AlloSeq™ Assign®

How to submit novel alleles to Genbank and IMGT

For use with AlloSeq Assign v1.0

TEC745
Version No: 1.0
Issue Date: June 2022

REF ASA1.0

CareDx Pty Ltd,
20 Collie Street,
Fremantle, WA 6160,
Australia

CareDx Pty Ltd

Table of Contents

1. Purpose

1.1. To submit novel alleles to IMGT they must first be submitted to Genbank, ENA or DDBJ.

1.2. This document describes the steps for generating sequences and feature locations for submission to Genbank.

2. Selecting a good allele candidate for submission

2.1. **The allele is a good candidate if it is completely phased and homozygous.**

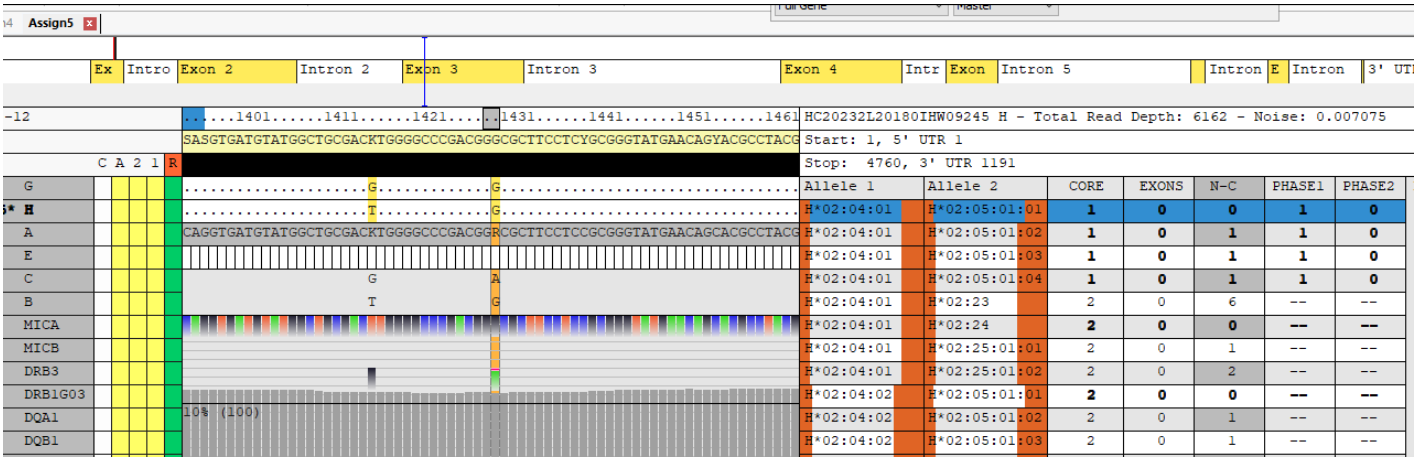


Figure 1. An example of a novel HLA-H allele with fully phased sequence.

2.2. **If the sample is partially phased, a good allele candidate should have identical sequence for both alleles in the region that is not phased.**

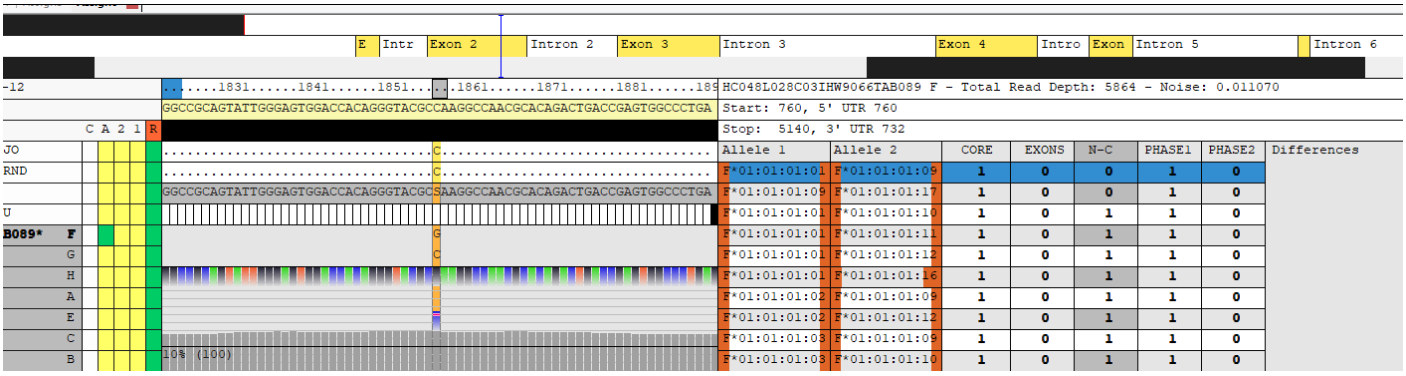


Figure 2. An example of a novel HLA-F allele with partially phased sequence. The sequence for exons 4-7 is homozygous.

2.3. **A good allele candidate may also have complete coverage with a single heterozygous position.**

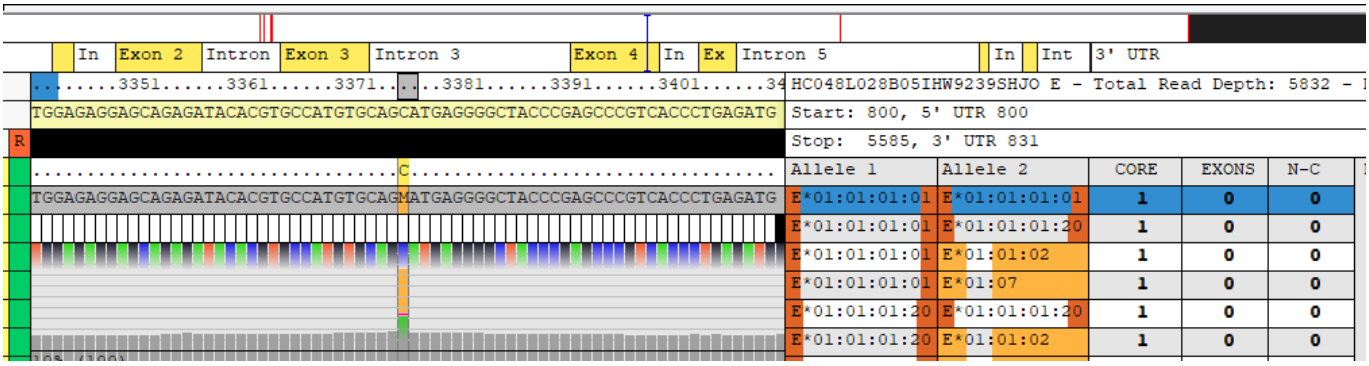


Figure 3. An example of a novel HLA-E sample with a single heterozygous position.

3. Check the sequence for errors

3.1. **Check for any base call errors and make edits as required.**

3.2. **Edit any masked positions.**

3.2.1. Masked positions are excluded from analysis will report as “N” in the fasta sequence report.

3.2.2. Editing these positions where a clear base call can be made is recommended.

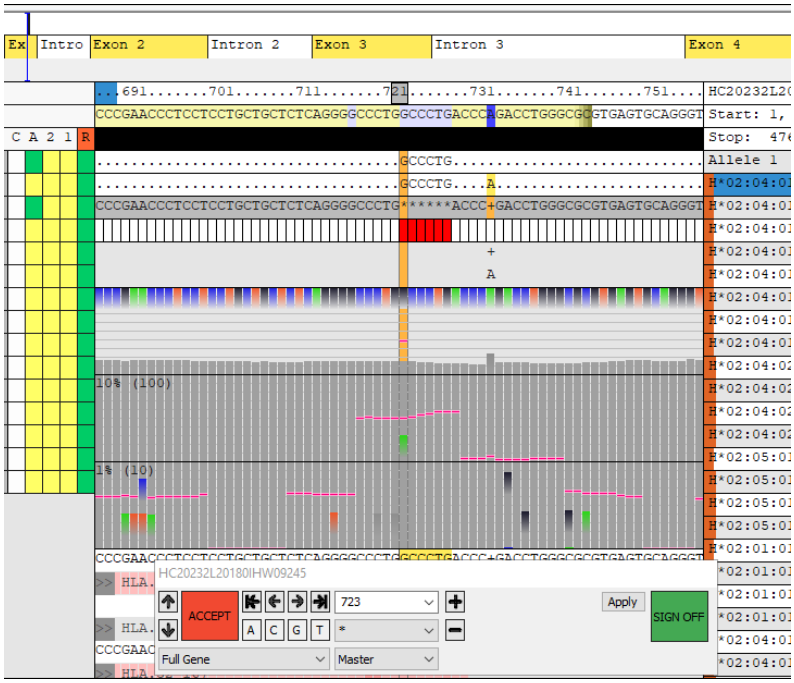


Figure 4. An example of a masked position. Masked positions are indicated by stars (*) in the sample consensus sequence.

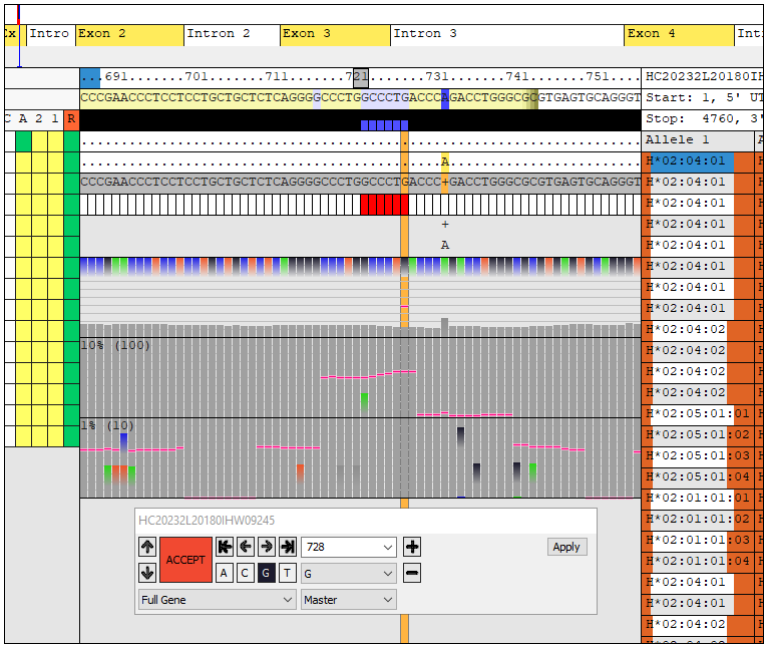


Figure 5.The sequence after editing the masked base calls.

4. Get the fasta sequence

Once you are satisfied with the accuracy of your sequence, use the fasta report function from Assign to generate the fasta sequence for submission.

4.1. **To generate the fasta sequence from Assign:**

4.1.1. Open the reports menu and select the FASTA tab.

4.1.2. Select the sample name from the drop down

4.1.3. Select the group from the dropdown. Choose _All_ to generate the sequence for the full gene, or cDNA to generate the sequence for exons only.

4.1.4. Click generate report.

4.1.5. Enter the name for the file and chose the location to save the file.

NOTE 1: Selecting the locus from the locus dropdown may generate a blank report for some genes. It is recommended to select _All_ in the locus dropdown and copy the sequence for the desired locus to a new file.

NOTE 2: When generating the sequence for the full gene, the start and end of the sequence will be automatically padded with “N’s” due to the masks in the 5’ and 3’ UTRs. These will need to be manually removed before locating the gene features.

Genbank will not accept non-IUPAC codes.

NOTE 3: It is recommended to submit only cDNA sequence for class II MICA and MICB due to limited non-coding sequence available with the AlloSeq Tx 17.1 assay.

4.2. **Fasta sequence output from AlloSeq Assign**

4.2.1. The fasta file output from Assign should be opened with a text editor such as Notepad.

4.2.2. The fasta file will contain sequences for each sample and gene as selected in standard fasta format, which consists of the header

>SampleIdentifier_IMGT/gene_Phase #

And the sequence starting on the following line.

NOTE: Update the header as desired. A suggested format for the novel sequence header is

>SampleID gene*##:##:new

Where ##:##:new is the local name for the novel allele.

4.2.3. A single sequence will be generated for homozygous loci.

4.2.4. For heterozygous samples 3 sequences will be generated for 1 locus:

1. Master- the sample consensus sequence with insertions expanded.
2. Phase 1- the sequence for phase layer 1, with insertions indicated by "+", homozygous deletions indicated as "- "and heterozygous deletions indicated in lowercase. This may correspond to allele 1 or 2, depending on how the phase track has been built.
3. Phase 2- the sequence for phase layer 2, with insertions indicated by "+", homozygous deletions indicated as "- "and heterozygous deletions indicated in lowercase. This may correspond to allele 1 or 2, depending on how the phase track has been built.

5. Indels in fasta sequence outputs from Assign

5.1. Due to how indels are handled in Assign, in certain cases it is necessary to edit indel locations in the fasta sequence outputs.

5.1.1. Figure 6 shows an example of a homozygous deletion in the Assign coverage pane.

Figure 6. Sample with homozygous deletion in Assign.

5.1.2. Figure 7 shows the fasta sequence output for the same sample. Note that the master layer sequence does not include the deletion, but phase 1 and phase 2 indicate the deletion as three dashes (---). These dashes will need to be deleted from the fasta sequence before submission.

>HC20223L20174C02163UCLA1000_IMGT/DQA1
ATGATCCTAAACAAAGCTCTGMTGCTGGGGGCCCTYGCCCTGACCACCGTGATGAGCCCYTGTGGAGGTGAAGACATTGTGGCTGACCAYGTTGCCTCTT
AYGGTGTAAACTTGTACCAGTCTTACGGTCCCTCTGGCCAGTTCACCCATGAATTTGATGGAGACGAGSAGTTCTAYGTGGACCTGGRGARGAAGGAGAC
TGTCTGGWRKTTGCCTSTKYTCMRMMRAYTTAGATTTGACCCGCAATTTGCACTGACAAACATCGCTGTGMYAAAACAYAACTTGAACATCCTGATTAAA
CGCTCCAACTCTACCGCTGCTACCAATGAGGTTCCTGAGGTCACAGTGTTTTCCAAGTCTCCCGTGACRCTGGGTCAGCCCAACACCCTCATCTGTCTTG
TGGACAACATCTTTCCTCCTGTGGTCAACATCACMTGGCTGAGCAATGGGCACTCAGTCACAGAAGGTGTTTCTGAGACCAGCTTCCTCTCCAAGAGTGA
TCATTCCTTCTTCAAGATCAGTTACCTCACCTTCCTCCCTTCTGCTGATGAGATTTATGACTGCAAGGTGGAGCACTGGGGCCTGGAYGAGCCTCTTCTG
AAACACTGGGAGCCTGAGATTCCAGCMCCTATGTCAGAGCTCACAGAGACTGTGGTCTGYGCCCTGGGRTTGTCTGTGGGCCTCGTGGGCATTGTGGTGG
GSACYGTCTTSATCATCCGAGGCCTGCGTTCAGTTGGTGCTTCCAGACACCAAGGGCCCTTGTGA
>HC20223L20174C02163UCLA1000_IMGT/DQA1_Phase 1
ATGATCCTAAACAAAGCTCTGATGCTGGGGGCCCTCGCCCTGACCACCGTGATGAGCCCTTGTGGAGGTGAAGACATTGTGGCTGACCACGTTGCCTCTT
ACGGTGTAAACTTGTACCAGTCTTACGGTCCCTCTGGCCAGTTCACCCATGAATTTGATGGAGACGAGGAGTTCTATGTGGACCTGGAGAGGAAGGAGAC
TGTCTGGAAGTTGCCTCTGTTCCACAGACTTAGA---TTTGACCCGCAATTTGCACTGACAAACATCGCTGTGCTAAAACATAACTTGAACATCCTGATT
AAACGCTCCAACTCTACCGCTGCTACCAATGAGGTTCCTGAGGTCACAGTGTTTTCCAAGTCTCCCGTGACACTGGGTCAGCCCAACACCCTCATCTGTC
TTGTGGACAACATCTTTCCTCCTGTGGTCAACATCACCTGGCTGAGCAATGGGCACTCAGTCACAGAAGGTGTTTCTGAGACCAGCTTCCTCTCCAAGAG
TGATCATTCCTTCTTCAAGATCAGTTACCTCACCTTCCTCCCTTCTGCTGATGAGATTTATGACTGCAAGGTGGAGCACTGGGGCCTGGATGAGCCTCTT
CTGAAACACTGGGAGCCTGAGATTCCAGCACCTATGTCAGAGCTCACAGAGACTGTGGTCTGTGCCCTGGGGTTGTCTGTGGGCCTCGTGGGCATTGTGG
TGGGGACCGTCTTGATCATCCGAGGCCTGCGTTCAGTTGGTGCTTCCAGACACCAAGGGCCCTTGTGA
>HC20223L20174C02163UCLA1000_IMGT/DQA1_Phase 2
ATGATCCTAAACAAAGCTCTGCTGCTGGGGGCCCTTGCCCTGACCACCGTGATGAGCCCCTGTGGAGGTGAAGACATTGTGGCTGACCATGTTGCCTCTT
ATGGTGTAAACTTGTACCAGTCTTACGGTCCCTCTGGCCAGTTCACCCATGAATTTGATGGAGACGAGCAGTTCTACGTGGACCTGGGGAAGAAGGAGAC
TGTCTGGTGTTTGCCTGTTCTCAGACAATTTAGA---TTTGACCCGCAATTTGCACTGACAAACATCGCTGTGACAAAACACAACTTGAACATCCTGATT
AAACGCTCCAACTCTACCGCTGCTACCAATGAGGTTCCTGAGGTCACAGTGTTTTCCAAGTCTCCCGTGACGCTGGGTCAGCCCAACACCCTCATCTGTC
TTGTGGACAACATCTTTCCTCCTGTGGTCAACATCACATGGCTGAGCAATGGGCACTCAGTCACAGAAGGTGTTTCTGAGACCAGCTTCCTCTCCAAGAG
TGATCATTCCTTCTTCAAGATCAGTTACCTCACCTTCCTCCCTTCTGCTGATGAGATTTATGACTGCAAGGTGGAGCACTGGGGCCTGGACGAGCCTCTT
CTGAAACACTGGGAGCCTGAGATTCCAGCCCCTATGTCAGAGCTCACAGAGACTGTGGTCTGCGCCCTGGGATTGTCTGTGGGCCTCGTGGGCATTGTGG
TGGGCACTGTCTTCATCATCCGAGGCCTGCGTTCAGTTGGTGCTTCCAGACACCAAGGGCCCTTGTGA

Figure 7. Fasta output for Assign indicating the locations of the deletions in the 3 layers.

5.1.3. Figure 8 shows an example of a homozygous insertion in the Assign coverage pane.

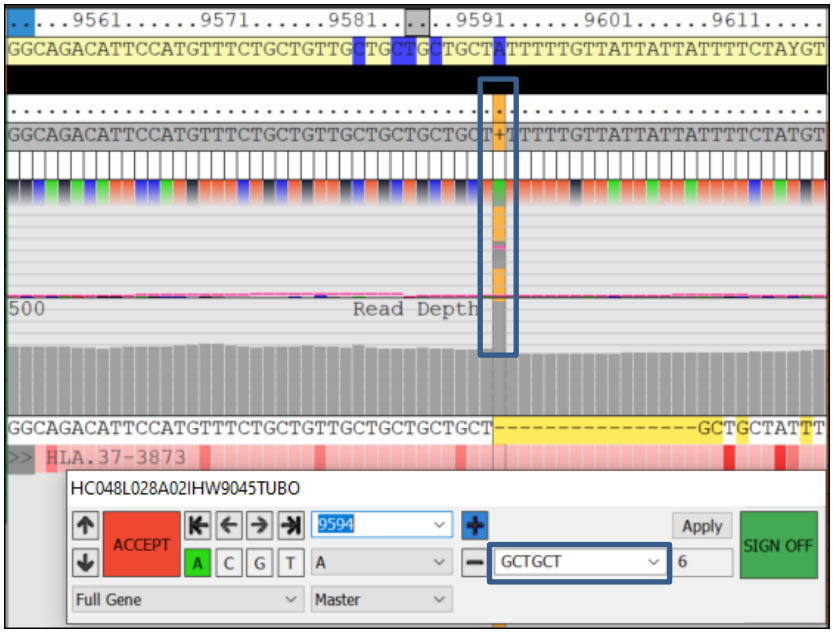


Figure 8. Sample with homozygous insertion in Assign.

5.1.4. Figure 9 shows the fasta sequence output for the same sample. As the sample is homozygous for the gene, a single fasta sequence is output, and the inserted based are expanded. No edits are required for this sequence.

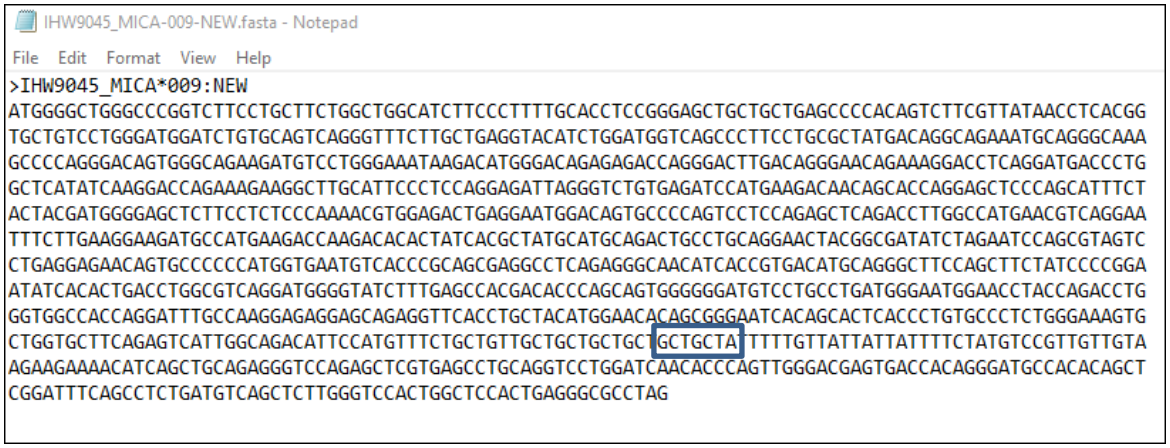


Figure 9. Fasta output from Assign indicating the location of the inserted bases.

5.1.5. Figure 10 shows an example of a heterozygous deletion in the Assign coverage pane.

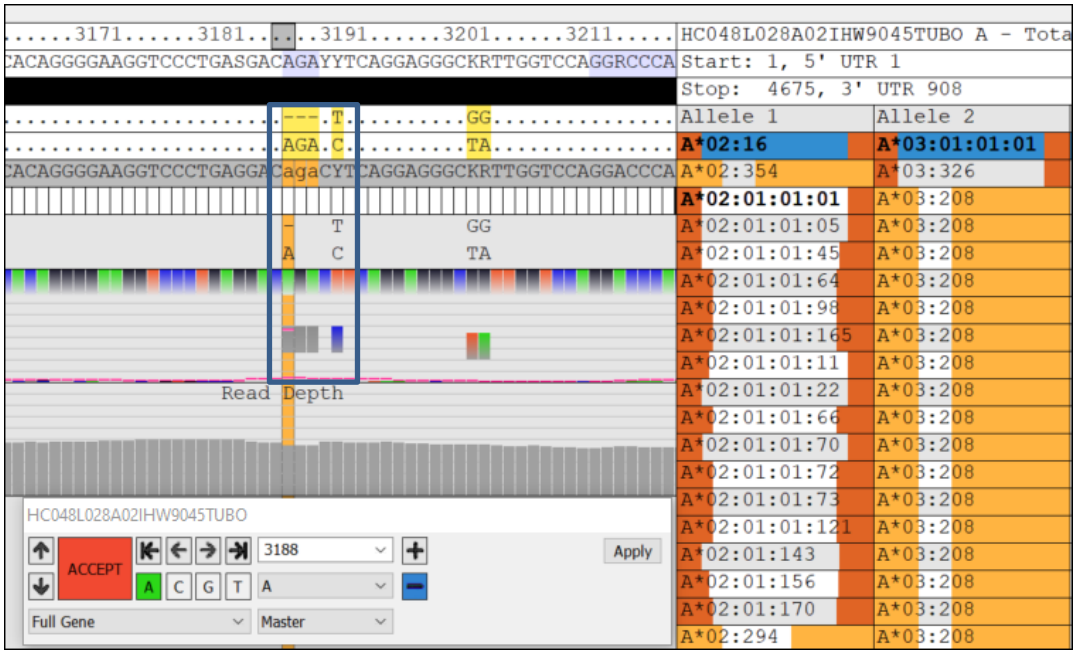


Figure 10. Example of a heterozygous deletion in Assign.

5.1.5.1. Figure 11 shows the fasta output Master layer of the same sample. The heterozygous deletion is indicated in lowercase.

```
>HC048L028A02IHW9045TUBO_IMGT/A
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCAGRAGCAGAGGGGTCAGGGCGAAGTCCCAGGGCCC
CAGGCGTGGCTCTCAGRGTCTCAGGCCCCGAAGGCGGTGTATGGATTGGGGAGTCCCAGCCTTGGGGATTCCCCAACTCCGCAGTTTCTTTTCTCCCTCT
CCCAACCTAYGTAGGGTCCTTCWTCCTGGATACTCACGACGCGGACCCAGTTCTCACTCCCATTGGGTGTCGGGTTTCCAGAGAAGCCAATCAGTGTCGT
CGCGGTCGCKGTTCTAAAGYCCGCACGCACCCACCGGGACTCAGATTCTCCCCAGACGCCGAGGATGGCCGTCATGGCGCCCCGAACCCTCSTCCTGCTA
CTCTCGGGGGCYCTGGCCCTGACCCAGACCTGGGCGGGTGAGTGCGGGGTCGGGAGGGAAACSGCCTCTGYGGGGAGAAGCAASGGGCCCKCCTGGCGGG
GGCGCAGGACCSGGGRAGCCGCGCCGGGASGAGGGTCGGGCRGGTCTCAGCCACTSCTCGYCCCCAGGCTCYCACTCCATGAGGTATTTCTTCACATCCG
TGTCCCGGCCCGGCCGCGGGGAGCCCCGCTTCATCGCMGTGGGCTACGTGGACGACACGCAGTTCGTGCGGTTCGACAGCGACGCCGCGAGCCAGAGGAT
GGAGCCGCGGGCGCCGTGGATAGAGCAGGAGGGKCCGGAGTATTGGGACSRGGAGACACGGAAWGTGAAGGCCCASTCACAGACTSACCGAGTGGACCTG
GGGACCCTGCGCGGCTACTACAACCAGAGCGAGGCCGGTGAGTGACCCCGGCCSGGGGCGCAGGTCASGACCYCTCATCCCCCACGGACGGGCCAGGTCG
CCCACAGTCTCCGGGTCCGAGATCCRCCCCGAAGCCGCGGGACCCCGAGACCCTTGCCCCGGGAGAGGCCCAGGCGCCTTTACCCGGTTTCATTTTCAGT
TTAGGCCAAAAATCCCCCCRGGTTGGTCGGGGCKGGGCGGGGCTCGGGGGACYGGGCTGACCGCGGGGTCSGGGCCAGGTTCTCACACCRTCCAGAKRAT
GTATGGCTGCGACGTGGGGTCGGACKGGCGCTTCCTCCGCGGGTACCRSCAGKACGCCTACGACGGCAAGGATTACATCGCCCTGAAMGAGGACCTGCGC
TCTTGGACCGCGGCGGACATGGCRGCTCAGAYCACCAAGCRCAAGTGGGAGGCGGCCCATGWGGCGGAGCAGTTGAGAGCCTACCTGGAKGGCRMGTGCG
TGGAGTGGCTCCGCAGATACCTGGAGAACGGGAAGGAGACGCTGCAGCGCACGGGTACCAGGGGCCACGGGGCGCCTCCCTGATCGCCTGTAGATCTCCC
GGGCTGGCCTCCCACAAGGAGGGGAGACAATTGGGACCAACACTAGAATATCRCCCTCCCTCTGGTCCTGAGGGAGAGGAATCCTCCTGGGTTtCCAGAT
CCTGTACCAGAGAGTGACTCTGAGGTTCCGCCCTGCTCTCTGACACAATTAAGGGATAAAATCTCTGAAGGARTGACGGGAAGACGATCCCTCGAATACT
GATGAGTGGTTCCCTTTGACACMCACAGGCAGCAGCCTTGGGCCCGTGACTTTTCCTCTCAGGCCTTGTTCTCTGCTTCACACTCAATGTGTGTGGGGGT
CTGAGTCCAGCACTTCTGAGTCCYTCAGCCTCCACTCAGGTCAGGACCAGAAGTCGCTGTTCCCTYYTCAGGGAMTAGAATTTTCCACGGAATAGGAGAT
TATCCCAGGTGCCTGTGTCCAGGCTGGTGTCTGGGTTCTGTGCTCYCTTCCCCATCCCRGGTGTCCTGTCCATTCTCAAGATRGCCACATGYGTGCTGGW
GGAGTGTCCCATGACAGATGCAAAATGCCTGAATKWTCTGACTCTTCCYGWCAGACSCCCCCAARACRCATATGACYCACCACSCYRTCTCTGACCATGA
RGCCACCCTGAGGTGCTGGGCCCTGRGCTTCTACCCTGCGGAGATCACACTGACCTGGCAGCGGGATGGGGAGGACCAGACCCAGGACACGGAGCTCGTG
GAGACCAGGCCTGCAGGGGATGGAACCTTCCAGAAGTGGGCGGCTGTGGTGGTGCCTTCTGGASAGGAGCAGAGATACACCTGCCATGTGCAGCATGAGG
GTYTGCCCAAGCCCCTCACCCTGAGATGGGGTAAGGAGGGAGAYGGGGGTGTCATGTCTYTTAGGGAAAGCAGGAGCCTCTCTGgagACCTTTAGCAGGG
TCAGGGCCCCTCACCTTCCCCTCTTTTCCCAGAGCYGTCTTCCCAGCCCACCATCCCCATCGTGGGCATCATTGCTGGCCTGGTTCTCYTTGGAGCTGTG
ATCACTGGAGCTGTGGTCGCTGCYGTGATGTGGAGGAGGAAGAGCTCAGGTGGRGAAGGGGTGAAGGGTGGGgTCTGAGATTTCTTGTCTCACTGAGGGT
TCCAAGMCCCAGSTAGAARTGTGCCCTGYCTCRTTACTGGGAAGCACCRYCCACAATYATGGGCCTACCCAGYCTGGGCCCTGTGTGCCAGCACTTACTC
TTTTGTAAAGCACCTGTTAAAATGAAGGACAGATTTATCACCTTGATTACRGCGGTGATGGGACCTGATCCCAGCAGTCACAAGTCACAGGGGAAGGTCC
CTGAGGACagaCYTCAGGAGGGCKRTTGGTCCAGGACCCACACCTGCTTTCTTCATGTTTCCTGATCCCGCCCTGGGTCTGCAGTCACACATTTCTGGAA
```

Figure 11. Fasta sequence for the master layer indicating the location of the heterozygous deletion.

5.1.5.2. Figure 12 shows the fasta output for phase 1. The view in Assign indicates that phase 1 matches allele 1, so it is expected that the deletion belongs to allele 1. Note that due to how Assign phases deletions, the first base of the deletion is not displayed. If this allele was to be submitted to 2 bases indicated would need to be deleted to make the sequence consistent with the Assign coverage display.

```
>HC048L028A02IHW9045TUBO_IMGT/A_Phase 1
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCAGAAGCAGAGGGGTCAGGGCGAAGTCCCAGGGCCC
CAGGCGTGGCTCTCAGGGTCTCAGGCCCCGAAGGCGGTGTATGGATTGGGGAGTCCCAGCCTTGGGGATTCCCCAACTCCGCAGTTTCTTTTCTCCCTCT
CCCAACCTATGTAGGGTCCTTCTTCCTGGATACTCACGACGCGGACCCAGTTCTCACTCCCATTGGGTGTCGGGTTTCCAGAGAAGCCAATCAGTGTCGT
CGCGGTCGCGGTTCTAAAGTCCGCACGCACCCACCGGGACTCAGATTCTCCCCAGACGCCGAGGATGGCCGTCATGGCGCCCCGAACCCTCGTCCTGCTA
CTCTCGGGGGCTCTGGCCCTGACCCAGACCTGGGCGGGTGAGTGCGGGGTCGGGAGGGAAACGGCCTCTGTGGGGAGAAGCAACGGGCCCGCCTGGCGGG
GGCGCAGGACCCGGGAAGCCGCGCCGGGAGGAGGGTCGGGCGGGTCTCAGCCACTCCTCGTCCCCAGGCTCTCACTCCATGAGGTATTTCTTCACATCCG
TGTCCCGGCCCGGCCGCGGGGAGCCCCGCTTCATCGCAGTGGGCTACGTGGACGACACGCAGTTCGTGCGGTTCGACAGCGACGCCGCGAGCCAGAGGAT
GGAGCCGCGGGCGCCGTGGATAGAGCAGGAGGGTCCGGAGTATTGGGACGGGGAGACACGGAAAGTGAAGGCCCACTCACAGACTCACCGAGTGGACCTG
GGGACCCTGCGCGGCTACTACAACCAGAGCGAGGCCGGTGAGTGACCCCGGCCCGGGGCGCAGGTCACGACCTCTCATCCCCCACGGACGGGCCAGGTCG
CCCACAGTCTCCGGGTCCGAGATCCGCCCCGAAGCCGCGGGACCCCGAGACCCTTGCCCCGGGAGAGGCCCAGGCGCCTTTACCCGGTTTCATTTTCAGT
TTAGGCCAAAAATCCCCCCAGGTTGGTCGGGGCGGGGCGGGGCTCGGGGGACCGGGCTGACCGCGGGGTCCGGGCCAGGTTCTCACACCGTCCAGAGGAT
GTATGGCTGCGACGTGGGGTCGGACTGGCGCTTCCTCCGCGGGTACCACCAGTACGCCTACGACGGCAAGGATTACATCGCCCTGAAAGAGGACCTGCGC
TCTTGGACCGCGGCGGACATGGCAGCTCAGACCACCAAGCACAAGTGGGAGGCGGCCCATGTGGCGGAGCAGTTGAGAGCCTACCTGGAGGGCGAGTGCG
TGGAGTGGCTCCGCAGATACCTGGAGAACGGGAAGGAGACGCTGCAGCGCACGGGTACCAGGGGCCACGGGGCGCCTCCCTGATCGCCTGTAGATCTCCC
GGGCTGGCCTCCCACAAGGAGGGGAGACAATTGGGACCAACACTAGAATATCGCCCTCCCTCTGGTCCTGAGGGAGAGGAATCCTCCTGGGTTTCCAGAT
CCTGTACCAGAGAGTGACTCTGAGGTTCCGCCCTGCTCTCTGACACAATTAAGGGATAAAATCTCTGAAGGAATGACGGGAAGACGATCCCTCGAATACT
GATGAGTGGTTCCCTTTGACACACACAGGCAGCAGCCTTGGGCCCGTGACTTTTCCTCTCAGGCCTTGTTCTCTGCTTCACACTCAATGTGTGTGGGGGT
CTGAGTCCAGCACTTCTGAGTCCTTCAGCCTCCACTCAGGTCAGGACCAGAAGTCGCTGTTCCCTCTTCAGGGACTAGAATTTTCCACGGAATAGGAGAT
TATCCCAGGTGCCTGTGTCCAGGCTGGTGTCTGGGTTCTGTGCTCCCTTCCCCATCCCAGGTGTCCTGTCCATTCTCAAGATAGCCACATGTGTGCTGGA
GGAGTGTCCCATGACAGATGCAAAATGCCTGAATGATCTGACTCTTCCTGACAGACGCCCCCAAAACGCATATGACTCACCACGCTGTCTCTGACCATGA
AGCCACCCTGAGGTGCTGGGCCCTGAGCTTCTACCCTGCGGAGATCACACTGACCTGGCAGCGGGATGGGGAGGACCAGACCCAGGACACGGAGCTCGTG
GAGACCAGGCCTGCAGGGGATGGAACCTTCCAGAAGTGGGCGGCTGTGGTGGTGCCTTCTGGACAGGAGCAGAGATACACCTGCCATGTGCAGCATGAGG
GTTTGCCCAAGCCCCTCACCCTGAGATGGGGTAAGGAGGGAGACGGGGGTGTCATGTCTTTTAGGGAAAGCAGGAGCCTCTCTGagACCTTTAGCAGGGT
CAGGGCCCCTCACCTTCCCCTCTTTTCCCAGAGCCGTCTTCCCAGCCCACCATCCCCATCGTGGGCATCATTGCTGGCCTGGTTCTCTTTGGAGCTGTGA
TCACTGGAGCTGTGGTCGCTGCTGTGATGTGGAGGAGGAAGAGCTCAGGTGGGGAAGGGGTGAAGGGTGGGTCTGAGATTTCTTGTCTCACTGAGGGTTC
CAAGACCCAGGTAGAAGTGTGCCCTGCCTCGTTACTGGGAAGCACCACCCACAATTATGGGCCTACCCAGCCTGGGCCCTGTGTGCCAGCACTTACTCTT
TTGTAAAGCACCTGTTAAAATGAAGGACAGATTTATCACCTTGATTACAGCGGTGATGGGACCTGATCCCAGCAGTCACAAGTCACAGGGGAAGGTCCCT
GAGGACgaCTTCAGGAGGGCGGTTGGTCCAGGACCCACACCTGCTTTCTTCATGTTTCCTGATCCCGCCCTGGGTCTGCAGTCACACATTTCTGGAAACT
```

Figure 12. Fasta sequence output for phase 1 indicating the location of the deletion.

5.1.5.3. Figure 13 shows the fasta output for phase 2. Note that the bases indicated match the sequence for allele 2 in Assign. These bases would need to be capitalised before submitting this allele.

```
>HC048L028A02IHW9045TUBO_IMGT/A_Phase 2
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCAGGAGCAGAGGGGTCAGGGCGAAGTCCCAGGGCCC
CAGGCGTGGCTCTCAGAGTCTCAGGCCCCGAAGGCGGTGTATGGATTGGGGAGTCCCAGCCTTGGGGATTCCCCAACTCCGCAGTTTCTTTTCTCCCTCT
CCCAACCTACGTAGGGTCCTTCATCCTGGATACTCACGACGCGGACCCAGTTCTCACTCCCATTGGGTGTCGGGTTTCCAGAGAAGCCAATCAGTGTCGT
CGCGGTCGCTGTTCTAAAGCCCGCACGCACCCACCGGGACTCAGATTCTCCCCAGACGCCGAGGATGGCCGTCATGGCGCCCCGAACCCTCCTCCTGCTA
CTCTCGGGGGCCCTGGCCCTGACCCAGACCTGGGCGGGTGAGTGCGGGGTCGGGAGGGAAACCGCCTCTGCGGGGAGAAGCAAGGGGCCCTCCTGGCGGG
GGCGCAGGACCGGGGAGCCGCGCCGGGACGAGGGTCGGGCAGGTCTCAGCCACTGCTCGCCCCCAGGCTCCCACTCCATGAGGTATTTCTTCACATCCG
TGTCCCGGCCCGGCCGCGGGGAGCCCCGCTTCATCGCCGTGGGCTACGTGGACGACACGCAGTTCGTGCGGTTCGACAGCGACGCCGCGAGCCAGAGGAT
GGAGCCGCGGGCGCCGTGGATAGAGCAGGAGGGGCCGGAGTATTGGGACCAGGAGACACGGAATGTGAAGGCCCAGTCACAGACTGACCGAGTGGACCTG
GGGACCCTGCGCGGCTACTACAACCAGAGCGAGGCCGGTGAGTGACCCCGGCCGGGGGCGCAGGTCAGGACCCCTCATCCCCCACGGACGGGCCAGGTCG
CCCACAGTCTCCGGGTCCGAGATCCACCCCGAAGCCGCGGGACCCCGAGACCCTTGCCCCGGGAGAGGCCCAGGCGCCTTTACCCGGTTTCATTTTCAGT
TTAGGCCAAAAATCCCCCCGGGTTGGTCGGGGCTGGGCGGGGCTCGGGGGACTGGGCTGACCGCGGGGTCGGGGCCAGGTTCTCACACCATCCAGATAAT
GTATGGCTGCGACGTGGGGTCGGACGGGCGCTTCCTCCGCGGGTACCGGCAGGACGCCTACGACGGCAAGGATTACATCGCCCTGAACGAGGACCTGCGC
TCTTGGACCGCGGCGGACATGGCGGCTCAGATCACCAAGCGCAAGTGGGAGGCGGCCCATGAGGCGGAGCAGTTGAGAGCCTACCTGGATGGCACGTGCG
TGGAGTGGCTCCGCAGATACCTGGAGAACGGGAAGGAGACGCTGCAGCGCACGGGTACCAGGGGCCACGGGGCGCCTCCCTGATCGCCTGTAGATCTCCC
GGGCTGGCCTCCCACAAGGAGGGGAGACAATTGGGACCAACACTAGAATATCACCCTCCCTCTGGTCCTGAGGGAGAGGAATCCTCCTGGGTTCCAGATC
CTGTACCAGAGAGTGACTCTGAGGTTCCGCCCTGCTCTCTGACACAATTAAGGGATAAAATCTCTGAAGGAGTGACGGGAAGACGATCCCTCGAATACTG
ATGAGTGGTTCCCTTTGACACCGGCAGCAGCCTTGGGCCCGTGACTTTTCCTCTCAGGCCTTGTTCTCTGCTTCACACTCAATGTGTGTGGGGGTCTGAG
TCCAGCACTTCTGAGTCCCTCAGCCTCCACTCAGGTCAGGACCAGAAGTCGCTGTTCCCTTCTCAGGGAATAGAAGATTATCCCAGGTGCCTGTGTCCAG
GCTGGTGTCTGGGTTCTGTGCTCTCTTCCCCATCCCGGGTGTCCTGTCCATTCTCAAGATGGCCACATGCGTGCTGGTGGAGTGTCCCATGACAGATGCA
AAATGCCTGAATTTTCTGACTCTTCCCGTCAGACCCCCCAAGACACATATGACCCACCACCCCATCTCTGACCATGAGGCCACCCTGAGGTGCTGGGCC
CTGGGCTTCTACCCTGCGGAGATCACACTGACCTGGCAGCGGGATGGGGAGGACCAGACCCAGGACACGGAGCTCGTGGAGACCAGGCCTGCAGGGGATG
GAACCTTCCAGAAGTGGGCGGCTGTGGTGGTGCCTTCTGGAGAGGAGCAGAGATACACCTGCCATGTGCAGCATGAGGGTCTGCCCAAGCCCCTCACCCT
GAGATGGGGTAAGGAGGGAGATGGGGGTGTCATGTCTCTTAGGGAAAGCAGGAGCCTCTCTGGagACCTTTAGCAGGGTCAGGGCCCCTCACCTTCCCCT
CTTTTCCCAGAGCTGTCTTCCCAGCCCACCATCCCCATCGTGGGCATCATTGCTGGCCTGGTTCTCCTTGGAGCTGTGATCACTGGAGCTGTGGTCGCTG
CCGTGATGTGGAGGAGGAAGAGCTCAGGTGGAGAAGGGGTGAAGGGTGGGGTCTGAGATTTCTTGTCTCACTGAGGGTTCCAAGCCCCAGCTAGAAATGT
GCCCTGTCTCATTACTGGGAAGCACCGTCCACAATCATGGGCCTACCCAGTCTGGGCCCTGTGTGCCAGCACTTACTCTTTTGTAAAGCACCTGTTAAAA
TGAAGGACAGATTTATCACCTTGATTACGGCGGTGATGGGACCTGATCCCAGCAGTCACAAGTCACAGGGGAAGGTCCCTGAGGACAgaCCTCAGGAGGG
```

Figure 13. Fasta sequence output for phase 2 indicating the location of the deletion.

5.1.6. Figure 14 shows an example of a heterozygous insertion in the Assign coverage pane.

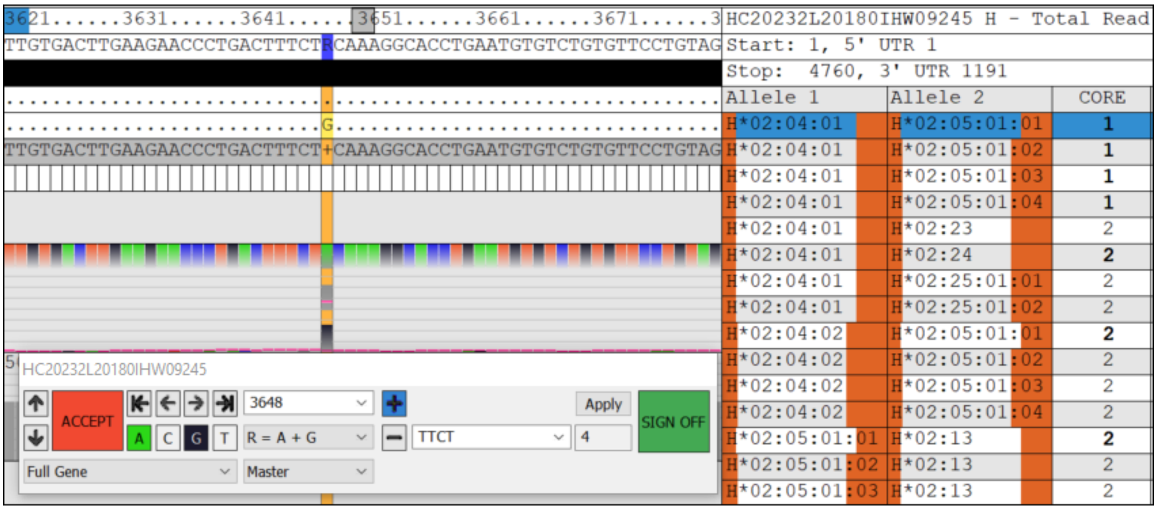


Figure 14. Example of a heterozygous insertion in Assign.

5.1.7. Figure 15 shows the fasta output for the master layer for the same sample. The R indicates the A+G at position 3648, the inserted bases are expanded.

```
>HC20232L20180IHW09245_IMGT/H
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGCACAGGAGGAGCGGGTCAGGGCGAAGT
CCCAGGGCCCCAGGCGTGGCTCTCAGGGTCTCAGGCCCCGAAGGCGGTGTATGGATTGGGGAKGCCCCGCCTTRGGGATTCGCCACCTCCGCAGTTTCTC
TTCTTCTCACAACCTGCGACGGGTCCTTYTTCCTKGATACTCACGAAGCGGRCACAGTTCTCATTCCCACTAGGTGTCGGGTTTCTAGAGAAGCCAATCG
GTGCCGCCGCGGTCCCGGTTCTAAAGTCCCCACGCACCCACCGGGACTCAGATTCTCCCCAGACGCCGAGGATGGTGCTCATGGCGCCCCGAACCCTCCT
CCTGCTGCTCTCAGGGGCCCTGGCCCTGACCCTGACCCAGACCTGGGCGCGTGAGTGCAGGGTCTGCAGGGAAATGGTCGGGAGGAGCGAGGGGCCCGCC
CGGCSGGGGCGCAGGACCCRGGGAKCCGCGCAGGGAGGAGGGTCGGGCSGGTCTCAGCTCCTCCTCGCTCCCAGGCTCCCACTCCATGAGGTATTTCTAC
ACCACCATGTCCCGGCCCGGCCGCGGGGAGCCCCGCTTCATCTCCGTCGGCTACGTGGACGATACGCAGTTCGTGCGGTTCGACAGCGACGACGCGAGTC
CGAGAGAGGAGCCGCGGGCGCCGTGGATGGAGCGGGAGGGGCCRGAGTATTGGGACCGGAACACACAGATCTGCAAGGCCCARGCACRGACTGAACGAGA
GAACCTGCGGATCGCGCTCCGCTACTACAACCAGAGCGAGGGCGGTGAGTGACCCCGGCCCGGGGCGCAGGTCACGACCCCTCCCCATCCCCCACGGAGG
GCCGGGTCGCCTCGAGTCTCTGGGTCCGAGATCCWCCCCGAAACCGCGGGACCCCGAGACCCTTGACCTGGGAGAGGCCCAGGCGCCTTTACCCGGTTTC
ATTTTCAGTTTAGGCCAAAAYCCCCGCGGGTTGGTCGGGGCAGGGCGGGGCTCGGGGGACCGGGCTGACCGCGGGGGCGGGGCCAGGTTCTCACACCATG
CAGGTGATGTATGGCTGCGACKTGGGGCCCGACGGRCGCTTCCTCCGCGGGTATGAACAGCACGCCTACGACGGCAAGGATTACATCGCYCTGAACGAGG
ACCTGCGCTCCTGGACCGCGGCGGACATGGCAGCTCAGATCACCAAGCGCAAGTGGGAGGCGGCCCGTCGGGCGGAGCAGCKGAGAGYCTACCTGGAGGG
CGAGTTCGTGGAGTGGCTCCGCAGATACCTGGAGAACGGGAAGGAGACGCTGCAGCGCGCGGGTACCAGGGGCCACAGGGCGCCTCCCGGATSGYCTGTA
GATCTCCGGGGCTGGCCTCCCACAAGAAAGGGAGACAAATGGACCAACACTATAATATCGCCCTCCCTCTGGTCYTGAGGGAGAAGAATCCTCCTGGGT
TTCCAGAGAGTGACTCTGAGGGTCCGCCSTGCTCTCTGACACAATTAAGGGATGAAATCTSTGAGGAAATGAAGGGAAGACAATCCCTGGAATACTGATG
AGTGGTTCCCTTTGACACACACYGGCAGCAGCCTTGGGCCCCGTGACTTTTCCTCTCAGGCCTTGTTCTCTGCTTCACACTCAATGTGCSTGGGGGTCTG
AGTCCAGCTCTTCTGAGTCCCTCAGCCTCCACTCAGGTCAGGACCAGAAGTCGCTGTTCCCTCYTCAGGGACTAGAATTTTCCACGGAATAGGAGATTAT
CCCAGGTGCCTGTGTCCAGGCTGTTGTCTGGGTTCTGTGCTCCCTTCCCCACCCCAGGCRTCCTGTCCATTCTCAAGATGGCCACATGCGTGCTGGWGGA
GTGTCCCATGACAGATGCAAAATGCCTGAATTTTCTGACTCTTCCCGTCAGACCCCCCCCAAGACACATATGACCCACCACCCCATCTCTGACCATGAGG
CCACCCTGAGGTGCTGGGCCCTGGGCTTCTACCCTGCGGAGATCACACTGACCTGGCAGCGGGATGGGGAGGACCAGACCCASGACACGGAGCTCGTGGA
GACCAGGCCTGCAGGGGATGGAACCTTCCAGAAGTGGGCGGCTGTGGTGGTGCCTTCTGGAGAGGAGCAGAGATACACCTGCCATGTGCAGCATGAGGGT
CTGCCMGAGCCCCTCACCCTGAGATGGGGTAAGGAGGGAGATGGGGGTGTCATGTCYCTTAGGGAAAGCCRGAGCCTCTCTGGAGAGCTTTAGCAGGGTC
AGGGTCCCTCACCTTCCCCCCTTTTCCCAGAGCCATCTTCCCAGCCCAMCRTCCCCATCGTGGGCATCGTTGCTGGCCTGGTTCTACTTGTAGCTGTGGT
CACTGGAGCTGTGGTCGCTGCTGTAATGTGGAGGAAGAAGAGCTCAGGTAAGGAAGGGGTGAGGAGTGTGGTCTGAGATTTCTTGTCTCACTGAGAGTTC
CAAGCCCCAGGTAGAAGTGCCCTGCCTGGTTACTGGGAAGCACCATCCACACTCATGGGCCTACCCAGCCTGGGCCSTGTGTGCCAGCACTTACTCTTTT
GTAAAGCACCTGTTACAATGAGGGACAGATTTATCACCTTGATGACTGTGGTGATGGGACCTGATCCCAGCAGTCACAAGTCACAGGGGAAGGTCCCCRA
GGACAGACCTCAGAAGGGCGGTTGGTCCAGGACCCAYATCTGCTTTCYTCATGTTTCCTGATCCCRCCCTGGGTCTGCAGTTGCACATTTCTGGAAACTT
CTCTGGGGTCCRAGACTTGGAGGTTCCTCTAGGACCTTATGGCCCTGGCTTCTTTCTGGCATCTCACAGGACATTTTCTTCCCACAGATAGAAAAGGAGG
GAGCTACTCTCAGGCTGCAAGTAAGTATGAAGGAGGCTGATCCCTGAAATCCTTTGGATATTGTGTTTGGGAGCCCATGGGGGAGCTCACCCACCCCACA
ATTCTTCCTCTAGCCACATCTACTGTGGGATCTGACCAGGTCCTGTTTTTATTCTACTCSAGGCRGCAACAGTGCCCAGGGCTCTGATGTGTCTCTCACG
GCKTGAAAGGTGAGACCTTGGGGGGCCTGATGTGTGGGGGRTGTTGGGGGGGAACAGTGGACACAGCTGTGCTATGGGGTTCTTTGAATTTGATGTTTTG
AGCRTGCGATGGGCTGCCAAAGTGTCATCCATTACTGGGACAGATATGAATTTGTTCATGAATATTTTTTCTATAGTGTGAGACAGCTGCCTTGTGTGGG
ACTGAGAGGCAAGAKTTGTTCMYRCCTTCCCTTTGTGACTTGAAGAACCCTGACTTTCTTTCTRCAAAGGCACCTGAATGTGTCTGTGTTCCTGTAGGCA
TAATGTGTGGAGGAGGGGAGACCAACCCACCCTCATGTCCACCATGACCCTCTTCCCCACGCTGATCTGTGTTCCCTCCCCAATCATCTTTCCTGTTCCA
```

Figure 15. Fasta sequence output for the master layer indicating the location of the insertion.

5.1.8. Figure 16 shows the fasta output for phase 1. From the Assign coverage pane it can be determined that the insertions belongs to phase 1, so this sequence would require editing to include the inserted bases before submission.

```
>HC20232L20180IHW09245_IMGT/H_Phase 1
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGCACAGGAGGAGCGGGGTCAGGGCGAAGT
CCCAGGGCCCAGGCGTGGCTCTCAGGGTCTCAGGCCCCGAAGGCGGTGTATGGATTGGGGATGCCCCGCCTTAGGATTCGCCACCTCCGCAGTTTCTC
TTCTTCTCACAACCTGCGACGGGTCCTTCTTCCTTGATACTCACGAAGCGGACACAGTTCTCATTCCCACTAGGTGTCGGGTTTCTAGAGAAGCCAATCG
GTGCCGCCGCGGTCCCGGTTCTAAAGTCCCCACGCACCCACCGGGACTCAGATTCTCCCCAGACGCCGAGGATGGTGCTCATGGCGCCCCGAACCCTCCT
CCTGCTGCTCTCAGGGGCCCTGGCCCTGACCCTGACCCAGACCTGGGCGCGTGAGTGCAGGGTCTGCAGGGAAATGGTCGGGAGGAGCGAGGGGCCCGCC
CGGCGGGGGCGCAGGACCCGGGGATCCGCGCAGGGAGGAGGGTCGGGCCGGTCTCAGCTCCTCCTCGCTCCCAGGCTCCCACTCCATGAGGTATTTCTAC
ACCACCATGTCCCGGCCCGGCCGCGGGGAGCCCCGCTTCATCTCCGTCGGCTACGTGGACGATACGCAGTTCGTGCGGTTCGACAGCGACGACGCGAGTC
CGAGAGAGGAGCCGCGGGCGCCGTGGATGGAGCGGGAGGGGCCAGAGTATTGGGACCGGAACACACAGATCTGCAAGGCCCAGGCACAGACTGAACGAGA
GAACCTGCGGATCGCGCTCCGCTACTACAACCAGAGCGAGGGCGGTGAGTGACCCCGGCCCGGGGCGCAGGTCACGACCCCTCCCCATCCCCCACGGAGG
GCCGGGTCGCCTCGAGTCTCTGGGTCCGAGATCCACCCCGAAACCGCGGGACCCCGAGACCCTTGACCTGGGAGAGGCCCAGGCGCCTTTACCCGGTTTC
ATTTTCAGTTTAGGCCAAAACCCCCGCGGGTTGGTCGGGGCAGGGCGGGGCTCGGGGGACCGGGCTGACCGCGGGGGCGGGGCCAGGTTCTCACACCATG
CAGGTGATGTATGGCTGCGACGTGGGGCCCGACGGACGCTTCCTCCGCGGGTATGAACAGCACGCCTACGACGGCAAGGATTACATCGCCCTGAACGAGG
ACCTGCGCTCCTGGACCGCGGCGGACATGGCAGCTCAGATCACCAAGCGCAAGTGGGAGGCGGCCCGTCGGGCGGAGCAGCTGAGAGCCTACCTGGAGGG
CGAGTTCGTGGAGTGGCTCCGCAGATACCTGGAGAACGGGAAGGAGACGCTGCAGCGCGCGGGTACCAGGGGCCACAGGGCGCCTCCCGGATCGCCTGTA
GATCTCCGGGGCTGGCCTCCCACAAGAAAGGGAGACAAATGGGACCAACACTATAATATCGCCCTCCCTCTGGTCTTGAGGGAGAAGAATCCTCCTGGGT
TTCCAGAGAGTGACTCTGAGGGTCCGCCCTGCTCTCTGACACAATTAAGGGATGAAATCTGTGAGGAAATGAAGGGAAGACAATCCCTGGAATACTGATG
AGTGGTTCCCTTTGACAC+GGCAGCAGCCTTGGGCCCCGTGACTTTTCCTCTCAGGCCTTGTTCTCTGCTTCACACTCAATGTGCGTGGGGGTCTGAGTC
CAGCTCTTCTGAGTCCCTCAGCCTCCACTCAGGTCAGGACCAGAAGTCGCTGTTCCCTCTTCAGGGACTAGAATTTTCCACGGAATAGGAGATTATCCCA
GGTGCCTGTGTCCAGGCTGTTGTCTGGGTTCTGTGCTCCCTTCCCCACCCCAGGCATCCTGTCCATTCTCAAGATGGCCACATGCGTGCTGGAGGAGTGT
CCCATGACAGATGCAAAATGCCTGAATTTTCTGACTCTTCCCGTCAGACCCCCCCAAGACACATATGACCCACCACCCCATCTCTGACCATGAGGCCAC
CCTGAGGTGCTGGGCCCTGGGCTTCTACCCTGCGGAGATCACACTGACCTGGCAGCGGGATGGGGAGGACCAGACCCAGGACACGGAGCTCGTGGAGACC
AGGCCTGCAGGGGATGGAACCTTCCAGAAGTGGGCGGCTGTGGTGGTGCCTTCTGGAGAGGAGCAGAGATACACCTGCCATGTGCAGCATGAGGGTCTGC
CCGAGCCCCTCACCCTGAGATGGGGTAAGGAGGGAGATGGGGGTGTCATGTCTCTTAGGGAAAGCCGGAGCCTCTCTGGAGAGCTTTAGCAGGGTCAGGG
TCCCTCACCTTCCCCCCTTTTCCCAGAGCCATCTTCCCAGCCCAACGTCCCCATCGTGGGCATCGTTGCTGGCCTGGTTCTACTTGTAGCTGTGGTCACT
GGAGCTGTGGTCGCTGCTGTAATGTGGAGGAAGAAGAGCTCAGGTAAGGAAGGGGTGAGGAGTGTGGTCTGAGATTTCTTGTCTCACTGAGAGTTCCAAG
CCCCAGGTAGAAGTGCCCTGCCTGGTTACTGGGAAGCACCATCCACACTCATGGGCCTACCCAGCCTGGGCCGTGTGTGCCAGCACTTACTCTTTTGTAA
AGCACCTGTTACAATGAGGGACAGATTTATCACCTTGATGACTGTGGTGATGGGACCTGATCCCAGCAGTCACAAGTCACAGGGGAAGGTCCCCAAGGAC
AGACCTCAGAAGGGCGGTTGGTCCAGGACCCATATCTGCTTTCTTCATGTTTCCTGATCCCACCCTGGGTCTGCAGTTGCACATTTCTGGAAACTTCTCT
GGGGTCCGAGACTTGGAGGTTCCTCTAGGACCTTATGGCCCTGGCTTCTTTCTGGCATCTCACAGGACATTTTCTTCCCACAGATAGAAAAGGAGGGAGC
TACTCTCAGGCTGCAAGTAAGTATGAAGGAGGCTGATCCCTGAAATCCTTTGGATATTGTGTTTGGGAGCCCATGGGGGAGCTCACCCACCCCACAATTC
TTCCTCTAGCCACATCTACTGTGGGATCTGACCAGGTCCTGTTTTTATTCTACTCGAGGCAGCAACAGTGCCCAGGGCTCTGATGTGTCTCTCACGGCTT
GAAAGGTGAGACCTTGGGGGGCCTGATGTGTGGGGGATGTTGGGGGGGAACAGTGGACACAGCTGTGCTATGGGGTTCTTTGAATTTGATGTTTTGAGCA
TGCGATGGGCTGCCAAAGTGTCATCCATTACTGGGACAGATATGAATTTGTTCATGAATATTTTTTCTATAGTGTGAGACAGCTGCCTTGTGTGGGACTG
AGAGGCAAGAGTTGTTCCTGCCTTCCCTTTGTGACTTGAAGAACCCTGACTTTCT+CAAAGGCACCTGAATGTGTCTGTGTTCCTGTAGGCATAATGTGT
```

Figure 16. Fasta sequence output for phase 1 indicating the location of the insertion.

5.1.9. Figure 17 shows the fasta output for phase 2. The nucleotide at the indicated position requires editing before submitting the sequence. From the Assign coverage pane, it can be determined that the correction call is “G”.

```
>HC20232L20180IHW09245_IMGT/H_Phase 2
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGCACAGGAGGAGCGGGGTCAGGGCGAAGT
CCCAGGGCCCCAGGCGTGGCTCTCAGGGTCTCAGGCCCCGAAGGCGGTGTATGGATTGGGGAGGCCCCGCCTTGGGGATTCGCCACCTCCGCAGTTTCTC
TTCTTCTCACAACCTGCGACGGGTCCTTTTTCCTGGATACTCACGAAGCGGGCACAGTTCTCATTCCCACTAGGTGTCGGGTTTCTAGAGAAGCCAATCG
GTGCCGCCGCGGTCCCGGTTCTAAAGTCCCCACGCACCCACCGGGACTCAGATTCTCCCCAGACGCCGAGGATGGTGCTCATGGCGCCCCGAACCCTCCT
CCTGCTGCTCTCAGGGGCCCTGGCCCTGACCCAGACCTGGGCGCGTGAGTGCAGGGTCTGCAGGGAAATGGTCGGGAGGAGCGAGGGGCCCGCCCGGCCG
GGGCGCAGGACCCAGGGAGCCGCGCAGGGAGGAGGGTCGGGCGGGTCTCAGCTCCTCCTCGCTCCCAGGCTCCCACTCCATGAGGTATTTCTACACCACC
ATGTCCCGGCCCGGCCGCGGGGAGCCCCGCTTCATCTCCGTCGGCTACGTGGACGATACGCAGTTCGTGCGGTTCGACAGCGACGACGCGAGTCCGAGAG
AGGAGCCGCGGGCGCCGTGGATGGAGCGGGAGGGGCCGGAGTATTGGGACCGGAACACACAGATCTGCAAGGCCCAAGCACGGACTGAACGAGAGAACCT
GCGGATCGCGCTCCGCTACTACAACCAGAGCGAGGGCGGTGAGTGACCCCGGCCCGGGGCGCAGGTCACGACCCCTCCCCATCCCCCACGGAGGGCCGGG
TCGCCTCGAGTCTCTGGGTCCGAGATCCTCCCCGAAACCGCGGGACCCCGAGACCCTTGACCTGGGAGAGGCCCAGGCGCCTTTACCCGGTTTCATTTTC
AGTTTAGGCCAAAATCCCCGCGGGTTGGTCGGGGCAGGGCGGGGCTCGGGGGACCGGGCTGACCGCGGGGGCGGGGCCAGGTTCTCACACCATGCAGGTG
ATGTATGGCTGCGACTTGGGGCCCGACGGGCGCTTCCTCCGCGGGTATGAACAGCACGCCTACGACGGCAAGGATTACATCGCTCTGAACGAGGACCTGC
GCTCCTGGACCGCGGCGGACATGGCAGCTCAGATCACCAAGCGCAAGTGGGAGGCGGCCCGTCGGGCGGAGCAGCGGAGAGTCTACCTGGAGGGCGAGTT
CGTGGAGTGGCTCCGCAGATACCTGGAGAACGGGAAGGAGACGCTGCAGCGCGCGGGTACCAGGGGCCACAGGGCGCCTCCCGGATGGTCTGTAGATCTC
CGGGGCTGGCCTCCCACAAGAAAGGGAGACAAATGGGACCAACACTATAATATCGCCCTCCCTCTGGTCCTGAGGGAGAAGAATCCTCCTGGGTTTCCAG
AGAGTGACTCTGAGGGTCCGCCGTGCTCTCTGACACAATTAAGGGATGAAATCTCTGAGGAAATGAAGGGAAGACAATCCCTGGAATACTGATGAGTGGT
TCCCTTTGACAC+GGCAGCAGCCTTGGGCCCCGTGACTTTTCCTCTCAGGCCTTGTTCTCTGCTTCACACTCAATGTGCCTGGGGGTCTGAGTCCAGCTC
TTCTGAGTCCCTCAGCCTCCACTCAGGTCAGGACCAGAAGTCGCTGTTCCCTCCTCAGGGACTAGAATTTTCCACGGAATAGGAGATTATCCCAGGTGCC
TGTGTCCAGGCTGTTGTCTGGGTTCTGTGCTCCCTTCCCCACCCCAGGCGTCCTGTCCATTCTCAAGATGGCCACATGCGTGCTGGTGGAGTGTCCCATG
ACAGATGCAAAATGCCTGAATTTTCTGACTCTTCCCGTCAGACCCCCCCAAGACACATATGACCCACCACCCCATCTCTGACCATGAGGCCACCCTGAGG
TGCTGGGCCCTGGGCTTCTACCCTGCGGAGATCACACTGACCTGGCAGCGGGATGGGGAGGACCAGACCCACACACGGAGCTCGTGGAGACCAGGCCTGC
AGGGGATGGAACCTTCCAGAAGTGGGCGGCTGTGGTGGTGCCTTCTGGAGAGGAGCAGAGATACACCTGCCATGTGCAGCATGAGGGTCTGCCAGAGCCC
CTCACCCTGAGATGGGGTAAGGAGGGAGATGGGGGTGTCATGTCCCTTAGGGAAAGCCAGAGCCTCTCTGGAGAGCTTTAGCAGGGTCAGGGTCCCTCAC
CTTCCCCCCTTTTCCCAGAGCCATCTTCCCAGCCCACCATCCCCATCGTGGGCATCGTTGCTGGCCTGGTTCTACTTGTAGCTGTGGTCACTGGAGCTGT
GGTCGCTGCTGTAATGTGGAGGAAGAAGAGCTCAGGTAAGGAAGGGGTGAGGAGTGTGGTCTGAGATTTCTTGTCTCACTGAGAGTTCCAAGCCCCAGGT
AGAAGTGCCCTGCCTGGTTACTGGGAAGCACCATCCACACTCATGGGCCTACCCAGCCTGGGCCCTGTGTGCCAGCACTTACTCTTTTGTAAAGCACCTG
TTACAATGAGGGACAGATTTATCACCTTGATGACTGTGGTGATGGGACCTGATCCCAGCAGTCACAAGTCACAGGGGAAGGTCCCCGAGGACAGACCTCA
GAAGGGCGGTTGGTCCAGGACCCACATCTGCTTTCCTCATGTTTCCTGATCCCGCCCTGGGTCTGCAGTTGCACATTTCTGGAAACTTCTCTGGGGTCCA
AGACTTGGAGGTTCCTCTAGGACCTTATGGCCCTGGCTTCTTTCTGGCATCTCACAGGACATTTTCTTCCCACAGATAGAAAAGGAGGGAGCTACTCTCA
GGCTGCAAGTAAGTATGAAGGAGGCTGATCCCTGAAATCCTTTGGATATTGTGTTTGGGAGCCCATGGGGGAGCTCACCCACCCCACAATTCTTCCTCTA
GCCACATCTACTGTGGGATCTGACCAGGTCCTGTTTTTATTCTACTCCAGGCGGCAACAGTGCCCAGGGCTCTGATGTGTCTCTCACGGCGTGAAAGGTG
AGACCTTGGGGGGCCTGATGTGTGGGGGGTGTTGGGGGGGAACAGTGGACACAGCTGTGCTATGGGGTTCTTTGAATTTGATGTTTTGAGCGTGCGATGG
GCTGCCAAAGTGTCATCCATTACTGGGACAGATATGAATTTGTTCATGAATATTTTTTCTATAGTGTGAGACAGCTGCCTTGTGTGGGACTGAGAGGCAA
GATTTGTTCACACCTTCCCTTTGTGACTTGAAGAACCCTGACTTTCT+CAAAGGCACCTGAATGTGTCTGTGTTCCTGTAGGCATAATGTGTGGAGGAGG
GGAGACCAACCCACCCTCATGTCCACCATGACCCTCTTCCCCACGCTGATCTGTGTTCCCTCCCCAATCATCTTTCCTGTTCCAGAGAGGCGGGGCTGAG
```

Figure 17. Fasta sequence output for phase 2 indicating the location of the insertion.

5.1.10. Once the fasta file has been edited and verified for accuracy, copy the phase layer of the novel allele including the header and paste into a new text editor. Save the file as a .fasta file. This is the file that will be used in the GenBank submission.

6. Locate the gene features

Gene features include the locations of the start and end of the exons in the context of the fasta sequence.

6.1. cDNA sequence

6.1.1. In IMGT, align the closest matched allele to the novel. https://www.ebi.ac.uk/ipd/imgt/hla/align.html

6.1.2. For best results, use the closest allele as the reference sequence and align a single allele. Align in blocks of 10 bases.

6.1.3. follow the steps in the alignment tool.

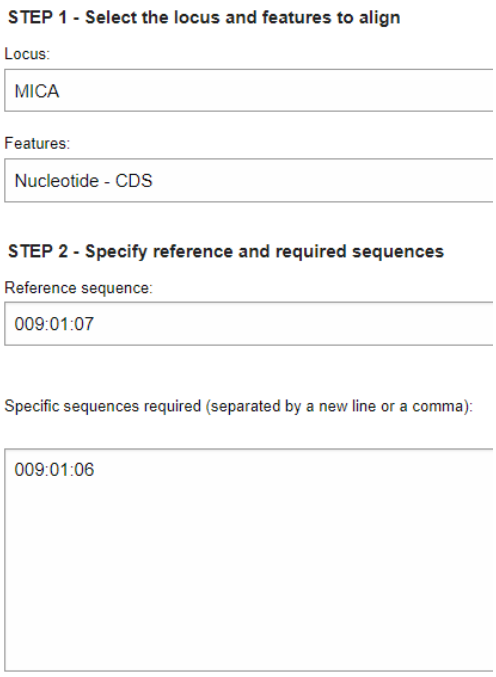


Figure 18. Alignment form in IMGT.

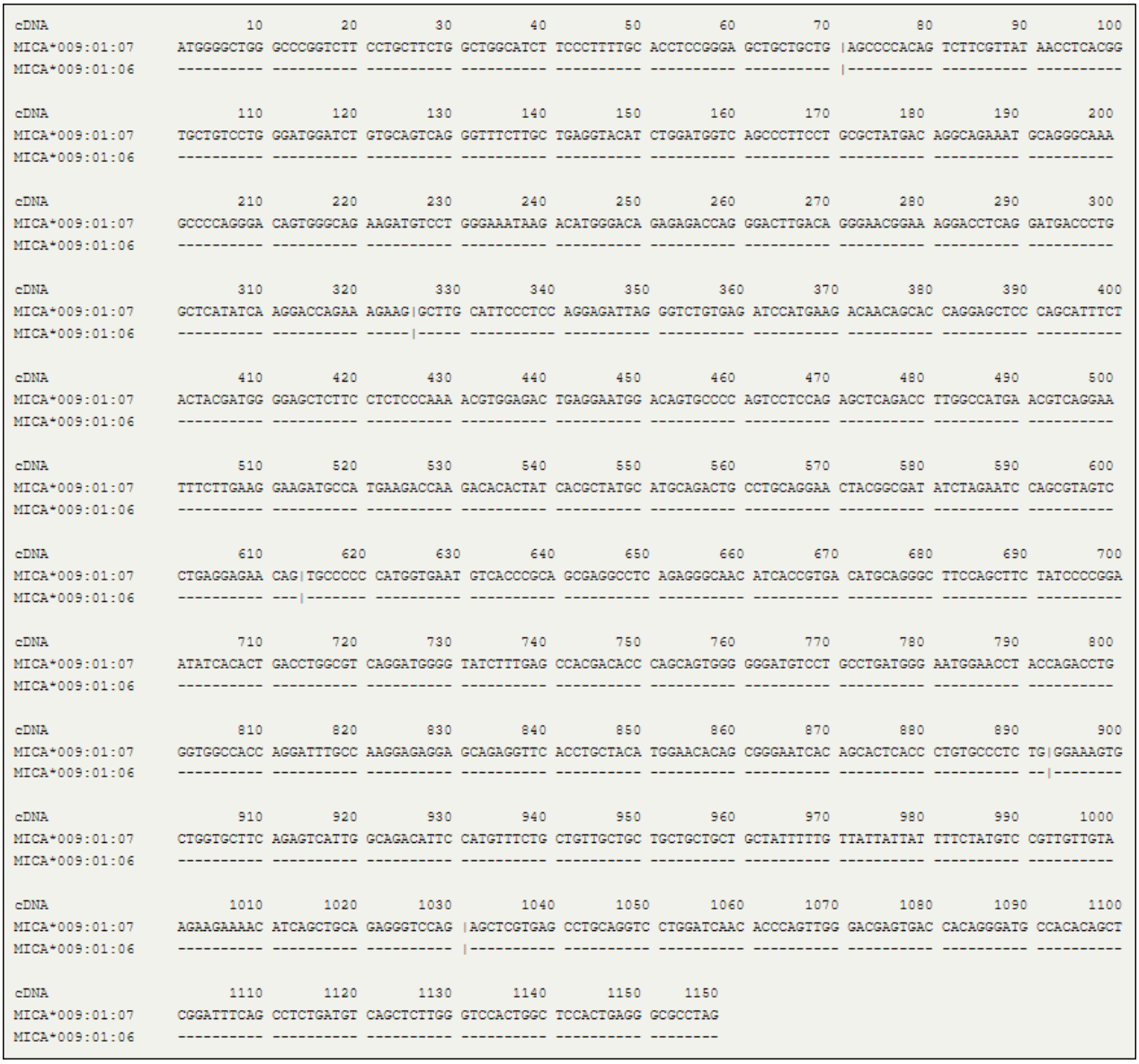


Figure 19. Alignment of MICA alleles in IMGT.

6.1.4. Use the numbering in the alignment to determine the feature locations. The vertical lines (Pipes) indicate the exon boundaries.

Table 1. Feature locations for MICA as determined from the alignment.

Feature	**Location**
Exon 1	1-70
Exon 2	71-325
Exon 3	326-613
Exon 4	614-892
Exon 5	893-1030
Exon 6	1031-1150

6.2. **gDNA sequence**

6.2.1. To determine the feature locations, BLAST the fasta sequence output from Assign using IMGT BLAST *https://www.ebi.ac.uk/Tools/services/web_ncbiblast/toolform.ebi?tool=ncbiblast&context=nucleotide&database=imgthlagen/*

6.2.2. Select the appropriate IMGT database to search. i.e., for a full gene sequence select IMGT> IMGT/HLA (genomic).

6.2.3. Paste the fasta nucleotide sequence for the novel allele into IMGT Nucleotide BLAST and submit the job.

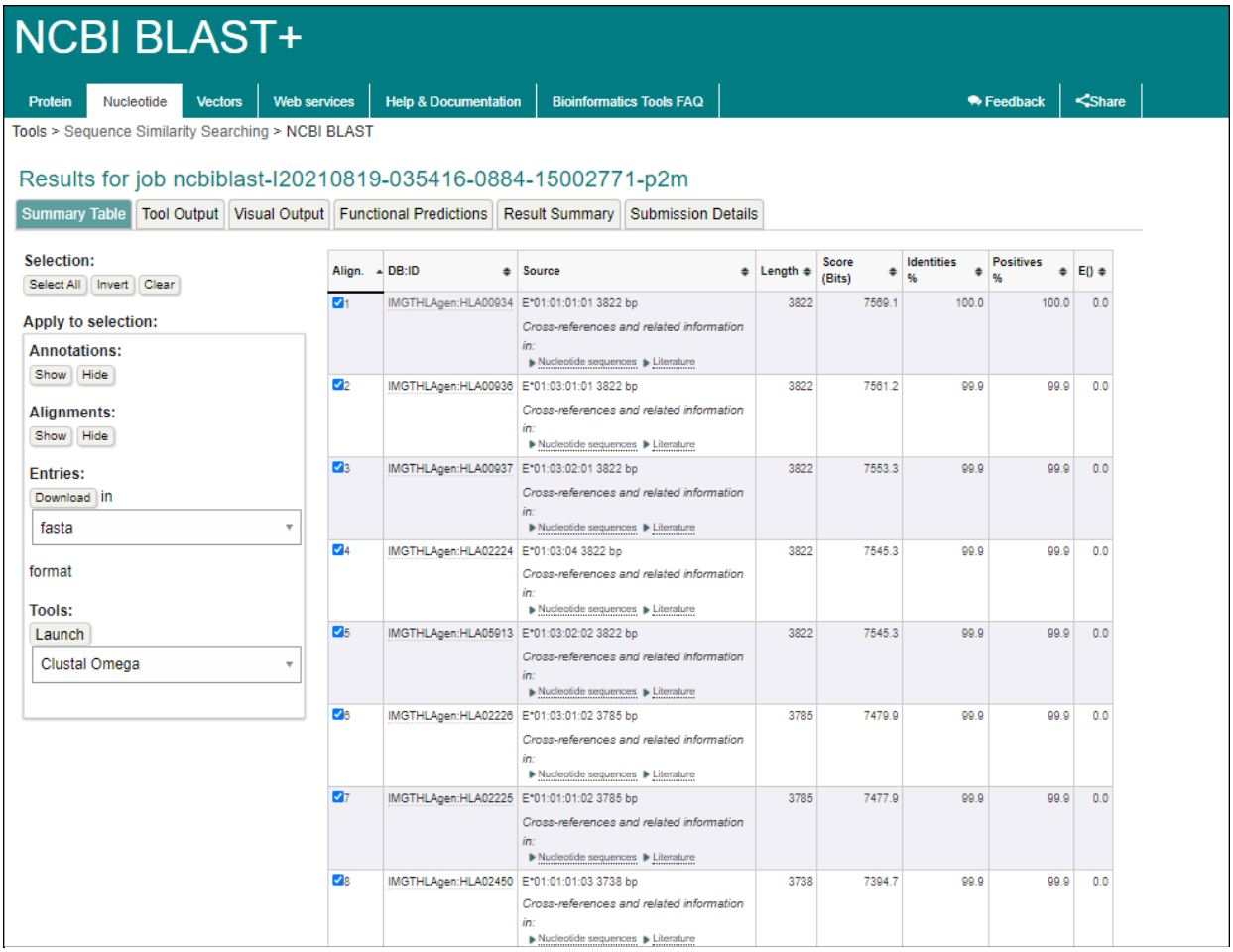


Figure 20. Summary table of hits for novel HLA-E sequence.

6.2.4. IMGT BLAST will then display the matches for the sequence.

6.2.5. By default, all the matches will be selected. Click Clear under Selection to deselect all matches.

6.2.6. Select the top result, then click show under Annotations and show under Alignments to review the results for the top result.

6.2.7. If the length of the Subject sequence is the same length as the novel fasta sequence (or query) then the annotations will correspond to the novel sequence. Check the length in the alignment and locate the novel position(s).

```
>IMGTHLAgen:HLA00934 E*01:01:01:01 3822 bp
Length=3822

 Score = 7569 bits (3818),  Expect = 0.0
 Identities = 3821/3822 (99%), Gaps = 0/3822 (0%)
 Strand=Plus/Plus

Query  1     CAAAGTGCTGAGATTACAGGCGTGAGCCACCGCGCCCAGCCAGGACTAATTTCTAAGAGT  60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1     CAAAGTGCTGAGATTACAGGCGTGAGCCACCGCGCCCAGCCAGGACTAATTTCTAAGAGT  60
```

Figure 21. The start of the alignment for the novel HLA-E sequence and top match E*01:01:01:01 matches.

```
Query  3781  GGCAGAGTGCGGCAGCTCATGCCTGTAATCCCAGCACTTAGG  3822
              |||||||||||||||||||||||||||||||||||||||||
Sbjct  3781  GGCAGAGTGCGGCAGCTCATGCCTGTAATCCCAGCACTTAGG  3822
```

Figure 22. The end of the alignment for the novel HLA-E sequence and top match E*01:01:01:01 matches.

```
Query  2101  TGGTGCCTTCTGGAGAGGAGCAGAGATACACGTGCCATGTGCAGAATGAGGGGCTACCCG  2160
              ||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Sbjct  2101  TGGTGCCTTCTGGAGAGGAGCAGAGATACACGTGCCATGTGCAGCATGAGGGGCTACCCG  2160
```

Figure 23. The location of the novel SNP as indicated by the absence of the pipe in the alignment for a novel HLA-E sequence and top match E*01:01:01:01.

6.2.8. The alignment will also highlight any errors in the fasta sequence due to N's or "+". If any errors are identified these can be edited in the fasta sequence.

The screenshot below shows an example of an HLA-H sequence error due to "+". The "+" has been converted to N, and there is missing sequence in the fasta file.

```
Query  3241  TTGTTCCTGCCTTCCCTTTGTGACTTGAAGAACCCTGA----CTTTCTNCAAAGGCACCT  3296
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3241  TTGTTCCTGCCTTCCCTTTGTGACTTGAAGAACCCTGACTTTCTTTCTACAAAGGCACCT  3300
```

Figure 24. Alignment of novel HLA-H*02:04:new sequence and H*02:04:01.

6.2.9. Error! Reference source not found. shows that at the position where Assign has put a "+" there is an A and 4nt insertion TTCT, matching the A*02:04:01 allele.

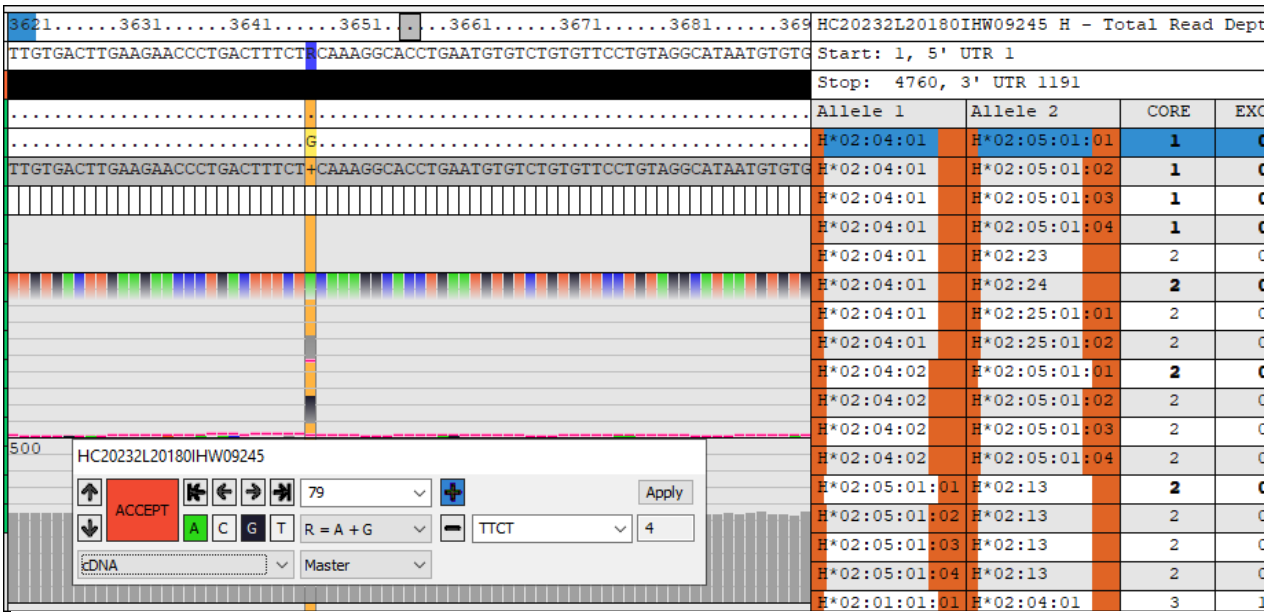


Figure 25. Alignment of a novel HLA-H sample in Assign highlighting the insertion.

6.2.10. This position can be edited in the fasta sequence to match the display in Assign, by locating the "+" in the fasta sequence, then replacing it with "TTCTA".

```
>IHW09245_IMGT/H_Phase 1
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGCACAGGAGGAGCGGGGTCAGGGCGAAGT
CCCAGGGCCCCAGGCGTGGCTCTCAGGGTCTCAGGCCCCGAAGGCGGTGTATGGATTGGGGATGCCCCGCCTTAGGGATTCGCCACCTCCGCAGTTTCTC
TTCTTCTCACAACCTGCGACGGGTCCTTCTTCCTTGATACTCACGAAGCGGACACAGTTCTCATTCCCACTAGGTGTCGGGTTTCTAGAGAAGCCAATCG
GTGCCGCCGCGGTCCCGGTTCTAAAGTCCCCACGCACCCACCGGGACTCAGATTCTCCCCAGACGCCGAGGATGGTGCTCATGGCGCCCCGAACCCTCCT
CCTGCTGCTCTCAGGGGCCCTGGCCCTGACCCTGACCCAGACCTGGGCGCGTGAGTGCAGGGTCTGCAGGGAAATGGTCGGGAGGAGCGAGGGGCCCGCC
CGGCGGGGGCGCAGGACCCGGGGATCCGCGCAGGGAGGAGGGTCGGGCCGGTCTCAGCTCCTCCTCGCTCCCAGGCTCCCACTCCATGAGGTATTTCTAC
ACCACCATGTCCCGGCCCGGCCGCGGGGAGCCCCGCTTCATCTCCGTCGGCTACGTGGACGATACGCAGTTCGTGCGGTTCGACAGCGACGACGCGAGTC
CGAGAGAGGAGCCGCGGGCGCCGTGGATGGAGCGGGAGGGGCCAGAGTATTGGGACCGGAACACACAGATCTGCAAGGCCCAGGCACAGACTGAACGAGA
GAACCTGCGGATCGCGCTCCGCTACTACAACCAGAGCGAGGGCGGTGAGTGACCCCGGCCCGGGGCGCAGGTCACGACCCCTCCCCATCCCCCACGGAGG
GCCGGGTCGCCTCGAGTCTCTGGGTCCGAGATCCACCCCGAAACCGCGGACCCCGAGACCCTTGACCTGGGAGAGGCCCAGGCGCCTTTACCCGGTTTC
ATTTTCAGTTTAGGCCAAAACCCCCGCGGGTTGGTCGGGGCAGGGCGGGGCTCGGGGGACCGGGCTGACCGCGGGGGCGGGGCCAGGTTCTCACACCATG
CAGGTGATGTATGGCTGCGACGTGGGGCCCGACGGACGCTTCCTCCGCGGGTATGAACAGCACGCCTACGACGGCAAGGATTACATCGCCCTGAACGAGG
ACCTGCGCTCCTGGACCGCGGCGGACATGGCAGCTCAGATCACCAAGCGCAAGTGGGAGGCGGCCCGTCGGGCGGAGCAGCTGAGAGCCTACCTGGAGGG
CGAGTTCGTGGAGTGGCTCCGCAGATACCTGGAGAACGGGAAGGAGACGCTGCAGCGCGCGGGTACCAGGGGCCACAGGGCGCCTCCCGGATCGCCTGTA
GATCTCCGGGGCTGGCCTCCCACAAGAAAGGGAGACAAATGGGACCAACACTATAATATCGCCCTCCCTCTGGTCTTGAGGGAGAAGAATCCTCCTGGGT
TTCCAGAGAGTGACTCTGAGGGTCCGCCCTGCTCTCTGACACAATTAAGGGATGAAATCTGTGAGGAAATGAAGGGAAGACAATCCCTGGAATACTGATG
AGTGGTTCCCTTTGACAC+GGCAGCAGCCTTGGGCCCCGTGACTTTTCCTCTCAGGCCTTGTTCTCTGCTTCACACTCAATGTGCGTGGGGGTCTGAGTC
CAGCTCTTCTGAGTCCCTCAGCCTCCACTCAGGTCAGGACCAGAAGTCGCTGTTCCCTCTTCAGGGACTAGAATTTTCCACGGAATAGGAGATTATCCCA
GGTGCCTGTGTCCAGGCTGTTGTCTGGGTTCTGTGCTCCCTTCCCCACCCCAGGCATCCTGTCCATTCTCAAGATGGCCACATGCGTGCTGGAGGAGTGT
CCCATGACAGATGCAAAATGCCTGAATTTTCTGACTCTTCCCGTCAGACCCCCCCAAGACACATATGACCCACCACCCCATCTCTGACCATGAGGCCAC
CCTGAGGTGCTGGGCCCTGGGCTTCTACCCTGCGGAGATCACACTGACCTGGCAGCGGGATGGGGAGGACCAGACCCAGGACACGGAGCTCGTGGAGACC
AGGCCTGCAGGGGATGGAACCTTCCAGAAGTGGGCGGCTGTGGTGGTGCCTTCTGGAGAGGAGCAGAGATACACCTGCCATGTGCAGCATGAGGGTCTGC
CCGAGCCCCTCACCCTGAGATGGGGTAAGGAGGGAGATGGGGGTGTCATGTCTCTTAGGGAAAGCCGGAGCCTCTCTGGAGAGCTTTAGCAGGGTCAGGG
TCCCTCACCTTCCCCCCTTTTCCCAGAGCCATCTTCCCAGCCCAACGTCCCCATCGTGGGCATCGTTGCTGGCCTGGTTCTACTTGTAGCTGTGGTCACT
GGAGCTGTGGTCGCTGCTGTAATGTGGAGGAAGAAGAGCTCAGGTAAGGAAGGGGTGAGGAGTGTGGTCTGAGATTTCTTGTCTCACTGAGAGTTCCAAG
CCCCAGGTAGAAGTGCCCTGCCTGGTTACTGGGAAGCACCATCCACACTCATGGGCCTACCCAGCCTGGGCCGTGTGTGCCAGCACTTACTCTTTTGTAA
AGCACCTGTTACAATGAGGGACAGATTTATCACCTTGATGACTGTGGTGATGGGACCTGATCCCAGCAGTCACAAGTCACAGGGGAAGGTCCCCAAGGAC
AGACCTCAGAAGGGCGGTTGGTCCAGGACCCATATCTGCTTTCTTCATGTTTCCTGATCCCACCCTGGGTCTGCAGTTGCACATTTCTGGAAACTTCTCT
GGGGTCCGAGACTTGGAGGTTCCTCTAGGACCTTATGGCCCTGGCTTCTTTCTGGCATCTCACAGGACATTTTCTTCCCACAGATAGAAAAGGAGGGAGC
TACTCTCAGGCTGCAAGTAAGTATGAAGGAGGCTGATCCCTGAAATCCTTTGGATATTGTGTTTGGGAGCCCATGGGGGAGCTCACCCACCCCACAATTC
TTCCTCTAGCCACATCTACTGTGGGATCTGACCAGGTCCTGTTTTTATTCTACTCGAGGCAGCAACAGTGCCCAGGGCTCTGATGTGTCTCTCACGGCTT
GAAAGGTGAGACCTTGGGGGGCCTGATGTGTGGGGGATGTTGGGGGGGAACAGTGGACACAGCTGTGCTATGGGGTTCTTTGAATTTGATGTTTTGAGCA
TGCGATGGGCTGCCAAAGTGTCATCCATTACTGGGACAGATATGAATTTGTTCATGAATATTTTTTCTATAGTGTGAGACAGCTGCCTTGTGTGGGACTG
AGAGGCAAGAGTTGTTCCTGCCTTCCCTTTGTGACTTGAAGAACCCTGACTTTCT+CAAAGGCACCTGAATGTGTCTGTGTTCCTGTAGGCATAATGTGT
GGAGGAGGGGAGACCAACCCACCCTCATGTCCACCATGACCCTCTTCCCCACGCTGATCTGTGTTCCCTCCCCAATCATCTTTCCTGTTCCAGAGAGGAG
GGGCTGAGATGTCTCCATCTTTTTCTCAACTTTATGTGCACTGAGCTGTAACTTCTTACTTCCCTCTTAAAATTAGANNNNNNNNNNNNNNNNNNNNNNN
```
Figure 26. Location of the “+” in the fasta sequence output from Assign for novel HLA-H*02:04:New allele.

6.2.11. It is recommended to BLAST the edited sequence in IMGT BLAST to confirm the edits.

```
>IHW09245_IMGT/H_Phase 1
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGCACAGGAGGAGCGGGTCAGGGCGAAGT
CCCAGGGCCCCAGGCGTGGCTCTCAGGGTCTCAGGCCCCGAAGGCGGTGTATGGATTGGGGATGCCCCGCCTTAGGGATTCGCCACCTCCGCAGTTTCTC
TTCTTCTCACAACCTGCGACGGGTCCTTCTTCCTTGATACTCACGAAGCGGACACAGTTCTCATTCCCACTAGGTGTCGGGTTTCTAGAGAAGCCAATCG
GTGCCGCCGCGGTCCCGGTTCTAAAGTCCCCACGCACCCACCGGGACTCAGATTCTCCCCAGACGCCGAGGATGGTGCTCATGGCGCCCCGAACCCTCCT
CCTGCTGCTCTCAGGGGCCCTGGCCCTGACCCTGACCCAGACCTGGGCGCGTGAGTGCAGGGTCTGCAGGGAAATGGTCGGGAGGAGCGAGGGGCCCGCC
CGGCGGGGGCGCAGGACCCGGGGATCCGCGCAGGGAGGAGGGTCGGGCCGGTCTCAGCTCCTCCTCGCTCCCAGGCTCCCACTCCATGAGGTATTTCTAC
ACCACCATGTCCCGGCCCGGCCGCGGGGAGCCCCGCTTCATCTCCGTCGGCTACGTGGACGATACGCAGTTCGTGCGGTTCGACAGCGACGACGCGAGTC
CGAGAGAGGAGCCGCGGGCGCCGTGGATGGAGCGGGAGGGGCCAGAGTATTGGGACCGGAACACACAGATCTGCAAGGCCCAGGCACAGACTGAACGAGA
GAACCTGCGGATCGCGCTCCGCTACTACAACCAGAGCGAGGGCGGTGAGTGACCCCGGCCCGGGGCGCAGGTCACGACCCCTCCCCATCCCCCACGGAGG
GCCGGGTCGCCTCGAGTCTCTGGGTCCGAGATCCACCCCGAAACCGCGGGACCCCGAGACCCTTGACCTGGGAGAGGCCCAGGCGCCTTTACCCGGTTTC
ATTTTCAGTTTAGGCCAAAACCCCCGCGGGTTGGTCGGGGCAGGGCGGGGCTCGGGGGACCGGGCTGACCGCGGGGGCGGGGCCAGGTTCTCACACCATG
CAGGTGATGTATGGCTGCGACGTGGGGCCCGACGGACGCTTCCTCCGCGGGTATGAACAGCACGCCTACGACGGCAAGGATTACATCGCCCTGAACGAGG
ACCTGCGCTCCTGGACCGCGGCGGACATGGCAGCTCAGATCACCAAGCGCAAGTGGGAGGCGGCCCGTCGGGCGGAGCAGCTGAGAGCCTACCTGGAGGG
CGAGTTCGTGGAGTGGCTCCGCAGATACCTGGAGAACGGGAAGGAGACGCTGCAGCGCGCGGGTACCAGGGGCCACAGGGCGCCTCCCGGATCGCCTGTA
GATCTCCGGGGCTGGCCTCCCACAAGAAAGGGAGACAAATGGGACCAACACTATAATATCGCCCTCCCTCTGGTCTTGAGGGAGAAGAATCCTCCTGGGT
TTCCAGAGAGTGACTCTGAGGGTCCGCCCTGCTCTCTGACACAATTAAGGGATGAAATCTGTGAGGAAATGAAGGGAAGACAATCCCTGGAATACTGATG
AGTGGTTCCCTTTGACAC+GGCAGCAGCCTTGGGCCCCGTGACTTTTCCTCTCAGGCCTTGTTCTCTGCTTCACACTCAATGTGCGTGGGGGTCTGAGTC
CAGCTCTTCTGAGTCCCTCAGCCTCCACTCAGGTCAGGACCAGAAGTCGCTGTTCCCTCTTCAGGGACTAGAATTTTCCACGGAATAGGAGATTATCCCA
GGTGCCTGTGTCCAGGCTGTTGTCTGGGTTCTGTGCTCCCTTCCCCACCCCAGGCATCCTGTCCATTCTCAAGATGGCCACATGCGTGCTGGAGGAGTGT
CCCATGACAGATGCAAAATGCCTGAATTTTCTGACTCTTCCCGTCAGACCCCCCCCAAGACACATATGACCCACCACCCCATCTCTGACCATGAGGCCAC
CCTGAGGTGCTGGGCCCTGGGCTTCTACCCTGCGGAGATCACACTGACCTGGCAGCGGGATGGGGAGGACCAGACCCAGGACACGGAGCTCGTGGAGACC
AGGCCTGCAGGGGATGGAACCTTCCAGAAGTGGGCGGCTGTGGTGGTGCCTTCTGGAGAGGAGCAGAGATACACCTGCCATGTGCAGCATGAGGGTCTGC
CCGAGCCCCTCACCCTGAGATGGGGTAAGGAGGGAGATGGGGGTGTCATGTCTCTTAGGGAAAGCCGGAGCCTCTCTGGAGAGCTTTAGCAGGGTCAGGG
TCCCTCACCTTCCCCCCTTTTCCCAGAGCCATCTTCCCAGCCCAACGTCCCCATCGTGGGCATCGTTGCTGGCCTGGTTCTACTTGTAGCTGTGGTCACT
GGAGCTGTGGTCGCTGCTGTAATGTGGAGGAAGAAGAGCTCAGGTAAGGAAGGGGTGAGGAGTGTGGTCTGAGATTTCTTGTCTCACTGAGAGTTCCAAG
CCCCAGGTAGAAGTGCCCTGCCTGGTTACTGGGAAGCACCATCCACACTCATGGGCCTACCCAGCCTGGGCCGTGTGTGCCAGCACTTACTCTTTTGTAA
AGCACCTGTTACAATGAGGGACAGATTTATCACCTTGATGACTGTGGTGATGGGACCTGATCCCAGCAGTCACAAGTCACAGGGGAAGGTCCCCAAGGAC
AGACCTCAGAAGGGCGGTTGGTCCAGGACCCATATCTGCTTTCTTCATGTTTCCTGATCCCACCCTGGGTCTGCAGTTGCACATTTCTGGAAACTTCTCT
GGGGTCCGAGACTTGGAGGTTCCTCTAGGACCTTATGGCCCTGGCTTCTTTCTGGCATCTCACAGGACATTTTCTTCCCACAGATAGAAAAGGAGGGAGC
TACTCTCAGGCTGCAAGTAAGTATGAAGGAGGCTGATCCCTGAAATCCTTTGGATATTGTGTTTGGGAGCCCATGGGGGAGCTCACCCACCCCACAATTC
TTCCTCTAGCCACATCTACTGTGGGATCTGACCAGGTCCTGTTTTTATTCTACTCGAGGCAGCAACAGTGCCCAGGGCTCTGATGTGTCTCTCACGGCTT
GAAAGGTGAGACCTTGGGGGGCCTGATGTGTGGGGGATGTTGGGGGGGAACAGTGGACACAGCTGTGCTATGGGGTTCTTTGAATTTGATGTTTTGAGCA
TGCGATGGGCTGCCAAAGTGTCATCCATTACTGGGACAGATATGAATTTGTTCATGAATATTTTTTCTATAGTGTGAGACAGCTGCCTTGTGTGGGACTG
AGAGGCAAGAGTTGTTCCTGCCTTCCCTTTGTGACTTGAAGAACCCTGACTTTCTTTCTACAAAGGCACCTGAATGTGTCTGTGTTCCTGTAGGCATAATGTGT
GGAGGAGGGGAGACCAACCCACCCTCATGTCCACCATGACCCTCTTCCCCACGCTGATCTGTGTTCCCTCCCCAATCATCTTTCCTGTTCCAGAGAGGAG
GGGCTGAGATGTCTCCATCTTTTTCTCAACTTTATGTGCACTGAGCTGTAACTTCTTACTTCCCTCTTAAAATTAGANNNNNNNNNNNNN
```

Figure 27. The sequence edited to include the inserted bases in the novel HLA-H*02:04:New allele.

6.2.12. Once satisfied that the annotation locations match the novel allele sequence, they can be used to create the 5-column feature table if submitting to Genbank.

7. Genbank submission

7.1. the 5-column feature table for Genbank submission

7.1.1. The Genbank Bankit tool has 2 functions- *manual entry using a form* or *upload a 5-column feature table*- for adding gene feature locations to a sequence, but the same information is required for each.

7.1.2. Creating a 5-column feature table allows the inclusion of exons, CDS and details of the gene, as well as other features described in the help links below from Genbank.

7.1.3. This page *https://www.ncbi.nlm.nih.gov/WebSub/html/help/feature-table.html* describes the requirements for creating the feature table.

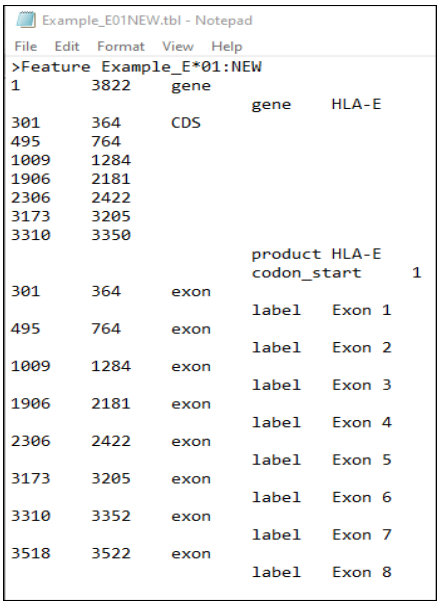
7.1.4. This page *https://www.insdc.org/files/feature_table.html* contains detailed information on the feature table.

7.1.5. 5-column feature tables are saved as .tbl or .txt format and may be edited with notepad.

7.1.6. Some examples of feature table files have been included with this SOP.

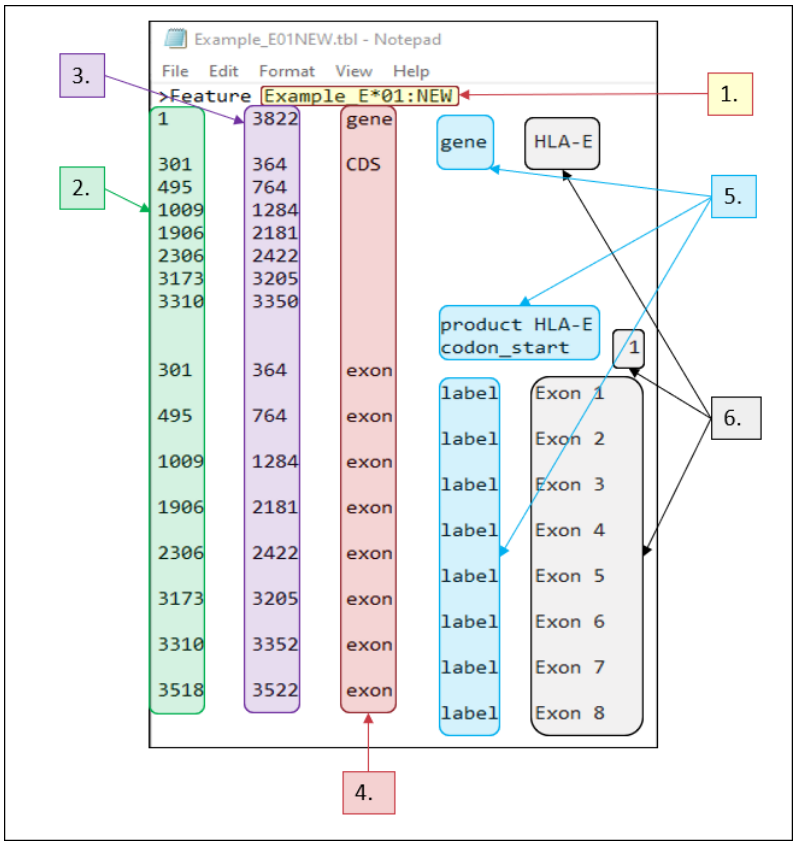
7.1.7. Key features of the 5-column feature table:

Example HLA-E



The 5-column feature table as viewed in notepad.

Figure 28. 5-column feature table example for HLA-E



Column descriptions:

1.Sequence Identifier

Line 1

2. Column 1: Start location (first nucleotide) of a feature

3. Column 2: Stop location (last nucleotide) of a feature

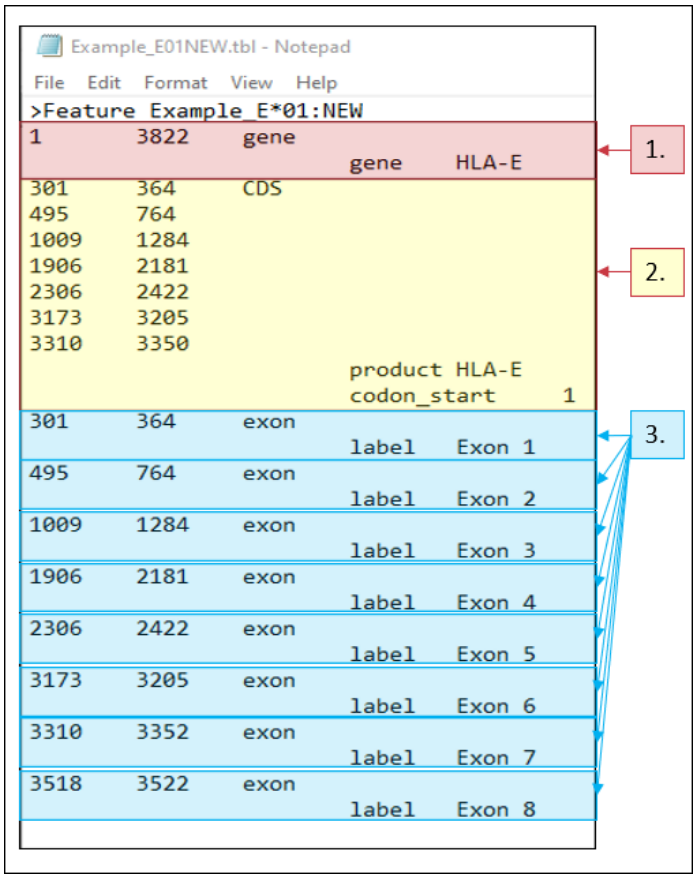
4. Column 3: Feature name (for example, 'CDS' or 'mRNA' or 'rRNA' or 'gene' or 'exon')

Line 2

5. Column 4: Qualifier name (for example, 'product' or 'number' or 'gene' or 'note')

6. Column 5: Qualifier value

Figure 29. 5-column feature table with columns indicated.



Feature examples:

1.Gene Feature

2.CDS Feature

3.Exon Features

Figure 30. 5-column feature table with features indicated.

7.1.8. Tab stops are used to separate each column in the feature table.

7.2. **Create the 5-column feature table**

7.2.1. For cDNA sequences- feature table and gapped sequence

7.2.2. Use the template *TEC745-1_Create feature table cDNA.xlsx* to create the 5-column feature table.

7.2.3. Enter the feature locations into the first tab- exon_locations.

7.2.4. Copy the cDNA fasta sequence output for the novel allele including the header and paste into the paste_fasta tab.

7.2.5. The template will add a string of N's (gaps) to the fasta sequence to represent to introns.

7.2.6. The updated sequence can then be copied and pasted to a new text document and saved as .fasta or.txt file.

7.2.7. The template will generate the 5-column feature table with updated exon locations and gap locations in the tab tbl_output.

7.2.8. The feature table can then be copied and pasted into a new text document and saved as .tbl or .txt file.

7.2.9. For gDNA sequences- feature table

7.2.10. Use the template *TEC745-2_Create feature table gDNA.xlsx* to create the 5-column feature table.

7.2.11. In the first tab exon_locations, select the gene name from the list to populate the default exon locations, or manually enter the exon locations in the table.

NOTE: the default exon locations have been calculated based on common alleles and will not be suitable for all sequences. It is recommended to carefully check the feature locations for each novel sequence.

7.2.12. Copy the gDNA fasta sequence output including the header and paste into the paste_fasta tab.

NOTE: the fasta sequence is used for creating the feature table only and is not modified.

7.2.13. The template will generate the 5-column feature table in the tbl_output tab.

7.2.14. The feature table can then be copied and pasted into a new text document and saved as .tbl or .txt file.

7.3. **NCBI account**

7.3.1. Go to *https://www.ncbi.nlm.nih.gov/* and follow the instructions to set up an account.

7.4. **Submit with Bankit**

7.4.1. Click submit on the NCBI homepage and follow the prompts for submission or go directly to Bankit *https://submit.ncbi.nlm.nih.gov/about/bankit/*

7.4.2. Follow the prompts on the Bankit forms to enter contact information and references.

7.4.3. In the Sequencing Technology form, select "Other" for Method, and enter "AlloSeq Tx17.1 Hybrid Capture". Select assembled sequences and enter "AlloSeq Assign" for the Assembly program and enter the version of AlloSeq Assign used.

7.4.4. In the Nucleotide form, enter the desired release date. Select genomic DNA for the molecule type, Linear topology and No to complete sequence submission. Either paste the fasta sequence for the novel allele or upload the file.

7.4.5. Enter the Organism name (homo sapiens), Submission category and Source modifiers as appropriate.

7.4.6. In the Features form, select Add features by uploading 5 column feature table. Choose the file, then upload. Once uploaded the features will populate.

7.4.7. Review the Submission and Submit.

8. IMGT submission

8.1. **To submit to IMGT the following are required:**

8.1.1. Accession number for the sequence.

8.1.1.1. This will be provided via email by Genbank once they have performed initial processing of the submitted sequence.

8.1.2. Fasta sequence

8.1.2.1. use the sequence prepared for the Genbank/ENA/DDBJ submission.

8.1.3. The name of the closest known allele

8.1.3.1. The top result as determined by the IMGT BLAST search.

8.1.4. The locations of the differences to the closest known allele.

8.1.4.1. The location of the differences to the closest known allele can be determined using the IMGT BLAST alignment, or by performing a sequence alignment of the closest known allele *https://www.ebi.ac.uk/ipd/imgt/hla/align.html*

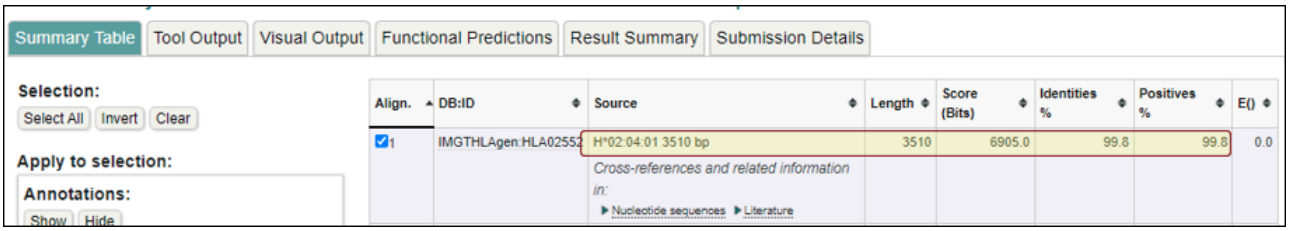


Figure 31. Top result for H*02:04:NEW from the IMGT BLAST search.

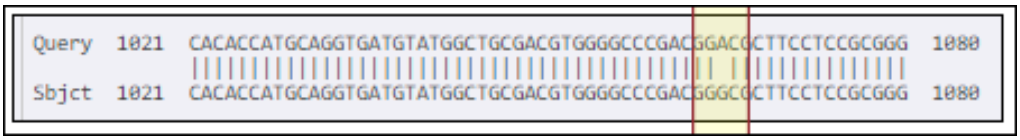


Figure 32. Alignment from the HLA-H*02:04:NEW BLAST search indicating the novel position at nt 1065.

8.1.5. The genotypes of other HLA loci such as A, B and DRB1.

8.2. Use the submission tool on the IMGT website to submit the novel allele.

8.2.1. *https://www.ebi.ac.uk/ipd/imgt/hla/subs/submit.html*

8.2.2. Once the novel sequence has been accepted and processed, IMGT will email the official name for the novel sequence.

9. References

NCBI

National Center for Biotechnology Information (NCBI)[Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; [1988] – [cited 2022 Jan 04]. Available from: https://www.ncbi.nlm.nih.gov/

IMGT

Robinson, J, Barker, DJ, Georgiou, X, Cooper, MA, Flicek, P, Marsh, SGE, The IPD-IMGT/HLA Database, Nucleic Acids Research (2020) 43:D948-D955

10. Resources

DDBJ Nucleotide Sequence Submission System (NSSS)
https://www.ddbj.nig.ac.jp/assets/files/pdf/ddbj/websubHelp_full-e.pdf

Submissions to GenBank

https://www.ncbi.nlm.nih.gov/genbank/submit/

General Guide on ENA Data Submission

https://ena-docs.readthedocs.io/en/latest/submit/general-guide.html#

AlloSeq Assign Instructions for Use

https://labproducts.caredx.com/software/assign/alloseq-assign/manuals/

11.Customer Support

Website: https://labproducts.caredx.com/

For Technical Support please email: techsupport-global@caredx.com

For ordering details, please refer to the CareDx website: https://labproducts.caredx.com/

12.Appendix

Table 2. Amino acids

Amino Acid	DNA codons	Amino Acid	DNA codons
Ala A	GCT, GCC, GCA, GCG	**Lys K**	AAG, AAA
Arg R	CGT, CGC, CGA, CGG, AGA, AGG	**Met M**	ATG
Asn N	AAT, AAC	**Phe F**	TTT, TTC
Asp D	GAT, GAC	**Pro P**	CCT, CCC, CCA, CCG
Cys C	TGT, TGC	**Ser S**	TCT, TCC, TCA, TCG, AGT, AGC
Gln Q	CAA, CAG	**Thr T**	ACT, ACC, ACA, ACG
Glu E	GAA, GAG	**Trp W**	TGG
Gly G	GGT, GGC, GGA, GGG	**Tyr Y**	TAT, TAC
His H	CAT, CAC	**Val V**	GTT, GTC, GTA, GTG
Ile I	ATT, ATC, ATA	**Start**	ATG
Leu L	TTA, TTG, CTT, CTC, CTA, CTG	**Stop**	TAA, TGA, TAG

Revision History

Version	Date	Modification
1.0	18-Aug-21	Drafted document. Issued 14 Jun 2022