem in communication is the illusion that it has taken place". If we are not precise in our language, we will inevitably mislead each other and our patients.

We encourage professional organizations like the ACMG and AMP to continue efforts to actively shape the precise and accurate use of nomenclature and terminology in our field through existing and novel efforts to define and refine the language of genetics. We also welcome readers to contact us at (gim@acmg.net) with their own concerns, pet peeves and examples. In the meantime, *Genetics in Medicine* will continue efforts to ensure that vocabulary and terminology used in the manuscripts we publish is as precise, informative and unambiguous as possible.

References

1. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015; 17(5):405–424. [PubMed: 25741868]

2. Maquat LE. Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics. Nature reviews. Molecular cell biology. 2004; 5(2):89–99. [PubMed: 15040442]

3. Conti E, Izaurralde E. Nonsense-mediated mRNA decay: molecular insights and mechanistic variations across species. Current opinion in cell biology. 2005; 17(3):316–325. [PubMed: 15901503]

4. Isidor B, Lindenbaum P, Pichon O, et al. Truncating mutations in the last exon of NOTCH2 cause a rare skeletal disorder with osteoporosis. Nat Genet. 2011; 43(4):306–308. [PubMed: 21378989]