



## Filling out the Submission Worksheets

Created: June 6, 2005; Updated: June 15, 2010.

### Which Worksheet to Use

**I would like to submit some SNPs, but I'm having difficulty understanding the SNP Submission examples. What worksheets do I use?**

dbSNP has three submission options for information about variations:

dbSNP has three submission templates that are used to submit information about variations:

1. The SNPASSAY template is used to describe a new variation. dbSNP accepts all variations types, including single and multiple nucleotide substitutions, insertions and deletions, microsatellites, and named variations.
2. The SNPPOPUSE template is used to submit allele and/or genotype frequencies for a sampled population. The name "SNPPOPUSE" derives from the fact that the data submitted using this form is data for a SNP that is tested or assayed in a population.
3. The SNPINDUSE template is used to submit actual genotypes for assayed individuals. The name "SNPINDUSE" derives from the fact that the data submitted using this form is data for a SNP that is tested or assayed in an individual.

All three submission templates are in "templates\_SNPsub.xls", located in the [submission file](#), which can be found in the /specs subdirectory of the dbSNP FTP site. (ud: 3/21/08)

**Do I need to fill out all three parts of the SNPSubmission (SNPSub) worksheet?**

You will need to fill out at least the SNPASSAY worksheet, and depending on whether you have individual or frequency data, you can fill out either the SNPINDUSE or the SNPPOPUSE worksheets. SNPINDUSE would be preferred if you have individual genotype data.

**I assayed 4 distinct SNPs in 56 individuals. I selected these individuals at random from various geographical regions, so they can not be identified as populations. Do I fill out the SNPAssay form for each of these 56 individuals?**

Submit all four SNPs using just a single SNPAssay worksheet. There is a good [example](#) of such a submission available online. (7/23/07)

### General Instructions for all Three Worksheets

**I am new to using the dbSNP submission template. Do I enter my information in column A of the spreadsheet, or in column B?**

Please look at the dbSNP Submission Quick Start for [information on copying the template](#); once you have a copy, fill in column “B” so that your submission, when submitted, will be formatted similarly to this [example](#). (05/15/08)

### **Do I need to fill out all three parts of the SNP Submission (SNPSub) form?**

You will need to fill out at least the SNPASSAY worksheet, and depending on whether you have individual or frequency data, you can fill out either the SNPINDUSE or the SNPPOPUSE worksheets. SNPINDUSE would be preferred if you have individual genotype data.

### **The submission worksheets ask for the same information several times. Do I have to repeat the information?**

You do not need to repeat the CONTACT, METHOD, or POPULATION sections if they are the same for the SNPASSAY, SNPINDUSE, and SNPPOPUSE worksheets.

### **How do I create submitted SNP (ss) and RefSNP (rs) numbers for my submission?**

dbSNP assigns the ss and rs numbers after your submission has been loaded (ss) and clustered (rs). In the meantime, you can use your local SNP ID/name instead of an ss number or rs number when reporting genotype (SNPINDUSE) or frequency data (SNPPOPUSE). (5/16/05)

## **Submitting Additional Information**

### **I have a SNP resulting in a confirmed amino acid change. Should I include this information in a submission?**

You can put the description of amino acid change in the COMMENT tag for each SNP. Here is an example:

```
SNP: lp02849
ACCESSION: AF050163
LENGTH: 101
COMMENT: a nucleotide substitution in exon 6 resulting in an N291S amino
acid change in LPL
5'_assay:CTGCTTGAGTTGTAGAAAGAACCGCTGCAACAATCTGGGCTATGAGATCA
OBSERVED: A/G
3'_assay:TAAAGTCAGAGCCAAAAGAAGCAGCAAAATGTACCTGAAGACTCGTTCTC
```

(2/25/05)

## **The SNPASSAY Worksheet**

### **Do the CONT, PUB, and METHOD submission sections need to appear before each SNP, or just once at the top of the submission file?**

The CONT, PUB, METHOD sections should appear once. Place them on top of the submission in the order listed.

### **When we fill out the dbSNP submission form, what do we enter in the “ID” space of the METHOD section, the “ID” space in the POPULATION section, and the BATCH space in the SNPASSAY section?**

For the “ID” space in the METHOD section, you or your lab would choose a unique name/ID or local identifier for the method (e.g., “Sequencing Protocol 123123”, “LAB Sequencing Protocol”, or “SNP detection”). For the “ID” space in the POPULATION section, you or your lab would again choose a unique name/ID or local identifier—this time for the population (e.g., “Male Caucasians”). For the BATCH space in the SNPASSAY section, you or your lab would again choose a unique name/ID or local identifier—this time for the SNPASSAY batch (e.g., “07092004” or “BRCA1-SNPs”).

**If the same SNP is found in 3 different sources do we submit a SNP for each of the 3 sources?**

Yes. Submit for each of the 3 sources.

**Please define the term “BATCH”, which is found in the SNPASSAY section of a submission. How is the batch identifier chosen?**

The BATCH ID is a user-defined name for the batch of SNPs being submitted. The BATCH ID should be unique for each submission.

## The SNPINDUSE Worksheet

**We have genotype data on 5 different standard laboratory stocks of *Drosophila melanogaster*. How should I specify these stocks in the SNPINDUSE form?**

My understanding is that there are no genetic variations within *Drosophila melanogaster* stocks, and that variations are therefore assayed between the stocks. Since this is the case, you can treat each stock as an individual.

The submission format looks like this:

TYPE: INDIVIDUAL

IND:handle|loc\_pop\_id|loc\_ind\_id|tax\_id|sex|breed\_structure|loc\_ind\_grp

SOURCE: src\_type|source\_name|src\_ind\_id|src\_ind\_grp

PEDIGREE: curator|curator\_ped\_id|curator\_ind\_id|ma\_ind\_id|pa\_ind\_id ||

Please note: src\_ind\_grp is any grouping that the Source gives to this sample (e.g. coriell calls CEPH samples "Caucasian"). loc\_ind\_grp is any grouping that the submitter gives to the sample (e.g. HapMap calls CEPH samples "European". You could call it "Plate\_1" or whatever makes sense. (9/12/07)

**What are the "SNPINDUSE" and "SNPPOPUSE" submission worksheets used for?**

The SNPINDUSE worksheet is for reporting the genotype of the breed or strain of a SNP reported in the SNPASSAY section. The SNPPOPUSE worksheet is for reporting SNP allele frequency observed in a population. (5/16/05)

**I used subjects from a number of populations housed in the Human Diversity Collection, Coriell Cell Repository, in a study and want to submit the data to dbSNP. How do I address subject IDs in my submission?**

You can use the individual or sample IDs provided by the source (i.e. CEPH, Coriell, etc.). Please see the [“Individual Description Section”](#) of the online dbSNP submission documentation (9/19/06)

**How do I create submitted SNP (ss) and RefSNP (rs) numbers for my submission?**

dbSNP assigns the ss and rs numbers after your submission has been loaded (ss) and clustered (rs). In the meantime, you can use your local SNP ID/name instead of an ss number or rs number when reporting genotype (SNPINDUSE) or frequency data (SNPPOPUSE). (5/16/05)

**Do I need to fill out all three parts of the SNPSubmission (SNPSub) worksheet?**

You will need to fill out at least the SNPASSAY worksheet, and depending on whether you have individual or frequency data, you can fill out either the SNPINDUSE or the SNPPOPUSE section. SNPINDUSE would be preferred if you have individual genotype data.

**Can the gender field and the Pedigree data elements in the SNPINDUSE worksheet be omitted?**

Yes, gender can be blank and the Pedigree elements are optional. (7/27/07)

## The SNPPOPUSE Worksheet

### What are the the "SNPINDUSE" and "SNPPOPUSE" worksheets used for?

The SNPINDUSE field is for reporting the genotype of the breed or strain of a SNP reported in the SNPASSAY section. The SNPPOPUSE field is for reporting SNP allele frequency observed in a population. (5/16/05)

### Does the SNPPOPUSE submission section need to be placed before each ALLELECOUNT entry?

The SNPPOPUSE header is defined once for each batch and appears only once and is followed immediately by the frequency data for the batch.

### Do I need to fill out all three parts of the SNPSubmission (SNPSub) worksheet?

You will need to fill out at least the SNPASSAY worksheet, and depending on whether you have individual or frequency data, you can fill out either the SNPINDUSE or the SNPPOPUSE section. SNPINDUSE would be preferred if you have individual genotype data.

### How do I create submitted SNP (ss) and RefSNP (rs) numbers for my submission?

dbSNP assigns the ss and rs numbers after your submission has been loaded (ss) and clustered (rs). In the meantime, you can use your local SNP ID/name instead of an ss number or rs number when reporting genotype (SNPINDUSE) or frequency data (SNPPOPUSE). (5/16/05)