Filter.py – filters tumor variants based on read depth and VAF

#! /bin/python3.7

import sys

if len(sys.argv)<3:

print ('Usage: input1.vcf filteredsnps.vcf')

sys.exit()

else:

input1=sys.argv[1]

output1=sys.argv[2]

inputfile = open(input1, "r")

passedsnps = open(output1, "w")

column =[]

for line in inputfile:

line = line.strip('\n')

if line[0] == '#' and line[1] == '#':

passedsnps.write(line+'\n')

continue

if line[0] == '#':

column = line.split('\t')

column[0] = column[0].replace("#", "")

passedsnps.write(line+'\n')

continue

row = line.split('\t')

lineDat = dict(zip(column, row))

# print(column)

# print(row)

# print(lineDat)

lineDat.update({'TUMOR': row[10]})

formatKey = lineDat['FORMAT'].split(':')

tumorVals = dict(zip(formatKey, lineDat['TUMOR'].split(':')))

if int(tumorVals['AD'].split(',')[1]) < 2 and float(tumorVals['DP']) > 0:

try:

vaf = int(tumorVals['AD'].split(',')[1])/float(int(tumorVals['AD'].split(',')[1])+int(tumorVals['AD'].split(',')[0]))

except ZeroDivisionError:

vaf=0.0

if vaf < 0.1:

row[7] = "PASS"

passedsnps.write("\t".join(row)+ '\n')