mission bio

Troubleshooting 'Missing Variants'

Introduction

This document describes procedures to investigate 'missing variants'. The term 'missing variants' refers to the inability to detect target variants that are expected to be present in the data (e.g., from orthogonal bulk sequencing data) when using Tapestri Insights and the default software settings.

1.   Confirm that target variant is covered by panel

Convert target variant into hg19 coordinate

If the target variant is not available in hg19 coordinate (e.g., chr13:28592642), but is only available in codon format (e.g., c.2503G>T) or protein format (e.g., p.D835Y) you need to convert the variant into hg19 coordinate format.

- 1.1.   Visit https://bioinformatics.mdanderson.org/transvar/
- 1.2.   Select a task
  - ○   Reverse Annotation: Protein for variants like p.I254N
  - ○   Reverse Annotation: cDNA for variants like c.761T>A
- 1.3.   Select a reference genome
  - ○   GRCh37/hg19
- 1.4.   Select one or more annotation databases
  - ○   UCSC
- 1.5.   Enter variant in white box on the right side (follow required format)
  - ○   TET2:c.3594+5G>A
  - ○   FLT3:p.D835Y
- 1.6.   Click *Submit* button
- 1.7.   Review TransVar Results table and copy relevant information (in red)
  - ○   In coordinates(gDNA/cDNA/protein) column the coordinate is listed chr13:g.**28592642**C>A.

Cross-reference target variant coordinate against panel file

1.8.   Locate **insert.bed** file and the **amplicon.bed** file of the catalog our custom panel

NOTE: The coordinates in the amplicon.bed file include the amplicon primer regions and insert regions. The coordinates in the insert.bed file excludes the amplicon primer regions and only contains the insert regions.
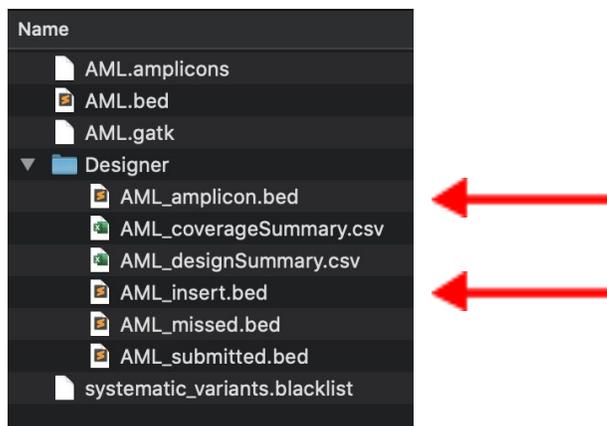
- ○ Catalog panel (AML, CLL, MYE, THP):
  - ■ V2 chemistry: Download panel file from https://tapestridesigner.com/ (files located inside *Designer* folder)
  - ■ V1 chemistry: Contact support@missionbio.com
- ○ Custom panel: Inside *Designer* folder that was shared at the time of panel design. Contact support@missionbio.com if you need assistance locating the file.

1.9.   Open the **amplicon.bed** file using a text editor like Sublime Text

1.10. Using the Finder function (command+F on MacOS) search for the amplicon that may harbor the target variant

1.11. Iteratively remove the last digit of the coordinate and search again until the coordinate overlaps partially with one value in the file

NOTE: If you remove 4 digits and still do not find a matching value, the coordinate is not covered by the panel

1.12. If the target coordinate is detectable, ensure that the coordinate is covered by Read1 and/or Read2 (assuming data was generated using the recommended PE 150 sequencing chemistry):

- ○ Add 100 nt to the first-column coordinate to account for Read1 sequencing data ➜ If the target coordinate is within the first 100 nt after the first-column coordinate, continue to **Step 1.13**
- ○ Subtract 150 nt from the second-column coordinate to account for Read2 sequencing data ➜ If the target coordinate is within the first 150 nt before the second-column coordinate, continue to **Step 1.13**
- ○ If the target coordinate is not covered by either read, it is located in the 'gap' region between both reads (and should be listed in the missed.bed file, located inside the *Designer* folder)

1.13. Open the **insert.bed** file and follow the **Steps 1.12** to confirm that the target coordinate is covered and not located inside the primer region.
- ○ Covered ➜ proceed to **Step 2 - Review coverage quality** below.
- ○ Not covered ➜ The target coordinate is located inside the amplicon primer sequence. Quality of the data for this region cannot be guaranteed and may impact the ability to call genotypes accurately and robustly.

NOTE: *If your target coordinate resides in the first nucleotide of either read or within the last 5 - 10 nucleotides of either read, the quality of the data may be impacted. Contact* [*support@missionbio.com*](mailto:support@missionbio.com) *for additional information.*

2. Review coverage quality

Ensure that recommended average coverage is being met

2.1. On Tapestri Pipeline account assess the number of reads/cell ➜ divide number by number of amplicons to compute per cell/per amplicon coverage (number of amplicons equals the number of rows in the **amplicon.bed** file in **Step 1.12**)
Example (AML): 6,441.78 / 127 amplicons = 51 X

2.2. If average coverage is below 30 X the sample is under-sequenced and may explain challenges to detect target variants


NOTE: *Tapestri Insights default filter settings may be inadequate if sample(s) are under/over-sequenced.*

Review target-specific amplicon variant coverage

2.3.   Download the barcode.cell.coverage.tsv file from your Tapestri Pipeline account (SAMPLE > ANALYSIS DETAILS)

2.4.   Open file in Microsoft Excel, reformat text data in first column to 3 columns (Data > Text to Columns)

2.5.   Order data by mean reads and identify amplicon that harbors target variant (amplicon name can be extracted from **Step 1.12** in **amplicon.bed** file)

	A	B	C
1	Amplicons	Mean Reads	Passed Thres
2	AMLV2_AMP15	11.88	TRUE
3	AMLV2_AMP46	12.46	TRUE
4	AMLV2_AMP27	14.79	TRUE
5	AMLV2_AMP25	21.62	TRUE
6	AMLV2_AMP5	23.6	TRUE
7	AMLV2_AMP113	25.6	TRUE
8	AMLV2_AMP28	27.14	TRUE
9	AMLV2_AMP9	27.25	TRUE
10	AMLV2_AMP99	31.6	TRUE
11	AMLV2_AMP109	35.42	TRUE
12	AMLV2_AMP77	35.56	TRUE
13	AMLV2_AMP13	35.82	TRUE
14	AMLV2_AMP14	36.51	TRUE

2.6.   If "Mean Reads" is < 10 the default filter settings in Tapestri Insights likely discard the variant due to "Read Depth" (2nd filter, default = 10) and/or due to "% of variants across cells" (4th filter, default = 50)

2.7.   Decreasing both filter threshold values in Tapestri Insights may recover the target variant (e.g., "Read Depth" from 10 to 4, and "% of variants across cells" from 50 % to 20%).
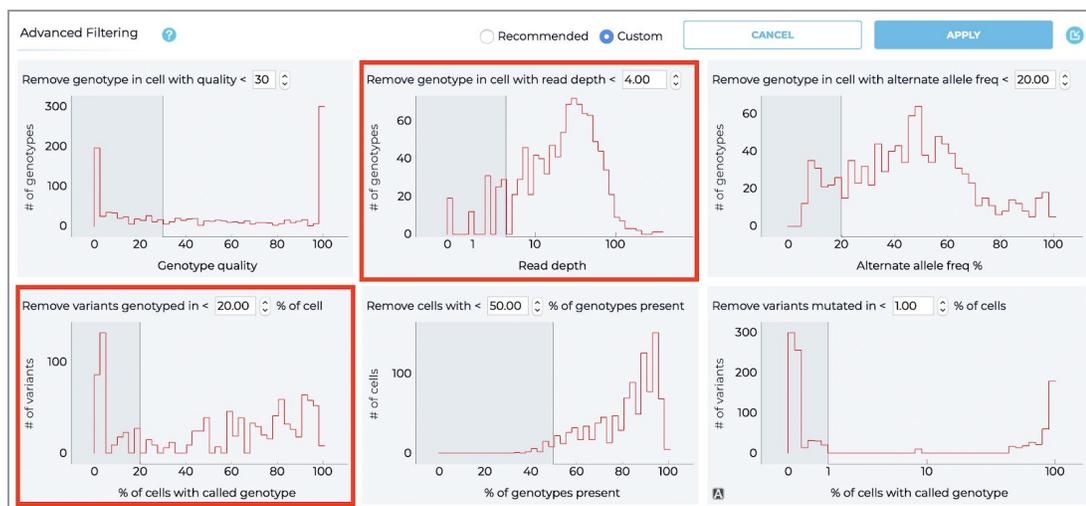
3. Work with BED file in Tapestri Insights

When working with bed files (or whitelists) in Tapestri Insights

Create custom BED file (whitelist)

3.1. Open text editor (e.g., Sublime Text editor) and create bed file
- First tab: chromosome number (e.g., chr17)
- Second tab: start of target coordinate (7577520)
- Third tab: end of target coordinate (7577525)

```
1    chr17    7577520 7577525
```

NOTE: Use the TAB key to separate each value. Do not use the SPACE key. If the target variant is a SNV the second and third tab coordinates are identical. Add as many target variants in separate rows.

3.2. Save file as .bed file (e.g., test.bed)

Review data in Tapestri Insights (TI) with custom BED file

3.3. Open the Tapestri Insights software, load the loom file(s) and add the custom bed file created in Step 3.1 under the *Whitelist* panel
3.4. Check the box *Remove all other variants*
3.5. Open the *Advanced Filtering* panel and set threshold settings for all 6 filters to 0
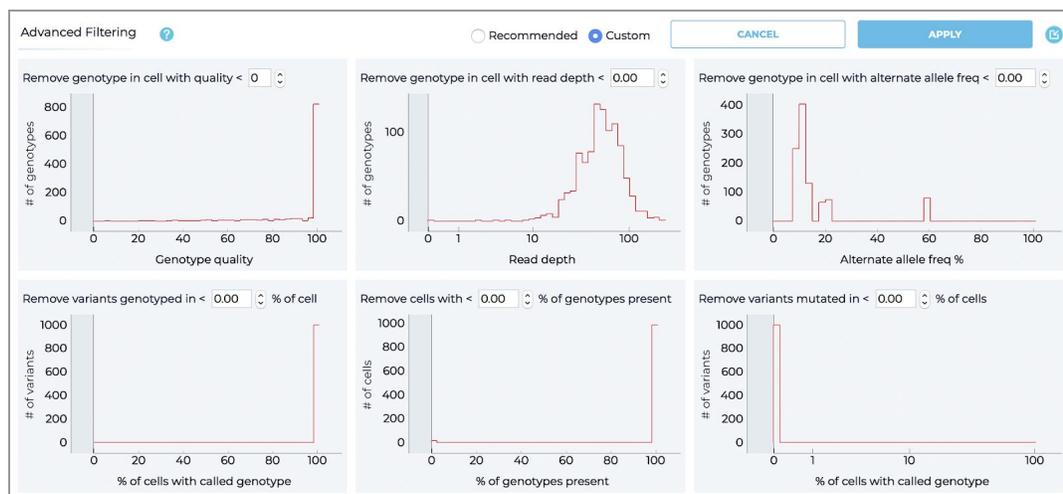
3.6. On the "Review Variants" page identify the target variant(s) and assess their quantitative metrics

Make sure to compare the "# Genotyped Cells" value on the "Review Variants" page to the total number of cells that is shown on the "Load Sample" page.

NOTE: *If the target variant is covered by the panel, each target nucleotide will result in at least two variants. For example a custom bed file with one target (row) and one nucleotide (SNV) will result in two variants detected in TI. This behavior is due to settings on the Tapestri Pipeline where GATK will report out 2 alternate alleles for each variant if at least one non-reference call is made.*

4.   Review BAM files

Prepare BAM files

4.1.   Download cells.bam file from your Tapestri Pipeline account

4.2.   Download and install the [Integrated Genome Viewer (IGV)](#) software

4.3.   Open IGV software, open IGV tools (Tools > Run igvtools) and index cells.bam file
  - In *Command* field select *Index* from dropdown menu
  - In *Input File* field select *cells.bam* file
  - With default parameters index files by clicking the *Run* button (process may take several minutes)
    The result is an index file (.BAI) in the same directory as the cells.bam file

4.4.   Close igvtools once completed

Review BAM files in IGV

4.5.   Open IGV software

4.6.   Load cells.bam file (File > Load from File) (indexed .BAI file will be loaded automatically)

4.7.   Enter target variant in search field (e.g., chr13:7577520)

4.8.   Wait for tracks to load and zoom out to evaluate Read1 and Read2 reads