



# Clinical Chemistry Trainee Council

## Pearls of Laboratory Medicine

[www.traineecouncil.org](http://www.traineecouncil.org)

**TITLE: DNA Sequence Nomenclature and Variant Interpretation**

**PRESENTER: Lora J. H. Bean, PhD, FACMG**

---

**Slide 1:**

Hello, my name is Lora Bean. I am an Assistant Professor and Senior Laboratory Director at Emory Genetics Laboratory, in the Department of Human Genetics at Emory University. Welcome to this Pearl of Laboratory Medicine on “DNA Sequence Nomenclature and Variant Interpretation.”

**Slide 2:**

The human genome consists of over 3 billion base pairs. Although the Human Genome Project was completed in 2003, gaps in the sequence still exist. As these gaps are filled in, the location or “map position” of a specific base pair often changes with each version of the human genome reference sequence. A reference sequence provides a framework on which to place variations in the genome – and there is a lot of variation. We refer to single nucleotide differences between individuals as single nucleotide polymorphisms (or SNPs) and greater gains or losses of sequence between people as copy number variants (or CNVs). We are interested in these variants for different reasons – some variants are benign, common changes that are used in association or copy number studies while others are deleterious changes involved in human disease. Pathogenic changes identified in the pre-genomics era are often referred to using historic non-standard nomenclature.

**Slide 3:**

Pathogenic variants in human genes cause heritable genetic disease. When one or a few pathogenic variants account for most disease alleles in a population, a targeted genotyping assay can be used to test for the pathogenic variants. In most cases of genetic disease, private or familial mutations comprise most of the mutation spectrum. Clinical laboratories use full gene sequencing to detect pathogenic changes. Novel (previously unreported) variants are often detected. In the clinical laboratory, we need to use resources that help us determine the pathogenic nature of such variants. We also need to use standard nomenclature to describe the variant.

**Slide 4:**

No matter the purpose of a study, it is very important that we are able to compare data across studies and between individuals. The best way to ensure that a specific DNA variant can be unambiguously identified is to use widely understood naming conventions. Current conventions utilize publicly available reference sequence accession numbers. Such conventions are particularly important when referring to

rare or familial gene mutations, since a generation or more may pass before additional family members present for testing, by which time known mutation carriers may no longer be available for confirmatory studies.

**Slide 5:**

Widely accepted recommendations for sequence nomenclature are given by the Human Genome Variation Society (HGVS). There are recommendations for naming genomic (as g dot), coding (as c dot), mitochondrial (as m dot), RNA (as r dot), and protein (as p dot) sequence changes compared to a reference. HGVS has naming conventions for single base pair variants, deletions, duplications, insertions, large rearrangements, intronic variants, and nearly any scenario that has been reported.

**Slide 6:**

Here are some examples. Changes from a reference sequence are described by naming the reference base or amino acid and noting the new one. Nucleotide and amino acid changes are meaningless without a reference sequence so their use is critical. Note that in the genomic nomenclature, the nucleotide change may be different than in the coding nomenclature. This is because genomic numbers refer to one strand of the chromosome, known as the “+” strand, and a gene may be in the opposite orientation on the chromosome, or on the “-” strand.

**Slide 7:**

Let's take the example of the most common sickle cell disease mutation. Most references in older literature, as well as medical records, use the historical name HbS, hemoglobin S, or *HBB* Glutamic acid 6 Valine, and these names are familiar to us. Using standard nomenclature, this mutation can be named using genomic, mRNA, or a protein reference. The pathogenic variant is a T to an A because the *HBB* gene is on the “-” strand of the chromosome. Using standard nomenclature, the protein change is now referred to as p. Glutamic acid 7 Valine. To obtain data on this mutation from large scale sequencing projects or genotyping platforms, the genomic location or a reference number, such as a dbSNP rs number, must be known.

**Slide 8:**

The HbS sickle cell mutation is well-established as being pathogenic. Not all variants in the genome are so easily interpretable. Clinical laboratories follow professional standards and guidelines from the American College of Medical Genetics to classify sequence variants into one of five categories – pathogenic, likely pathogenic, unknown clinical significance, likely benign, or benign. The guidelines are currently being revised and updated.

**Slide 9:**

Some variants are classified as pathogenic, even if they have not been previously reported, because they are predicted to destroy protein function. These include nonsense, frameshift, and splice site mutations. Missense variants and in-frame deletions or duplications, as well as silent changes (DNA variants predicted to not change an amino acid) or intronic changes outside of the plus / minus 1 and 2 positions, may also be classified as deleterious if there is sufficient evidence in the literature for pathogenicity of the change.

**Slide 10:**

Let's look at a simple example of a nonsense change in the *GJB2* gene identified by sequence analysis. The *GJB2* gene encodes the protein connexin 26. Mutations in this gene are associated with autosomal recessive hearing loss. The reference sequence is the bottom trace in which you can see the normal GAG that encodes a glutamic acid. The patient sequence is the top trace. You can see that the patient is heterozygous for a G to T change at coding position 139. In addition to the normal GAG-encoding glutamic acid, the patient also has a TAG which encodes a stop. This premature stop codon is interpreted as a pathogenic change.

**Slide 11:**

This is an example of a common splice site mutation in the *PAH* gene, which is mutated in patients with PKU. Remember that when RNA is transcribed from DNA, introns are present. In order to produce a mature mRNA, the introns must be removed from this pre-mRNA. The splicing machinery recognizes introns by their sequence. The "GU" sequence at the beginning of the intron, known as the splice donor, and the "AG" sequence at the end of the intron, known as the splice acceptor, are nearly invariant. Variants at these +1, +2, -1, and -2 positions are interpreted as pathogenic.

**Slide 12:**

At the other end of the spectrum, we have benign changes. We classify variants as benign if they are observed at a frequency in the population that is too common to be a disease-causing mutation. These data may come from large population studies or from control groups in published studies. Sometimes results from published functional studies demonstrate that a variant has no effect on protein function.

**Slide 13:**

Large scale sequencing projects have provided us with a much better view of common variation in the human population. For example, the NHLBI GO exome sequencing project has made allele frequency data from over 8,000 European-derived alleles and over 4,000 African-derived alleles publicly available. These are valuable when determining pathogenicity of a sequence variant.

For example, the C to T variant 34 nucleotides upstream of the ATG start in the *GJB2* gene (referred to as coding -34 C to T), is common in the African-derived population with over 23% of alleles being a T. This change is defined as a benign variant. The variant G to A at coding position 79 is more difficult to interpret since it is rare in both the European and African populations.

**Slide 14:**

Data from the HapMap project, a large-scale population study designed to study genetic variation, is from a more diverse population. The c.79G>A variant in *GJB2* gene is common in Asian populations, with over 15% of people studied being homozygous for this variant. We can conclude that the variant is not associated with hearing loss. These data underscore the importance of studying diverse populations.

**Slide 15:**

Many sequence variants have not been previously observed and their pathogenicity cannot be determined. Such changes include missense (or amino acid) changes that have not been previously reported, in-frame deletions or duplications (loss or addition of amino acids without a change in the reading frame), or, as now identified by exome sequencing, nucleotide changes in genes not known to be associated with human disease.

**Slide 16:**

Here is an example of a variant of unknown significance in the *MLL2* gene, mutations in which cause Kabuki syndrome. Kabuki syndrome is an autosomal dominant dysmorphology syndrome. A single pathogenic variant in *MLL2* causes Kabuki syndrome. In this case, a G to A change at coding position 10740, the last base pair of exon 38, was identified.

**Slide 17:**

The coding 10740 G to A change has not been observed in population studies or in individuals with Kabuki syndrome. Additionally, this change is not predicted to change the amino acid glutamine at codon 3580. So how do we determine the clinical significance of this change? We could test the parents to see if the change was inherited. This would also tell us if this variant is a new variant in this individual. Variants found in a patient but not their parents are said to have occurred *de novo*. For a laboratory to say that a variant occurred *de novo*, parental identity should be confirmed using genetic markers.

**Slide 18:**

We tested the parents and found that neither had this variant. With the family's permission, we used molecular techniques to confirm parental relationships. We then interpreted this change as likely pathogenic.

Clinical information about the patient and the parents is also critical to interpretation. If this change had been inherited, we would need to determine if the parent carrying the change has features of Kabuki syndrome. Such information would help us to build a case for calling this variant likely benign or likely pathogenic, but a final determination of benign or pathogenic would likely require identifying these changes in other people.

**Slide 19:**

In the case I just showed you, we combined data from the literature and databases of genetic variation with data obtained by testing family members of an affected person and classified a variant as likely pathogenic. The terms likely pathogenic and likely benign are usually reserved for variants in genes with clearly associated clinical or metabolic phenotypes – it is important to be confident that you are in the right gene. If studying the parents of an affected person demonstrate that an amino acid change at an evolutionarily conserved site segregates with a known mutation (for a recessive disease) or occurred *de novo* (for a dominant disease), then this is evidence for the variant being a likely pathogenic variant. If a variant is found homozygous in unaffected relatives (for a recessive disorder) or heterozygous in unaffected relatives (for a dominant disorder), then this is evidence for the variant being likely benign. As with any variant, it is important to reassess pathogenicity of variants anytime new information becomes available.

**Slide 20:**

It is important to be aware of the process that laboratories use to determine the pathogenicity of sequence variants. Gathering accurate, complete information on a variant from the literature and from population studies requires use and understanding of standard nomenclature.

**Slide 21:**

Emerging efforts will allow clinical laboratories to centralize knowledge about sequence variants, including the NCBI ClinVar project and the Human Variome Project. These efforts will be especially useful for clinical laboratories as we move farther into the genomic era of whole exome and whole genome sequencing.

**Slide 22: References**

**Slide 23: Disclosures**

**Slide 24: Thank You from [www.TraineeCouncil.org](http://www.TraineeCouncil.org)**

Thank you for joining me on this Pearl of Laboratory Medicine on “DNA Sequence Nomenclature and Variant Interpretation.” I am Lora Bean.