Interpreting Discrepancies in Submitted SNP (ss) Detail Reports

Created: July 7, 2005; Updated: February 18, 2014.

ss43470589 is validated, but it looks as if this indel was based on homozygous occurrence in only one individual (NA19012). Is this a true indel or a genotyping artifact?

ss43470589 was submitted on July, 2005 with its variation listed as "-/C/T". This SNP was later genotyped by HapMap, but HapMap's genotyping method does not detect deletions. The genotypes HapMap reported for ss43470589 were CC, CT, and TT. This SNP is marked "validated" because it meets the criteria of minor alleles observed in at least two chromosomes.

This is a validated SNP in dbSNP based on the above reasoning, but we do not know whether the "deletion" is real based on the data in dbSNP; this brings up an interesting point:

- When a SNP has two alleles, dbSNP's the validation status applies to both alleles.

But

- When there are more than 2 alleles, and dbSNP does not have genotype data covering all them, then the validation status just means that "there is a validated variation" at the position and does not necessarily mean that each allele have been "validated" or observed in at least two chromosomes.

In this example, the refSNP cluster rs10989589 has 8 submitted SNPs. Only ss43470589 was reported to include a "deletion", while the remaining submitted SNPs reported a C/T. If you are in a lab, and have observed a "deletion" in your samples, we encourage you to submit this genotype data to dbSNP, as we welcome additional genotype data for any existing variation in dbSNP. (**02/07/08**)

rs3181457 is validated, yet neither of the member ss of this cluster is validated. How trustworthy is the validation for this cluster?

The genotype results for rs3181457 in the refSNP cluster report as provided by HapMap, show that European and Sub-Saharan African are homozygous on T. The "G" allele count is one in each of the Han Chinese and the Japanese populations. This SNP has a minor allele count of two — just meeting the minimum criteria for this SNP to be called "validated by frequency".

The ss24813116 assay data was submitted by SEQUENOM with neither genotype nor frequency information, so ss24813116 isn't flagged as validated in the refSNP cluster report. The reason this refSNP cluster shows up as "validated by frequency of genotype data" is that HapMap submitted genotype data separately for the rs3181457 cluster. When the genotype data was submitted by HapMap, we linked it directly to the exemplar for this cluster, which is ss24813116. You can see this genotype data if you do the following:

1. Click in the link for ss24813116 located in the "NCBI Assay ID" column of the "Submitter Records" section of the rs3181457 refSNP cluster report

2. Once you get the "Submitted SNP Detail" report for ss24813116, scroll down to the "Submitted Individual Genotype" section of the report to see the data.

Another possible source of validation information is to look at the method description for ss24813116, but unfortunately, no detailed method description was submitted by SEQUENOM in this case. The submitter only submitted a method ID called "disc_method-1". You can get more information on SEQUENOM's SNP discovery method, by contacting them using the contact information provided in the "Submitted SNP Detail" report for ss24813116. Just click on the "SEQUENOM" handle link. (**11/15/07**)

C and G are listed as alleles in the "Summary of Genotypes" section of the detail report for ss15377600, yet the "Allele" section in the same report shows that the observed alleles are -/T.

One of the idiosyncrasies of dbSNP is that genotype & frequency data need to be linked to one of the submitted-SNP(ss) records within a refSNP(rs) cluster — specifically the ss exemplar for that cluster (see FAQ regarding ss exemplars for a refSNP) — because a refSNP will sometimes merge away. Linking genotype & frequency data to the ss exemplar becomes a problem when different submitted SNPs contribute different variations to the refSNP cluster. This is the problem with the submitted SNP (ss15377600) you mention in your question. In this case, ss15377600 happens to be the exemplar for the refSNP cluster, and is an in/del variation, while all other ss in the rs2070922 cluster are true SNPs and contribute the allele frequencies you found in the "Summary of Genotypes" section of the report.

The SNPdev team is thinking about separating refSNP clusters if the exemplar submitted SNPs within that cluster is of a different class from the other submitted SNPs in the cluster (such as indel vs. true SNP, as in this example). I will try to determine how many ss exemplars do not have the alleles reported in their refSNP genotypes. In the meantime, please look at the refSNP allele list to see if a submitted SNP genotype allele is valid or not. (**2/28/05**)