CSHL SeqTech 2024
Variant Representation Lab (40 minutes)

1. Consider a genomic sequence (S):
   `GTTATTATC`
   And alternate sequence (S'):
   `GTTATTATTATC`
   What is the variant?

2.
   a. How would this be represented in VCF with the following fields:
      `POS  ID  REF  ALT`

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

      ii. An insertion

      iii. A Deletion-Insertion

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

   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?

   e. Based on S and S', convert this variant into SPDI notation (which you can learn about here).
      S:

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

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

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

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

6. What is the associated protein change HGVS expression?

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

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