

# Amplicon Indel Hunter (AIH/AID)

## aiHunter test data

Sabah Kadri<sup>\*1</sup>, Chao Jie Zhen<sup>1</sup>, Michelle N Wurst<sup>1</sup>, Bradley C Long<sup>1</sup>, Zifeng F Jiang<sup>1</sup>, Y. Lynn Wang<sup>1</sup>, Larissa V Furtado<sup>1</sup>, and Jeremy P Segal<sup>†1</sup>

<sup>1</sup>The University of Chicago, Chicago, Illinois

This document contains information about example test data for a 27bp *FLT3* ITD supplied to be tested with aiHunter v1.1.0. Please do not distribute.

## Publication (please cite if using software):

Sabah Kadri, Chao J. Zhen, Michelle N. Wurst, Bradley C. Long, Zi-Feng Jiang, Y. Lynn Wang, Larissa V. Furtado, Jeremy P. Segal, **Amplicon Indel Hunter Is a Novel Bioinformatics Tool to Detect Large Somatic Insertion/Deletion Mutations in Amplicon-Based Next-Generation Sequencing Data**, The Journal of Molecular Diagnostics, Volume 17, Issue 6, November 2015, Pages 635-643, ISSN 1525-1578, <http://dx.doi.org/10.1016/j.jmoldx.2015.06.005>.  
(<http://www.sciencedirect.com/science/article/pii/S1525157815001506>)

---

<sup>\*</sup>skadri@bsd.uchicago.edu

<sup>†</sup>jsegal5@bsd.uchicago.edu

## Contents

<b>1</b>	<b>Input Data</b>	<b>3</b>
1.1	Test data . . . . .	3
<b>2</b>	<b>Outputs from AIH and AID</b>	<b>3</b>
2.1	Output files and formats . . . . .	3

# 1 Input Data

aiHunter requires the following input files in order to run:

```
--read1      Read1 Fastq file.  
--read2      Read2 Fastq file.  
--amp        Tab-delimited amplicon info file. See below.  
--inserts    Fasta file with insert sequences. See below.
```

Please see the detailed documentation for optional parameters for mutant frequency thresholds etc.

## 1.1 Test data

The test data can be run as following:

```
aiHunter --read1 example_data_R1_001.fastq --read2 example_data_R2_001.fastq  
--amp example_data.amplicons.txt --inserts example_data.inserts.fa
```

# 2 Outputs from AIH and AID

## 2.1 Output files and formats

The expected output files are included in the folder.

### 1. **example\_data\_R1\_001.fastq.R2fastq.indelcalls.txt:**

This is the output of the AIH module of the software. The fourth column in this file is the fraction of reads in your input fastq files that have some length-affecting mutation. This is a very helpful column to flag potential indel-containing amplicons. If the counts in the 4th column are 0, you have an error in your data. For the example data, the results should be:

Amplicon	Reads_w_indel>5bp	#readpairspassingfilter	%indel
FLT3_14	1674	3534	0.473684210526

### 2. **example\_data\_R1\_001.fastq.R2fastq.indelcalls.significant.txt:**

With the default MAF cutoff being 5%, the amplicon passes the threshold, and the contents of the file should be:

```
FLT3_14
```

### 3. **example\_data\_R1\_001.fastq.R2fastq.finalindelstats:**

Output of the AID module of the software. The file has no header. If AID is able to decipher the exact indel sequence, it will be displayed in this file or you will see **Failed indel identification - manual review required** for **<amplicon name>** error. The contents of this file for example data will be:

```
chr13 28608257 3524 3524 T 0.527809307605 +27CATATTCATATTCTCTGAAATCATAAA
0.472190692395
```

which correspond to

```
chrom position depth depth ref_allele ref_freq indel indel_freq
```

#### 4. **example\_data\_R1\_001.fastq.R2fastq.finalindelstats.vcf:**

A VCF formatted output of the annotated indels (if AID was able to annotate the indel). Other than the header, the output should contain:

```
chr13 28608257 . T TCATATTCATATTCTCTGAAATCATAAA . .
DP=3524;DP30=3524;AF=0.473;REF=T;ALT=TCATATTCATATTCTCTGAAATCATAAA;
GT:DP :
```