# **Step 1**, prepare the data for segmentation by DNAcopy

# The genotype data are split into 20 files, one for each chromosome. We used the gt.convert function (see "gt\_convert.docx" in Github) to turn genotype codes "0/0", "0/1", and "1/1" into "0", "1", and "2", respectively, with missing or non-diploid genotypes converted into "NA".

# We extracted the sum of genotype counts for each site over the 12 BN samples

x1<-apply(conv.gt20.b[,16:27],1,function(x) sum(x==0,na.rm=T))

x2<-apply(conv.gt20.b[,16:27],1,function(x) sum(x==1,na.rm=T))

x3<-apply(conv.gt20.b[,16:27],1,function(x) sum(x==2,na.rm=T))

x4<-apply(conv.gt20.b[,16:27],1,function(x) sum(is.na(x)))

tmp20.bn.new<-cbind(x1,x2,x3,x4)

#here conv.gt20.b is the genotype table for chr 20, in which [,16:27] are the 12 BN samples. tmp20.bn.new contains genotype counts over 12 samples for the (0,1,2,NA) genotypes.

# We summed the hets and NA genotype counts, and used x for the genomic coordinates of the variant sites.

tmp<-tmp20.bn.new [,2]+ tmp20.bn.new [,4]

x<-as.numeric(rownames(tmp20.bn.new))

# we constructed the input data file per DNAcopy's requirements.

tmp.CNA<-CNA(log(tmp+2,2),rep(20,length(tmp)),as.numeric(x))

#CNA is the command to create the data object.

#here the final "track" is the hets + NA counts, with an added floor of 2, then undergoing log2 transformation. The components of "rep(20,length(tmp))" and "as.numeric(x)" are needed by DNAcopy.

#**Step 2**, run DNAcopy on the hets+NA track

library(DNAcopy)

peaks<- segment(tmp.CNA, min.width=5,alpha=0.001)

#segment is the command in DNAcopy. The parameters "min.width=5,alpha=0.001" were chosen to ensure a sensitive detection, to be followed by a merging step of the discovered segments.

#**Step 3**, merge the segments

#We used the merge.segments function (see "merge\_segment.docx in the Github folder) to join the segments.

#**Step 4**, repeat for all 20 chromosomes. This led to the final set of 673 segments.