

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
|-- 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

SnEff 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。