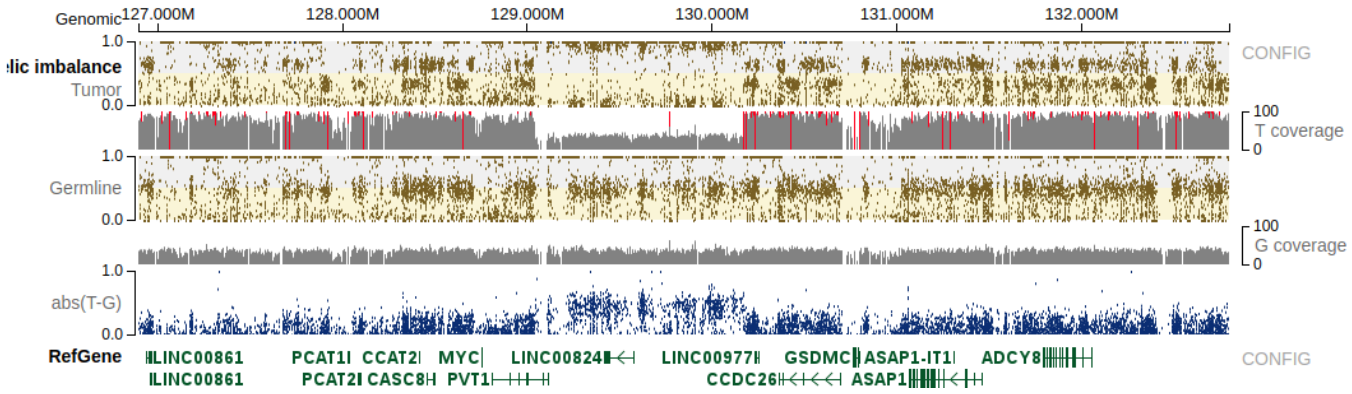
AIcheck track format and usage

“aicheck” is a term coined by Xiaotu Ma, who also designed the visualization to show the allelic imbalance of the heterozygous SNP markers in a tumor genome as compared to this patient’s germline genome, as a way of indicating loss-of-heterozygosity.



Live example: https://proteinpaint.stjude.org/?appcard=ai

The JSON definition for a aicheck track is:

```
{
      "type": "aicheck",
      "file": "files/hg19/example/aicheck/SJBALL020340_D1_sorted.gz",
      "name": "SJBALL020340 tumor allelic imbalance"
}
```

Submit this JSON text to the custom track panel to make it work. See custom track guide for details.

Additional JSON parameters:

- coveragemax: Integer
    - Maximum Y-axis value for coverage tracks of both tumor and normal; default: 100.
- vafheight: Integer
    - Variant allele fraction track height for both tumor and normal; default: 50.
- coverageheight: Integer

    - Coverage track height; default: 30.
- rowspace: Integer
    - Vertical spacing between rows; default: 5.

Following two optional parameters allow to filter markers.

- gtotalcutoff: Integer
    - Minimum total germline coverage. Markers with values below the cutoff will be excluded.
- gmafrestrict: float between 0 to 0.5
    - A limit on the germline B-allele fraction so as to only include markers with BAF close to 50%. E.g. by setting value 0.3 to this parameter, it will require 0.3 <= BAF <= 0.7. Markers with BAF outside this range will be excluded.

Aicheck track file has 6 columns, tab-separated. First line is a header, optional.

```
Chr     Pos     MinD    TinD    MinN    TinN
chr1    10003   3       38      0       23
chr1    10007   2       60      1       32
chr1    10009   3       58      1       32
chr1    10016   3       79      1       65
chr1    10019   1       94      2       76
chr1    10020   3       102     1       82
chr1    10025   2       127     1       91
chr1    10025   2       129     1       92
```

Each row is one germline heterozygous marker. All markers are from the same patient. Columns:

1. Chromosome name
2. Position of the SNP
3. Alternative allele read count in tumor DNA
4. Total read count of this SNP in tumor DNA
5. Alternative allele read count in germline DNA
6. Total read count of this SNP in germline DNA

After you assemble the marker data into a text file, do following to convert it to a track file:

```
$ sort -k1,1 -k2,2n input.file > input.sorted
$ bgzip input.sorted
$ tabix -c 'C' -s 1 -b 2 -e 2 input.sorted.gz
```

This generates two files “sorted.gz” and “sorted.gz.tbi”. Put both of them in the same path.