



**WORKSHOP  
DE BIOINFORMÁTICA  
APLICADA À GENÔMICA E  
MELHORAMENTO ANIMAL**



# **AULA PRÁTICA 1**

## **Controle de qualidade de marcadores SNPs**



**Fabyano Fonseca e Silva**

**Prof. Adjunto IV - Dep Zootecnia – UFV  
Estatística Genômica e Bioinformática**

**Campo Grande, 14/07 a 15/07 de 2014**

**#lendo os dados originais**

```
dados=read.table("exemplo1_illumina.txt",h=T)
```

```
head(dados)
```

**#criando coluna de genotipos**

```
genot=paste(dados$Allele1_AB,dados$Allele2_AB,sep="")
```

```
head(genot)
```

**#criando dados com informacao de SNP, ID e coluna de genotipos**

```
dados1=cbind(dados[,1:2],genot)
```

```
head(dados1)
```

```
Nsnp=230
```

```
Nid=200
```

**#criando arquivos com ID nas linhas e SNPs nas colunas**

```
dados2=noquote(cbind(  
matrix(unique(dados$SampleID)),matrix(dados1$genot,Nid,Nsnp,byrow=TRUE)))
```

```
dados2[,1:10]
```

**#especificando nome das colunas**

```
SNP_name=matrix(t(unique(dados[,1])))
```

```
colnames(dados2)=c("id",t(SNP_name))
```

```
dados2[,1:10]
```

```
snp=dados2 #renomeando
```

```
#padronizando genótipos (pg 8 manual HapEstXXR)
```

```
snp[snp==".."]<-0 #missing pelo pacote HapEstXXR
```

```
snp[snp=="AA"]<-1
```

```
snp[snp=="BB"]<-3
```

```
snp[snp=="AB"]<-2
```

```
snp[,1:10]
```

```
write.table(snp,"illumina_HapEstXXR.txt",row.names=FALSE,quote=FALSE)
```

```
> snp[,1:10]
```

	id	rs11102647	rs6695241	rs12567796	rs2810583	rs4654986	rs1567710	rs1320964	rs2205847	rs10923099
[1,]	B-1	0	2	2	3	1	2	1	2	3
[2,]	B-2	2	1	1	3	1	3	1	3	3
[3,]	B-3	1	1	0	3	0	2	1	1	3
[4,]	B-4	2	1	2	2	2	1	3	3	2
[5,]	B-5	0	2	2	1	0	0	0	1	1
[6,]	B-6	3	1	2	2	1	3	3	2	2

**Arquivo pronto para a análise de qualidade no pacote HapEstXXR**

```
library(HapEstXXR) #dados reais com 500 ID's e 1000 SNPs
```

```
snp0=as.matrix(read.table("HapEstXXR_real.txt",h=T)) #lendo os dados já formatados  
head(snp0)  
snp=snp0[,-1] #removendo coluna de indivíduos  
head(snp)
```

```
snp_names=colnames(snp) #especificando identificação de cada SNP  
snp_names
```

```
a=data.frame(maf(snp, marker.label=snp_names)) #função maf do pacote HapEstXXR  
head(a)
```

```
#ATENÇÃO NOS VALORES LIMITES
```

```
snp1<-snp[,a$call.rate>=0.95 & a$maf>=0.05 & a$hwe.chisq.p.value>=0.05/ncol(snp)]
```

```
snp1=t(snp1)  
dim(snp1)
```

```
snp2=t(na.omit(snp1)) #elinando as colunas com Nas (com os SNPs excluídos)
```

```
dim(snp2)  
head(snp2)
```

```
#padronizando genótipos conforme pacotes de análises (0, 1 e 2)
```

```
snp2[snp2==0]<-3 #substituindo missing (0 do pacote HapEstXXR) pelo genotipo  
heterozigoto (imputação?)
```

```
snp2[snp2==1]<-0 #aa  
snp2[snp2==3]<-1 #aA  
snp2[snp2==2]<-2 #AA  
dim(snp2)
```

```
write.table(snp2,"snp_fim.txt",row.names=FALSE,quote=FALSE) #exportar arquivo pronto
```

```
> snp2
      snp201 snp202 snp203 snp204 snp205 snp206 snp207 snp208 snp209 snp210
1          0          0          0          0          0          1          0          1          1          1
2          0          0          0          0          0          0          0          0          0          2
3          1          0          0          1          0          1          0          1          1          1
4          0          1          0          1          1          0          1          0          1          1
5          1          0          1          1          1          1          0          1          1          2
6          1          0          1          1          1          1          0          1          0          1
7          1          0          0          0          1          1          0          1          1          2
8          0          0          1          2          2          2          0          2          1          1
9          0          0          1          1          1          1          0          1          0          1
10         1          0          1          1          1          0          1          0          1          1
11         0          0          0          2          2          2          0          2          2          1
12         1          0          1          1          1          0          1          0          1          1
13         1          0          1          1          1          0          1          0          1          1
14         0          0          1          1          1          0          1          0          1          1
15         1          0          1          2          2          1          1          1          2          1
16         0          0          0          0          0          0          0          0          0          2
```

**Arquivo pronto para ser utilizado nas análises estatísticas!**