CSHL SeqTech 2024
Variant Representation Lab - Answers (40 minutes)

1. Consider a genomic sequence (S):

   `GTTATTATC`

   And alternate sequence (S'):

   `GTTATTATTATC`

   What is the variant?

   Answer (one of many possible) *This is an insertion of "`TTA`" or "`TAT`" ? that occurs between X and Y*

2.

   a. How would this be represented in VCF with the following fields:

POS	ID	REF	ALT
1	INS	G	GTTA

   b. Convert this same variant into HGVS notation, characterized as each of the following variant types:

      (Hint: find Recommendations>DNA pages on the hgvs-nomenclature.org website. Pay special attention to the one-line definitions for each term, syntax explanations, and examples.)

      i. A duplication

         Answer: `S:g.6_8dup`

      ii. An insertion

         Answer: `S:g.8_9insTAT`

      iii. A Deletion-Insertion

         Answer: `S:g.8_9delinsTTATC`

   c. Which of the three HGVS variants representations from b i-iii above are correct, and why or why not?

      Answer: it's a duplication, based on the post-hoc order on HGVS variant types, because it is immediately 3' of an identical nucleotide sequence

   d. Suppose someone were to report this variant as the following.

      `S:g.2_3insTAT`

      Is this a correct representation of the variant? Why or why not?

      If they have the same information as us, no, because HGVS arbitrarily assumes that insertions and deletions happen on the 3' most end of ambiguous regions.

e. Based on S and S', convert this variant into SPDI notation.

S:

Hint: there is one unique, unambiguous representation of this variant in SPDI.

Try googling "spdi insertion in ambiguous region"

Answer S:1:TTATTAT:TTATTATTAT

f. Which of the HGVS variant types does the SPDI representation most closely resemble?

Answer: a Delins, because all sequence variants are treated as (perhaps trivial) deletion-insertions in SPDI. This makes variant types in SPDI less 'specific' than in HGVS, but has the advantage of allowing an ambiguous insertion like S' to be represented unambiguously, and without overprecision.

3. The RefSeq identifier for chromosome 11 is `NC_000011.10`. What is the HGVS expression to describe the below VCF record fields?

CHR	POS	ID	REF	ALT
11	68032291	.	C	T

Answer: `NC_000011.10:g.68032291C>T`

4. What is the contextual SPDI expression for this variant?

Hint: Look up the HGVS expression in ClinVar.

Answer: `NC_000011.10:68032290:C:T`

5. Use the ClinGen Allele Registry to find the CAID for HGVS expression `NM_005228.5:c.2573_2574delinsGT`. What is the CAID?

Answer: `CA16602730`

6. What is the associated protein change HGVS expression?

Hint: They can use the links in the blue box to see mapped entries in other resources, which may contain additional details

Answer: `EGFR Leu858Arg` or `EGFR L858R`

7. What is the CIViC molecular profile score for the protein change you identified?

Answer `379`

8. What is the CAID linked to that CIViC record? Is it the same or different from the answer to #4? Why?

Answer `CA126713`

The former was a genomic-level substitution variant. The latter is represented as a protein-level delins.