



Flat File Submissions

Created: April 6, 2007; Updated: February 24, 2014.

Documentation Required for a Flat File Submission

Where are the submission worksheets I need to fill out?

Go to the [submission file](#), located in the /specs subdirectory of the dbSNP FTP site, and open “[templates_SNPsub.xls](#)”. This file contains three Excel templates (SNPASSAY, SNPPOPUSE and SNPINDUSE) that you can use to create submission worksheet(s) for your submission.

See the “Using the Excel Submission Templates” section of this Quick Start Guide for more information on how to use the templates.

Choosing which Flat File Excel Submission Template(s) to Use

What is the difference between the SNPASSAY, the SNPPOPUSE and the SNPINDUSE templates?

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.

The three flat file submission templates are located at “[templates_SNPsub.xls](#)” along with flat file examples in the [submission file](#), which can be found in the /specs subdirectory of the dbSNP FTP site.

When do I use each of the three flat file submission templates?

If you are reporting a **new SNP** and **do not have** either **allele/genotype frequencies** for a sampled population or **actual genotypes** for the SNP, use the SNPASSAY worksheet only.

If you are reporting a **new SNP** and need to **report allele/genotype frequencies** for the sampled population, use the SNPASSAY and SNPPOPUSE templates together (see the next question for instructions on how to use the two templates to create a single submission form).

If you are reporting a **new SNP** and need to **report a genotype** for SNP-individual pairs, use the SNPASSAY and SNPINDUSE templates together (see the next question for instructions on how to use the two templates to create a single submission form).

If you are **reporting allele/genotype frequencies for SNPs** that are **already in dbSNP**, use the SNPOPUSE template alone.

If you are **reporting actual genotypes for SNPs** that are **already in dbSNP**, use the SNPINDUSE template alone.

The three flat file submission templates are located at “[templates_SNPsub.xls](#)” along with flat file examples in the [submission file](#), which can be found in the /specs subdirectory of the dbSNP FTP site.

Using the Excel Submission Templates and Creating a Submission

I've downloaded the templates. How do I use them to make my own submission worksheet(s)?

Once you've finished downloading the templates from “[templates_SNPsub.xls](#)”, located in the [submission file](#), which can be found in the /specs subdirectory of the dbSNP FTP site, you can use them to make a submission worksheet for yourself by creating a new Excel spreadsheet and copying the first column (Gray area) of the submission template you wish to use to the new spreadsheet.

If you want to use more than one of the templates to report data for your variation, and want all of the data to be on a single page, Copy the first column (Gray area) of the SNPASSAY template to your new Excel spreadsheet, then copy the first column (Gray area) of either the SNPOPUSE template or the SNPINDUSE template (or both) and paste it (them) at the end of the SNPASSAY section of the spreadsheet.

If you are **combining the SNPASSAY and the SNPOPUSE worksheets on a single page, please note:** you do not need to copy the CONTACT, PUBLICATION, METHOD and POPULATION sections of the SNPOPUSE worksheet onto the page if these sections will be filled with the same information that was provided in the same sections of the SNPASSAY worksheet.

If you are **combining the SNPASSAY and the SNPINDUSE worksheets on a single page, please note:** You do not need to repeat the CONTACT and PUBLICATION sections of the SNPINDUSE worksheet onto the page if these sections will be filled with the same information that was provided in the same sections of the SNPASSAY worksheet.

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 is formatted similarly to this [example](#).

Do I have to fill out the same information for the CONTACT, METHOD, PUBLICATION or POPULATION sections if the information I am providing is the same for the SNPASSAY, SNPOPUSE, AND SNPINDUSE worksheets?

If you are using more than one section of the submission form, you do not need to repeat the CONTACT, METHOD, PUBLICATION OR POPULATION sections if they are the same for the SNPASSAY, SNPINDUSE, and SNPOPUSE worksheets.

The three flat file submission templates are located at “[templates_SNPsub.xls](#)” along with flat file examples in the [submission file](#), which can be found in the /specs subdirectory of the dbSNP FTP site.

Which fields in the submission worksheets are required, and which are optional?

The Excel submission templates, located in “[templates_SNPsub.xls](#)”, which can be found in the dbSNP FTP site's [submission file](#), have been annotated to indicate which fields are required and which are optional. Those fields in column A (the grey column) of the template that use **RED** font for the field tags are **required** fields, and must be included in your submission. Those fields in column A (the grey column) of the template

that use **BLACK** font for the field tags are fields that are **not required** for submission, but you are welcome to provide the information in these fields if you have it. You are also welcome to provide any additional information in your submission that is not delineated by the existing submission fields so long as it conforms to dbSNP submission format as described in the dbSNP [submission documentation](#).

Can I use the Flat File format to report the position of my variant using either an asserted position or flanking sequence?

The Flat file format is used for submitting the position of a variant using flanking sequence only.

The only time flanking sequence is permitted as a means of placing a variant is when you are unable to submit an asserted position* because your sequence does not have an NCBI Assembly ID, or you were unable to obtain a GenBank ID or RefSeq ID for reporting your asserted position should your sequence not have an NCBI Assembly ID.

***NOTE:** An “asserted position” is a statement, or assertion, based on experimental evidence that a variant is located at a particular position.

See the VCF: Reporting Variant Positions section of this document for submitting an asserted position using VCF format.

Before submitting flanking sequence to dbSNP, contact us at snp-sub@ncbi.nlm.nih.gov.

The flanking sequence that contains the position of my variant doesn't have an NCBI reference ID. What do I do?

dbSNP no longer accepts flanking sequence submissions for organisms that have a reference assembly housed within the [NCBI Assembly Resource](#) and whose sequence therefore has an associated NCBI Assembly ID.

1. If your sequence has an associated Assembly Resource ID, you must submit your variant with an asserted position* using the VCF submission format.
2. If you do not know whether or not your sequence has an associated NCBI Assembly resource ID see the FAQ in this document for instructions on how to find it.
3. If your sequence does not have an associated NCBI Assembly ID, we suggest the following:
 - a. If you are submitting an asserted position on a sequence that does not have an associated assembly ID but does have a GenBank ID or RefSeq ID, submit an asserted position on the GenBank ID or the RefSeq ID using the VCF submission format.
 - b. If you are submitting the position of your variant on a sequence that is not yet in GenBank, obtain a GenBank ID number (a gi or accession.version) by submitting the sequence to GenBank first. Once you have a GenBank ID, submit your variant using the VCF format and report the asserted position of your variant using the GenBank ID that you obtained for the sequence.
 - c. If for some reason it is not possible to obtain a GenBank ID (gi or accession.version), then you may submit your variant position using flanking sequence in the Flat file format.

***NOTE:** An “asserted position” is a statement, or assertion, based on experimental evidence that a variant is located at a particular position.

NOTE: If you report the position of the variant with a GenBank ID, a RefSeq ID or by using flanking sequence, there will be alterations to the assignment of ss and rs numbers to your variant since our submitted SNP (ss) and refSNP (rs) assignment policies have changed. Consult the “Reporting Variant Position and Resulting Submitted SNP (ss) and Reference SNP (rs) ID Assignment” section of this document to determine which dbSNP ID(s) will be assigned to your submission.

The VCF submission template (http://www.ncbi.nlm.nih.gov/projects/SNP/docs/vcf_template.xlsx) contains an example submission found using the “Example” tab at the bottom left corner of the template.

The three flat file submission templates are located at “[templates_SNPsub.xls](#)” along with flat file examples in the [submission file](#), which can be found in the /specs subdirectory of the dbSNP FTP site.

There are four examples of different flat file submission types in the in the dbSNP FTP site submission file. What kind of submission does each example file contain?

example_SNP_genocount.txt:

This example shows the correctly formatted submission of four distinct SNPs assayed in a single population and a genotype count for that population using a combination SNPASSAY/SNPPOPUSE worksheet.

example_snp_only.txt:

This example shows the correctly formatted submission of five distinct SNPs assayed in a single population using a SNPASSAY worksheet.

example_snpinduse_genocount.txt:

This example shows the correctly formatted submission of a single SNP assayed in a single population; the genotypes for four individuals assayed from within that population; and the allele count for the same population, submitted using a combination SNPASSAY/SNPINDUSE/SNPPOPUSE worksheet.

example_two_pop_freq.txt:

This example shows the correctly formatted submission of the allele frequency information for two distinct populations and 3 distinct SNPs for one of those populations, submitted using a combination SNPASSAY/SNPPOPUSE worksheet.

The three flat file submission templates are located at “[templates_SNPsub.xls](#)” along with flat file examples in the [submission file](#), which can be found in the /specs subdirectory of the dbSNP FTP site.