



VCF Submissions

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

Using the VCF Excel Submission Template to create a Submission

How do I create a dbSNP VCF Submission File?

As a convenience for submitters who have a [small](#) number of submissions, we provide a [VCF Submission Template](#) and example submission that you can use to create a submission file.

Those users who intend to send a large submission typically should have a computer program in place that can generate a VCF submission file for them. If you have questions about submitting large numbers of variants, contact dbSNP at snp-sub@ncbi.nlm.nih.gov

Using the VCF Submission template and example:

1. Insert data into the [VCF Submission template](#) using the instructions provided in the grey portions of the template and in the [dbSNP VCF Submission Format Guidelines](#). The VCF submission template contains an example submission you can access by clicking the “example” tab at the bottom left corner of the Excel Spreadsheet.
2. Once you have entered your data and have checked your submission for accuracy, use the “save as” command to save the file.

If you have any questions about creating a VCF submission file that cannot be answered using the provided documentation, contact snp-sub@ncbi.nlm.nih.gov.

Can I submit genotypes, allele frequencies and genotype frequencies using VCF?

Yes; if you have genotype or frequency data for new or existing variations, you can submit it using the dbSNP's [VCF Submission template](#).

Follow the instructions for frequency data submissions found in the [dbSNP VCF Submission Format Guidelines](#) and for genotype data, follow the [genotype submission example](#) provided by 1000 Genome Project in their description of VCF version 4.1.

VCF: Reporting Variant Positions

Why is the sequence I use to report the position of my variant required to have an associated NCBI Assembly ID?

When you submit a variant using the VCF format, you will be required to locate the position of your variant using an “asserted position”, which is a statement, or assertion, based on experimental evidence that a variant is located at a particular position.

The asserted position must be reported on a sequence that is part of an assembly housed in the [NCBI Assembly Resource](#) in order for dbSNP to:

1. Accurately map your variant
2. Validate the position and allele of your variant against the reported sequence.

Note: If the variant position and allele cannot be verified because they were submitted on a sequence not housed in NCBI's assembly resource, your submission may be delayed or not annotated on a sequence/assembly.

See the FAQ in this section that describes how to locate the NCBI Assembly ID for your sequence.

See the “Reporting Variant Position and Resulting Submitted SNP (ss) and Reference SNP (rs) ID Assignment” section to learn how the existence of an NCBI Assembly ID for your sequence will affect the assignment of ss and rs numbers to your variant.

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.

Can I submit an asserted position for my variant using either the VCF submission format or the Flat File submission format?

You must submit variant asserted positions* using the VCF format. We only accept Flat File submissions in the case where submission of flanking sequence is unavoidable. See the FAQ in this section that describes such a case.

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

How do I find the NCBI Assembly ID associated with the sequence(s) that I'm using to report the position of my variant?

1. Use the [NCBI Assembly Resource](#) to find Assembly IDs:

Currently, the NCBI Assembly resource is indexed for chromosomes only, so if you have either the GenBank or RefSeq chromosome accession.version number for your sequence, you can enter it into the search bar at the top of the NCBI Assembly Resource page, and click the “Search” button to see a list of associated assemblies and their respective assembly ID numbers. Select the appropriate Assembly whose date corresponds as closely as possible to the date of your experiment.

If you are reporting SNPs on multiple sequences that have GenBank or RefSeq chromosome accession.version numbers, enter all the accession.version numbers into the search bar at the top of the NCBI Assembly Resource page separated by spaces, and click the “Search” button. The assembly resource will return assemblies common to all the accession.version numbers you entered. Select the appropriate Assembly or Assemblies whose dates corresponds as closely as possible to the date of your experiment.

2. Use the [NCBI Gene Resource](#) to find Assembly IDs:

If you have either the GenBank or RefSeq Gene accession.version number for your sequence, you can enter it into the search bar at the top of the NCBI Gene Resource page, and click the “Search” button. The response page will include a list of available associated assemblies and their respective assembly ID numbers. Select the appropriate Assembly whose date corresponds as closely as possible to the date of your experiment.

What do I do if I can't find an NCBI Assembly ID for the sequence that I'm using to report the position of my variant?

1. We will accept an asserted position* on a sequence with a GenBank ID (gi or accession.version) or a RefSeq ID (accession.version). You can search for GenBank and RefSeq IDs using the [Nucleotide](#) resource.
2. If you have a novel sequence that does not yet have a GenBank ID, first submit your sequence to [GenBank](#). Once GenBank assigns a GenBank ID to your sequence, submit your variant with an asserted position using the newly assigned the GenBank sequence ID.
3. If for some reason it is not possible to obtain a GenBank ID (gi or accession.version), then you can submit your variant position using flanking sequence in the Flat File format, but before you try to submit flanking sequence, contact dbSNP at snp-sub@ncbi.nlm.nih.gov.

*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 submit the position of your variant on a sequence with a GenBank ID, a RefSeq ID or are using flanking sequence to position your variant, there will be alterations to the assignment of ss and rs numbers for 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.

I don’t have an NCBI Assembly ID for the sequence that I’m using to report the position of my variant. Can I use a GenBank ID instead?

We will accept an asserted position* on a sequence with a GenBank ID (gi or accession.version). You can search for GenBank and RefSeq IDs using the [Nucleotide](#) resource.

*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 submit the position of your variant on a sequence that only has a GenBank ID, 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.

I don’t have an NCBI Assembly ID for the sequence that I’m using to report the position of my variant. Can I use a RefSeq ID instead?

We will accept an asserted position* on a sequence with a RefSeq ID (accession.version). You can search for RefSeq IDs using the [Nucleotide](#) resource.

*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 submit the position of your variant on a sequence that only has a RefSeq ID, there will be alterations to the assignment of ss and rs numbers 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 sequence I'm using to report the position of my variant has never been submitted to a public database, and so does not have an ID number. How do I submit the position of my variant?

If you have a novel sequence that does not yet have a GenBank ID, first submit your sequence to [GenBank](#). Once GenBank assigns a GenBank ID to your sequence, submit your variant with an asserted position* using the newly assigned the GenBank sequence ID.

*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 submit the position of your variant on a sequence that only has a GenBank ID, 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.

I've tried to submit the sequence I'm using to report the position of my variant, but it is ineligible for submission and I can't get an ID number for it. How do I submit the position of my variant?

If for some reason it is not possible to obtain a GenBank ID (gi or accession.version) for your sequence from GenBank, and you are certain that your sequence does not have a RefSeq ID or associated NCBI Assembly ID, then you can submit your variant position using flanking sequence in the Flat File format.

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

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.

VCF: Reporting Variant Position and Resulting Submitted SNP (ss) and Reference SNP (rs) ID Assignment

If I submit the position of my variant on a sequence that has an associated NCBI Assembly Resource ID, will my submitted variant get ss and rs ID numbers?

Your variant will be assigned a submitted SNP (ss) number, and a will be assigned a Reference SNP (rs or RefSNP) number during the build that follows your submission if we can validate the submitted variant and can place it on an assembly.

Variations that are assigned a refSNP number are distributed as part of dbSNP, which allows the reported variation to appear on maps or graphic representations of the assembly, and be integrated with NCBI's other resources like Gene, ClinVar, dbGAP or PubMed.

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.

If I submit the position of my variant on a sequence that does not have an associated NCBI Assembly Resource ID because there isn't yet an assembly to which the sequence aligns, will my submitted variant get ss and rs ID numbers?

If your sequence doesn't have an associated Assembly Resource ID because there is not yet an assembly to which the sequence aligns, the reported variant positioned on that sequence will be assigned a submitted SNP (ss) number, but not a Reference SNP (RefSNP or rs) number.

Because the variation will not have an assigned rs number, it will not appear on maps or graphic representations of the assembly, and will not be integrated with NCBI's other resources.

The ss number of the variant will be reported on the "Submitted SNP" web report, will be available for search using dbSNP homepage's "ID search" tool, and will be made available on FTP site for download.

If at some future date a new assembly is created in the Assembly Resource to which the sequence aligns, an Assembly Resource ID will be created, and the reported variant will be assigned a rs number at that time. Once an rs number is assigned to the variant, the variant will be distributed as part of dbSNP, appear on maps or graphic representations of the assembly, and will be integrated with other NCBI resources.

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.

If I submit the position of my variant on a sequence that does not have an associated NCBI Assembly Resource ID because the sequence aligns to a gap in an existing assembly, will my submitted variant get ss and rs ID numbers?

If your sequence doesn't have an associated Assembly Resource ID because the sequence aligns to a gap in an existing assembly, the reported variant positioned on that sequence is initially assigned a submitted SNP (ss) number and not a Reference SNP (RefSNP or rs) number.

Because the variation will not have an assigned rs number, the reported variation will not appear on maps or graphic representations of the assembly, and will not be integrated with NCBI's other resources.

The ss number of the variant will be reported on the "Submitted SNP" web report, will be available for search using dbSNP homepage's "ID search" tool, and will be made available on FTP site for download.

If at some future date an existing assembly in the Assembly Resource is updated such that the sequence aligns, an Assembly Resource ID will be created, and the reported variant will be assigned a rs number at that time. Once an rs number is assigned to the variant, the variant will be distributed as part of dbSNP, appear on maps or graphic representations of the assembly, and will be integrated with other NCBI resources.

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.

If I submit the position of my variant on a sequence generated using an assay that provides just the variant and flanking sequence, will my submitted variant get ss and rs ID numbers?

If your sequence doesn't have an associated Assembly Resource ID because it was generated using an assay that provides just the variant and flanking sequence, **the reported "Assay Variant" will be assigned a submitted SNP (ss) number only.** The ss number will be reported in the "Submitted SNP" web report, will be made available on FTP site for download, and will be available for search using the dbSNP homepage "ID search" tool.

You can access Assay Variants only as archived data until such time as an assembly is available at NCBI that will allow mapping by BLAST and allow us to possibly assign an rs. dbSNP cannot predict when such an assembly will be made available, or when mapping by BLAST will occur – it could be delayed by months or possibly years. SS numbers can be used in publications describing Assay Variants.

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.