

This document details the requirements for submitting SNP's and/or variants to our Sequenom Genotyping service. This step is generally performed during initial discussions about your project and/or a quote request.

CONTENTS

1	Submitting your variants as RS numbers.....	1
2	Submitting your variants using other common annotation methods	2
3	Submitting your variants as Sequence.....	2
3.1	Proximal SNP detection.....	2
3.2	GC content.....	3
3.3	Duplicated Regions	3
3.4	Repetitive Regions	3
4	I have more questions.....	3
1	Submitting your variants as RS numbers	

The most common way to indicate what variant you want genotyped is to supply a dbSNP reference number (rs number). You can supply these as a simple list within a quote or an email.

RS (Reference SNP) numbers are unique ID codes given to SNPs by the National Center for Biotechnology Information (NCBI) and stored in a database (dbSNP).

<http://www.ncbi.nlm.nih.gov/SNP/>

We use the rs numbers to extract the SNP sequence information stored in the database. These sequences are what we base our designs on. One of the advantages of dbSNP records is that proximal SNPs which can interfere with our designs and generally well annotated and our designs can be modified to avoid them.

Not all valid SNPs have been assigned an RS number, and not all RS numbers refer to a valid SNP. However for the majority of SNPs (and some variants) in human and many other model organisms dbSNP provides an excellent resource.

IMPORTANT: Reported SNP Orientation

The SNP orientation (i.e. whether the SNP is reported in the + or – genomic strand) is dependent on sequence orientation provided by dbSNP. Unfortunately between different builds this orientation can change. For A/G or C/T SNPs, orientation changes can be easily picked up. However for A/T or C/G SNPs orientation changes can go unnoticed. If you are concerned about a potential orientation change between projects completed at different times, we can provide the original sequences from which the designs were based. Comparisons with this information can reveal orientation changes between A/T and C/G SNPs.

2 Submitting your variants using other common annotation methods

Generally we don't accept other SNP annotation methods. However if there is only a few SNPs where you only have a common annotation we will often try to help find the rs number or the relevant sequence.

3 Submitting your variants as Sequence

For non-model organisms or SNP's which are not annotated in dbSNP an rs number is sometimes not possible to get. In these cases you can supply the sequence information directly. We have the following criteria for supplying sequence information:

1. Approximately 100 bases flanking each side of the SNP locus (minimum of 75 bp) for efficient assay design
2. The sequence should not contain more than 1 ambiguous base ("N") within 25 bases of the SNP locus. If this is a limitation for a given SNP please discuss the options with us
3. The flanking sequence itself must only contain A, G, C, T, N or IUPAC ambiguous codes
4. The SNP should be bounded by square brackets as shown below:
ACCAACACCA[**A/C**]CTTCCTGGGCTACACCAGGACGC
5. Insertions/deletions should be marked with a dash as follows:
ACCAACACCA[**AGCT/**]CTTCCTGGGCTACACCAGGACGC
(Please note that the Sequenom platform can only deal with insertions or deletions up to around 10bp)
6. Trinucleotide polymorphisms, should be marked as follows:
ACGTCAGT[**A/G/T**]ACGTGCACCAC
7. There can be no spaces within the sequence
8. The SNP ID for each SNP can be the RS number or any label you want to use. The SNP ID can be up to 12 characters long and must not contain any spaces or any of the following (* ? / \ #. + : ;). An example of a formatted SNP list is shown below.
9. This information can be provided in a text file or Excel.

3.1 Proximal SNP detection

Proximal SNPs (SNPs next to or near to a SNP of interest) can cause assay failure or incorrect results. At AGRF we perform automatic proximal SNP detection for human SNPs only. For other organisms where submitters know the locations of proximal SNPs within 100bp of the SNP of interest, it is recommended that these be marked on the submitted sequence using ambiguous DNA codes.

3.2 GC content

GC content can affect the ability to develop a successful assay. The flanking sequence for the SNPs should have a GC content between 40% and 65% for ideal results.

3.3 Duplicated Regions

Duplicated regions of genomes are relatively common within the genomes of mammalian organisms. Some species may contain substantial amounts of duplicated regions. In most cases design in these regions is problematic and should be avoided. It is the submitter's responsibility to identify SNPs that fall in duplicated regions.

3.4 Repetitive Regions

SNP design in repetitive regions can be difficult or impossible. AGRF will attempt to design for SNPs in these regions but will fail them if the repeat regions cause significant problems.

4 I have more questions

If you have any further questions please contact us by phone or email.

PH: (07) 3365 4423

Email: SNPGroup@agrif.org.au