#$/usr/bin/perl -w

# This script outputs heterozygous sites

# along with the number of supporting reads

# from a vcf file with a single sample.

#

# Input: vcf format file with one sample

# Output:

# Chromosome

# position

# read coverage

# # reads supporting reference allele

# # reads supporting alternative allele

# call

#

# Usage: perl extract\_heterozygous\_sites.pl input.vcf

# Read input file or die

$file = $ARGV[0];

$ofile = $file;

$ofile =~ s/vcf/ac.vcf/;

if (length($file)>0){

open(IN,"$file") or die "Can't open\n";

open(OUT,">$ofile") or die "Can't open\n";

} else {

print "No input file\nUsage: perl extract\_heterozygous\_sites.pl input.vcf\n";

die;}

# Read header

do {

$line = <IN>;

} until $line =~ m/CHROM/;

chomp($line);

@line = split(/\t/,$line);

$line[0] =~ s/#//;

print OUT "Chromosome.position,read coverage,ref reads,alt reads,call\n";

# Read vcf

while ($line = <IN>){

chomp($line);

@line = split(/\t/,$line);

if ($line[4] !~ m/,/){

@geno = split(/:/,$line[9]);

if ($geno[0] =~ m/0\/1|1\/0/){

print OUT "$line[0].$line[1],$geno[2],$geno[4],$geno[6],heterozygous\n";}

elsif ($geno[0] =~ m/1\/1|0\/0/){

print OUT "$line[0].$line[1],$geno[2],$geno[4],$geno[6],homozygous\n";}

else {

next();}

}

}

close(IN);

close(OUT);