



## Submitting Various Data Types to dbSNP

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

### Mutation Submissions

**If dbSNP is a database containing Single Nucleotide Polymorphisms (SNP), it doesn't accept mutations, right?**

Think of dbSNP as a "dbVariation" since it contains not only Single Nucleotide Polymorphism (SNP), but also indels, STRs, MNPs, etc. Biologists have also been using the term "Polymorphism" to refer to common variations and the term "mutation" to refer to rare allele variation. dbSNP includes both polymorphisms and mutations. Please note, however, that the use of the word "mutation" is being phased out in dbSNP, and will be replaced by the term "Clinical/LSDB variation".

Starting in the Spring of 2008, dbSNP began accepting submissions of human Clinical/LSDB variations as well as annotations to existing human variations (including phenotype) on the [Human Variation: Search, Annotate, Submit](#) site (for single submissions) as well as on the [Human Variation: Annotate and Submit Batch Data](#) site (for multiple submissions). As of this date, there are a total of 1266 records in dbSNP that were submitted as Clinical/LSDB variations (select "Clinical/LSDB variation" in the Entrez SNP limits page and click "GO" with out entering a search term in the Entrez search box), and 1134 records submitted as Clinical/LSDB variations that also have OMIM links (select "Clinical/LSDB variation" and "OMIM" in the Entrez SNP limits page and click "GO" with out entering a search term in the Entrez search box).

Those SNPs with clinical association(s) will have a red "VarView" ([Variation Viewer](#)) link in the "allele" section at the upper right of the refSNP cluster report. Clicking the link will take you to [the Variation Viewer Report](#) for the gene in which the SNP is found.

We expect that the number of Clinical/LSDB variation records in dbSNP will grow rapidly as more users discover dbSNP's resources for submitting them.

(07/09/08)

**Can we submit SNPs that might be somatic mutations? Problem is, we don't know which are germline SNPs and which are the mutations since we didn't sequence matched normal DNA.**

dbSNP accepts SNPs and mutations. Make sure that you state in your methods, however, that you have no way of knowing which SNPs are somatic and which are germline SNPs/mutations.(2/13/07)

**We have identified novel in some cell lines. Since this is an in vitro system I am not sure whether these variations can be submitted to dbSNP**

dbSNP does accept data derived from an in vitro system; just specify in your submission that the SNP is known to be a somatic variation by adding the following line in the SNP section:

SOMATIC: YES

#

You can comment about the variation in the line that occurs after the “#” if you wish. Such a comment is optional. (01/10/08)

### Can I use the dbSNP submission instructions for polymorphisms and mutation submissions?

Yes.

## Negative Data Submissions

### We sometimes find that the SNPs in dbSNP are not polymorphic in the population(s) we study. Can we report these negative data?

Yes, you can submit frequency information on existing SNPs, including negative results. Please follow the instructions below. You can also [find these instructions](#) on the web.

1. [Request a submission handle](#).
2. Format the frequency data for submission.

Example:

```
TYPE:SNPPOPUSE
HANDLE:WHOEVER
BATCH:1-2002
METHOD:MY_FREQUENCY_METHOD
ID: <Your Handle>|<Your Population Name/ID>
SAMPLESIZE:100
ALLELEFREQ:NCBI|ss1:A=0.50/T=0.50|SS_STRAND_FWD
ALLELEFREQ:NCBI|ss10:T=1.00/C=0.00|SS_STRAND_FWD <-- N.B. No variation observed in sample
```

[Additional formatting instructions](#) are available on the web. A submission template will also be sent with the handle confirmation email.

Please note that in addition to the frequency submission, you can also submit the actual genotypes (even on the non-polymorphic SNPs). [Please see the instructions](#) on how to do so.

### Are there types of variations that I should not submit to dbSNP, such as deleterious mutations?

dbSNP accepts all variations types, including single and multiple nucleotide substitutions, insertions and deletions, microsatellites, and named variations. (5/4/05)

### Can I still submit a SNP if I don't have genotype or frequency data?

Go ahead and submit; just skip the SNPINDUSE and SNPPOPUSE sections if you don't have genotype or frequency data. (5/16/05)

## Reporting Alleles

### How do I report three instances of the same allele (normal homozygous, heterozygous, mutated homozygous), where there are never more than three individuals/SNP?

I would use the SNPPOPUSE section; however, I would suggest that for the genotype frequencies, you write the genotype. For example, use "T/T" or "G/T" or "G/G" rather than "normal homozygous", "heterozygous", "mutated homozygous".

Using the SNPPOPUSE section will allow you to enter a sample size for each of your populations, as well as the frequencies of the different genotypes in each of the different populations. (5/31/06)

**How do I report an AA deletion, and the insertion of GC in the place of the AA (or AA>GC)? It was reported as an insertion/deletion in a NEJM paper because of nomenclature requirements.**

This is a multiple nucleotides polymorphism (MNP). The variation should be reported as AA/GC on the "OBSERVED:" line in the submission file. (5/4/05)

**When I submit a SNP to dbSNP, must the submission contain an allele frequency for every SNP submitted?**

Allele frequency data are not required data for SNP submissions. (2/13/07)

## Genotype Submissions

**We identified 6 novel SNPs by sequencing a panel of 48 genomic samples. We designed a mass spectrometer SNP assay for these SNPs and genotyped 1000 controls from the same population. How do we submit?**

You can submit the two populations, and the two methods using a SNPASSAY form and a SNPPOUSE form in a single file. For the SNPASSAY portion of the submission, use the method\_id of sequencing method, and the population\_id of the 48 individuals. For the SNPPOUSE portion of the submission, you use other method and population id. Please look at this [example](#). (06/04/08)

**Should I submit individual data instead rather than population frequency data?**

In general, we encourage the submission of individual data over population frequencies. Genotypes have many more potential uses in research, and therefore give the data a longer useful lifespan. (9/18/07)

**I want to submit SNP genotype data for four different populations, but don't understand the concept of a batch and what I use as the sample size for a batch.**

We find it useful to use a single population per batch. The sample size is the number of chromosomes in the population for each batch. If you want to submit data for over 10,000 SNPs, we suggest that you create separate batches for each chromosome. (10/2/06)

**How do I submit Affymetrix SNP chip results?**

Currently, we're looking into supporting submission of output files from high density genotyping platforms such as Affymetrix. Your submission would be the first. [Request a submission handle](#) online, and once the handle request is processed, a FTP account will be created to which you can upload your submission. We'll see if we can process the files and load the data to dbSNP.(5/1/06)

**We need to submit SNPs that include genotype data on 450 chromosomes; do we need to mention the genotype of each patient, or can we send an excel file containing the genotype data of the population?**

The preferred submission is the actual genotypes for 450 patients. Can you send an excel file containing the genotype data to snp-sub@ncbi.nlm.nih.gov. We'll check to see if the file you send contains information sufficient for submission.(5/24/06)

**When we submit genotypes and frequency data for SNPs extracted from dbSNP and genotyped in a different population, do we create new ss IDs under our submitter handle, or do we add the data onto the existing ss IDs?**

Please use the existing dbSNP ss ID to report your genotype and frequency data.

**I have identified more than 35 SNPs, and have assayed more than 400 individuals (400\*35=14000 results), but the SNP submission documentation states that I should report the results for each individual assayed for each SNP. How do I do this?**

Although submission of individual genotypes is not required, it is preferred. Since typing 14K genotypes into a file is not practical, we expect that the submission would be created using a computer program or script that queries the files or database containing the stored genotypes, and then writes out our simple submission format. (5/31/06)

**I am submitting SNPs discovered in a study of 48 cell lines: 28 consanguineous individuals and 5 families. Because all of the SNPs found in the children are also found in the parents, should I include the children in the population data?**

We have two types of genotype submissions, one for actual allele assignments (genotypes) and another for genotype frequencies. If you submit the allele assignment that is preferred, we would like to have the children included.

If the children are included, we will need you to submit a map of the relationships between the samples—something like a linkage ped file that includes pedigree, individual, mother, and father. Subsequent allele and genotype frequencies will take the family relationships into account. Once the genotypes are in dbSNP, views of the data will be similar to those in this [example](#) of an individual Genotype Batch report and this [example](#) of this Genotype Detail report.

Your families will also be included in the [pedigree view](#).

If you plan to submit only the genotype frequencies, please exclude non-founders in the sample number and in the frequency estimates.

## Frequency Data Submissions

**I would like to submit allele frequency information by ethnic group. If we have four ethnic groups, do we create four POPULATION entries?**

Correct. (03/21/08)

**Should I submit individual data instead rather than population frequency data?**

In general, we encourage the submission of individual data over population frequencies. Genotypes have many more potential uses in research, and therefore give the data a longer useful lifespan. (9/18/07)

**Can we report allele frequency along with the SNPASSAY information? If so, can the allele frequencies be included in the same file or do they need to be separated into two files?**

Submit them as the same file: Use the SNPASSAY submission template to create your submission form, and then append the SNPPOPUSE section from the SNPPOPUSE submission template to the end of your SNPASSAY form. (2/14/07)

**What are the criteria for submitting allele frequencies (heterozygosities) to dbSNP?**

There is no minimum population size, but with any luck, your sample size is large enough so that the variation can be detected. (7/6/05)

**Does dbSNP use TOP/BOT nomenclature, and can I use the Top/Bottom strand designation information in a frequency submission?**

dbSNP does use Top/Bottom strand designation information for submitted SNPs when that information is easily obtained. However, we do not use the Top/Bottom strand designation information in frequency submissions for the following reasons:

- 1 The Top/Bottom strand designation rules cover four distinct situations. The first two situations (A/G and A/C) are unambiguous so the Top/Bottom strand designation rules work well. When alleles are

symmetric (A/T and C/G), however, the strand designation rules rely on flanking bases at equal distances to variation site being non-symmetrical. Relying on neighboring base symmetry poses the following problems:

After having scanned the whole of dbSNP, I found a small but significant percentage of SNPs did not meet this condition (flanking bases at equal distances to variation site being non-symmetrical), so the Top/Bottom strand could not be defined.

dbSNP takes frequency submissions for variation classes (indels, multiple case substitutions, and microsatellites), that have no rules for the Top/Bottom strand designation. For symmetrical alleles in these classes, I have tried using the same rule (flanking bases at equal distances to variation site being non-symmetrical) to determine Top/Bottom strand designation and again found that a small but significant percentage of SNPs did not meet this condition, so the Top/Bottom strand designation could not be defined.

Some genomic regions contain SNPs so densely packed that flanking sequences also contain variations; in such a case, the Top/Bottom strand designation would not be stable.

- 1 Although most of dbSNPs variations have two alleles, dbSNP also includes variations that have 3 or 4 alleles — such as SNPs in the highly variant HLA regions.

Although it will be difficult to use our strand code for a frequency submission, we'd be glad to discuss how the code could be applied to your frequency submission files if you contact us at [snp-admin@ncbi.nlm.nih.gov](mailto:snp-admin@ncbi.nlm.nih.gov) . (8/29/06)

### **Can we submit the frequency and sequence data for a novel SNP?**

Yes, you can submit novel SNPs as well as their frequency data. (2/21/05)

## **Microsatellite Data Submissions**

### **Can I submit microsatellites?**

We would very much like to have a microsatellite submission. How many can you submit? Do you also have associated Genotypes? (10/12/05)

### **How do I submit a new microsatellite marker that I have identified?**

Please see the instructions below for submitting microsatellites to dbSNP.

1. Request a [submission handle](#) online.
2. Format the submission data as shown in the SNPASSAY templates located in “templates\_SNPsub.xls”, located in the [submission file](#), which can be found in the /specs subdirectory of the dbSNP FTP site. Please review the online [how-to-submit document](#) for additional instructions and detailed descriptions of the fields and their values. The fields in the spreadsheet are the minimum requirement for your submission, but you can include additional fields from the online document.
3. Send your formatted submission to [snp-sub@ncbi.nlm.nih.gov](mailto:snp-sub@ncbi.nlm.nih.gov).

### **Our lab has found many new microsatellite markers. Do we submit them to dbSNP or dbSTS?**

You can submit these markers simultaneously to dbSTS and to dbSNP. Please review our [online documentation](#). Allele variations and individual genotypes should be submitted to dbSNP.

## **Structural Variant Data Submissions**

### **Can I submit Structural (copy number) variations?**

NCBI is in the process of creating a database to house structural (copy number) variations (STV) — we hope to see submissions for this database accepted in the next few months, and the database itself is expected to launch in about a year. (07/11/08)

**I want to submit human structural variants that include insertions, deletions, and VNTRs. Some of these are as long as 100kb and some as short as 22 bp. How should I report the alleles?**

The current submission schema can only accept alleles less < 256bp. In the past we have implemented a kludge (a clumsy solution to the problem) where the long variations are placed on the comment line and given a name that is reported as named variation on the OBSERVED line.

We're starting to get more structural variations like yours so we need to implement a general solution for this situation. We will discuss the problem in the SNPdev group, and send you a response. (6/21/07)

## Submission of Multiple Linked SNPs

**How do I submit a sequence that contains multiple SNPs, all of which are linked? Do I submit them together or should I submit each SNP alone?**

Please submit each SNP individually regardless of whether or not they are linked. If adjacent SNPs are part of the flanking sequence of the SNP you are submitting, you can put an "N" at that position, or other more specific IUPAC codes. (09/05/07)

## Submission of Data Found Using Various Methods

**We identified 6 novel SNPs by sequencing a panel of 48 genomic samples. We designed a mass spectrometer SNP assay for these SNPs and genotyped 1000 controls from the same population. How do we submit?**

You can submit the two populations, and the two methods using a SNPASSAY form and a SNPPPOUSE form in a single file. For the SNPASSAY portion of the submission, use the method\_id of sequencing method, and the population\_id of the 48 individuals. For the SNPPPOUSE portion of the submission, you use other method and population id. Please look at this [example](#). (06/04/08)

**I'm submitting SNPs identified using minisequencing, so have only the 5' adjacent region of the target SNP, and a 400 bp PCR product. Problem is, both the "5'\_flank" and "3'\_flank" fields are mandatory if the assayed sequence is less than 25bp.**

The flanking sequences that you submit will only be used to locate the genomic position of a SNP, so as long as you know the sequence, it does not matter how you got them. The total length of 5'\_Flank + (5'\_assay) + observed + (3'\_assay) + 3'\_flank must be at least 100 bp, with each component being at least 25 bp (except the "observed" component of course), and the assay sequences are optional. You can simply provide 100 bp of the sequence. In your case, since you can amplify a 400 bp region by PCR, you should have no problem finding the 100 bp sequence you need. (09/04/07)

**We've created a pipeline to carry out *in silico* SNP prediction using the public mouse EST dataset. Can we submit these *in silico* SNPs to dbSNP?**

You can submit *in silico* SNPs. First, request a submission handle for your lab [online](#). Once we have received your request, a submission template and instructions will be sent to you along with your handle confirmation.

When you fill out the submission form, be sure to include a detailed description of the *in silico* SNP prediction method and specify the `METHOD_CLASS` as "Computation" in the method section. (3/29/06)

**I have used in silico methods to predict SNPs from EST data, and have produced validation criteria for screening out false positives. How do we submit fill out the submission form for these SNPs?**

The only sections in the submission form you need to fillout for in silico SNPs are CONTACT, PUB (optional), METHOD, and SNPASSAY header and SNP(s). You can skip the other sections.(8/23/06)

**If a SNP has been discovered by an in silico method, does a new Genbank file containing the flanking sequence need to be created for the submission, or can the Accession number point to a genome sequence scaffold?**

The accession number can point to the genome sequence scaffold.(9/15/06)

**Can I submit SNPs found by direct sequencing cDNAs using the cDNA sequence as the 5' and 3' assay sequence, without regard to where these sequences fall on the genomic sequence?**

Yes. (2/11/05)

**I want to submit SNPs that were identified by aligning sequence fragments. Can I submit these SNPs even though they are computationally derived?**

Yes, we accept computed SNPs from multiple sequence alignments. You will need to classify your method as "COMPUTATION" under the [METHOD section](#).

**Can SNPs that were identified via clustering of ESTs sequenced from cDNA libraries be submitted to dbSNP?**

Yes, if you can provide us with at least 100 bp of flanking sequences on both sides of the SNP. Please review the submission instructions and guidelines for the [SNP Assay section](#).

**Can I submit computationally derived SNPs that have no experimental evidence of their existence? If so, how do I do it?**

You can submit the individual genotypes using the SNPINDUSE section.(6/12/06)

## Submission of Previously Published SNPs

**On occasion I run into important published papers describing human SNPs which are not in dbSNP. Can people send you literature references and ask that you include the published polymorphisms in dbSNP?**

Most authors submit their SNP sequences to dbSNP prior to publication in order to see their SNPs annotated on a genome, compare their SNPs to other SNPs in dbSNP, and to include the SNP accession numbers in their publication(s). Not all authors were aware of the importance of pre-publication submission when dbSNP began — hence, your problem.

You can now submit previously published human SNP records using two new online resources. One is called the [Human Variation: Search, Annotate, Submit](#) site, where you can submit a single variation or annotation to a single variation and the other is the [Human Variation: Annotate and Submit Batch Data](#) site, where you can submit a number of annotations or variations simultaneously.

Using these resources, you can submit human variations described in a publication not authored by you, but you must be able to describe the variations using [HGVS nomenclature](#). You will see that the submission templates for the resources have a field for the inclusion of the PubMed ID number, so that the original reference for the variation will be linked to the submission. **Please Note:** these online submission resources do not take submissions for SNPs of other species. (07/10/08)