

Annotation Guidelines by column

Created by Robert Rivera, MD

1. Allele column

- a. The first letter or sequence of letters will represent what type of mutation occur: s for SNV, del for deletion, ins for insertion, and mu for multi.
- b. After the string that identifies what type of mutation occurred, insert the number representing the blood system the allele is referring to. For example, you would enter 6 if you are labeling alleles corresponding to KEL. Followed by system #,
- c. The tail of the generated allele label will be a string that represents its sequential position in the table. The labeling should be as follows: a,b,c.....z,aa,ab,ac...,ba, bb, etc.
- d. So the 3rd deletion in the list of alleles for KEL would be labeled del6c, del for deletion, 6 for KEL antigen group, and c as this is the 3rd letter in the alphabet.

2. Description column

- a. There is some variability to this column but overall, the labels should adhere to the following format.
 - i. Take the allele name listed in the table and remove all *, replace with underscore “_”.
 - ii. Decimal points “.” should be replaced with underscores “_”.
 - iii. If there is a negative value in the allele name, remove the “-” and place the string “neg” at the end of the label, while adhering the above.
 - iv. For example KEL*02.-14.2 would become KEL02_14_2neg.
 - v. If 2 numbers are separated by * in the original allele name, the insert an underscore; for example: FUT2*01 becomes FUT2_01.
- b. If two alleles corresponding to SNVs are in the same genomic coordinate, then the labeling for those two alleles should be placed in a single row, with a backslash “/” separating the two allele labels.
 - i. For example, KEL*02.-14.2 and KEL*02.24 represent SNVs at the same coordinate, and so in the description column, they will be listed together in a single row as KEL02_14_2neg/KEL02_24

3. Class columns (Aclass, Cclass, Gclass, Tclass)

- a. There are exceptions to the rules below in some of the fields entered in the excel sheet; however, most follow this guideline:
 - i. A “-” will be used to indicate an abnormal value, or an error.
 - ii. “Not found” should be used when the reference allele does not correspond to the presence of a specific antigen, or when no weakening or null phenotype effect is predicted.
 - iii. When an allele variant corresponds to the presence of an antigen, enter the name of the antigen, removing all spaces and “*”, and replacing “.” with “_”. If there is a “-” symbol in the antigen name, replace it with “neg” at the end of the antigen name. If there is a “+” symbol in the antigen name, replace it with a “pos” at the end of the antigen name.
 - iv. If the mutation corresponds to a null mutation, enter the symbol of the blood group system, followed by an underscore “_”, followed by “null”, followed by

another underscore “_”, then followed by the allele name written in the format mentioned above.

1. Example of blood group system symbol: Kell=KEL, Lutheran=LU,
Duffy=FY
- v. If the mutation corresponds to a weak mutation, enter the symbol of the blood group system, followed by an underscore “_”, followed by “weak”, followed by another underscore “_”, then followed by the allele name written in the format mentioned above.
- vi. If multiple mutations are involved for a specific variant phenotype, enter “multi”