**BMK\_1\_SNP\_Analysis SNP分析结果**

**=====================================文件=====================================**

**|-- AllSample.SNP\_density.png SNP类型密度图**

横轴为基因上平均每1000bp序列中分布的SNP数目，纵轴为基因数。

**|-- AllSample.SNP\_density.stat.xls SNP类型密度统计结果文件**

|-- Sample：样品名称；

|-- Interval：SNP 密度区间(个/ Kbp ) ；

|-- GeneNum：在相应(第二列)密度区间的基因数。

**|-- AllSample.snp.stat.xls** **SNP位点统计表**

|-- #BMK-ID：百迈客对样品的统一编号；

|-- SNP Number： SNP位点总数；

|-- Genic SNP：基因区SNP位点总数；

|-- Intergenic SNP：基因间区SNP位点总数；

|-- Transition：转换类型的SNP位点数目在总SNP位点数目中所占的百分比；

|-- Transversion：颠换类型的SNP位点数目在总SNP位点数目中所占的百分比；

|-- Heterozygosity：杂合型SNP位点数目在总SNP位点数目中所占的百分比。

**|-- final.\*.anno.gatk.all.list.xls SNP/Indel位点信息文件**

|-- Chr：SNP/InDel位点所在染色体编号；

|-- Pos：SNP/InDel位点在染色体上的位置；

|-- Gene\_id：SNP/InDel位点所在的基因或原来未注释的基因区(表中用Intergenic表示)；

|-- Ref：所选参考基因组中的SNP/InDel等位；

|-- Alt：测序样品中识别到的SNP/InDel；

|-- \*\*：样品\*\*中SNP/InDel位点的分型，SNP类型见附录；

|-- Depth：样品\*\*中SNP/InDel位点的测序深度；

|-- AlleDp：样品\*\*中SNP/InDel位点各等位点的支持深度；

|-- Effect：SNP/InDel所在区域或类型,Effect具体说明详见：

[http：//snpeff.sourceforge.net/SnpEff\_manual.html](http：/snpeff.sourceforge.net/SnpEff_manual.html)

|-- Codon\_change：编码改变方式,未改变用点表示。

########################################

其中Alt列中的 SNP 位点的碱基缩写表见下:

Nucleic\_Acid\_Code Meaning Mnemonic 中文名称

A A Adenine 腺嘌呤

C C Cytosine 胞嘧啶

G G Guanine 鸟嘌呤

T T Thymine 胸腺嘧啶

U U Uracil 尿嘧啶

R A or G puRine 嘌呤

Y C,T or U pYrimidines 嘧啶

K G,T or U bases which are Ketones 酮基

M A or C bases with aMino groups 氨基

S C or G Strong interaction 强相互作用

W A,T or U Weak interaction 弱相互作用

B not A (i.e. C,G,T or U) B comes after A 非腺嘌呤

D not C (i.e. A,G,T or U) D comes after C 非胞嘧啶

H not G (i.e. A,C,T or U) H comes after G 非鸟嘌呤

V neither T nor U (i.e. A, C or G) V comes after U 既非胸腺嘧啶也非尿嘧啶

N A C G T U Nucleic acid 核酸

SnpEff是一款用于注释变异（SNP、InDel）和预测变异影响的软件。根据变异位点在参考基因组上的位置以及参考基因组上的基因位置信息，可以得到变异位点在基因组发生的区域（基因间区、基因区或CDS区等），以及变异产生的影响（同义非同义突变等）。

**=====================================文件夹1==================================**

**|-- BMK\_1\_InDel\_anno Indel注释结果文件夹**

**|-- all.indel.anno.stat.png** **Indel注释分类图**

纵轴为SNP/Indel所在区域或类型，横轴为分类数目。

**|-- \*.indel.anno.stat.png** **单个样本Indel注释分类图**

**|-- final\_Indel.anno.stat.xls** **Indel注释分类统计文件**

**=====================================文件夹2==================================**

**|-- BMK\_2\_SNP\_anno SNP注释结果文件夹**

**|-- all.snp.anno.stat.png SNP注释分类图**

纵轴为SNP/Indel所在区域或类型，横轴为分类数目。

**|-- \*.snp.anno.stat.png** **单个样本SNP注释分类图**

**|-- final\_SNP.anno.stat.xls**  **SNP注释分类统计文件，常见SNP类型分类见结题报告2.6.1**

**=====================================文件夹3==================================**

**|--** **BMK\_3\_SNP\_type SNP类型结果文件夹**

**|-- All.snp.type.png**  **SNP突变类型分布图**

横轴为SNP突变类型，纵轴为相应的SNP数目。

**|-- \*.snp.type.png** **单个样本SNP突变类型分布图**

**|-- All.snp\_type.stat.xls**  **SNP突变类型统计文件**

常见SNP类型：

###################################################

(1) INTERGENIC: intergenic\_region 突变发生在基因间区；

(2) INTRAGENIC: intragenic\_variant 突变发生在基因区，但是在转录本所有的属性区域外；

(3) INTRON: intron\_variant 突变发生在内含子区；

(4) UPSTREAM: upstream\_gene\_variant 突变发生在基因上游（默认长度：5K bases）；

(5) DOWNSTREAM: downstream\_gene\_variant 突变发生在基因下游（默认长度：5K bases）；

(6) UTR\_5\_PRIME: 5\_prime\_UTR\_variant 突变发生在5'UTR区；

(7) UTR\_3\_PRIME: 3\_prime\_UTR\_variant 突变发生在3'UTR区；

(8) SPLICE\_SITE\_ACCEPTOR: splice\_acceptor\_variant 突变发生在可变剪切受体位点（一般认为外显子起始位置的前两个碱基，第一个外显子除外）；

(9) SPLICE\_SITE\_DONOR: splice\_donor\_variant 突变发生在可变剪切施体位点（一般认为编码外显子结束位置的后两个碱基，最后一个外显子除外）；

(10) SPLICE\_SITE\_REGION: splice\_region\_variant 突变发生在可变剪切区域，外显子的1-3个碱基或者内含子的3-8个碱基；

(11)START\_GAINED: 5\_prime\_UTR\_premature start\_codon\_gain\_variant 突变发生在5‘UTR产生起始密码子的3碱基序列中；

(12) START\_LOST: start\_lost 突变发生在起始密码子中，并使得起始密码子突变为非起始密码子。

(13) SYNONYMOUS\_CODING: synonymous\_variant 同义突变；

(14) NON\_SYNONYMOUS\_CODING: missense\_variant 错义突变；

(15) START\_LOST: start\_lost 突变发生在起始密码子中，并使得起始密码子突变为非起始密码子；

(16) SYNONYMOUS\_STOP：stop\_retained\_variant 突变使得终止密码子突变为另一个终止密码子；

(17)STOP\_GAINED：stop\_gained 突变产生一个终止密码子；

(18)STOP\_LOST：stop\_lost 突变使得终止密码子突变为非终止密码子。