

Genome re-sequencing pipeline @ GATC Biotech

SNP and InDel report



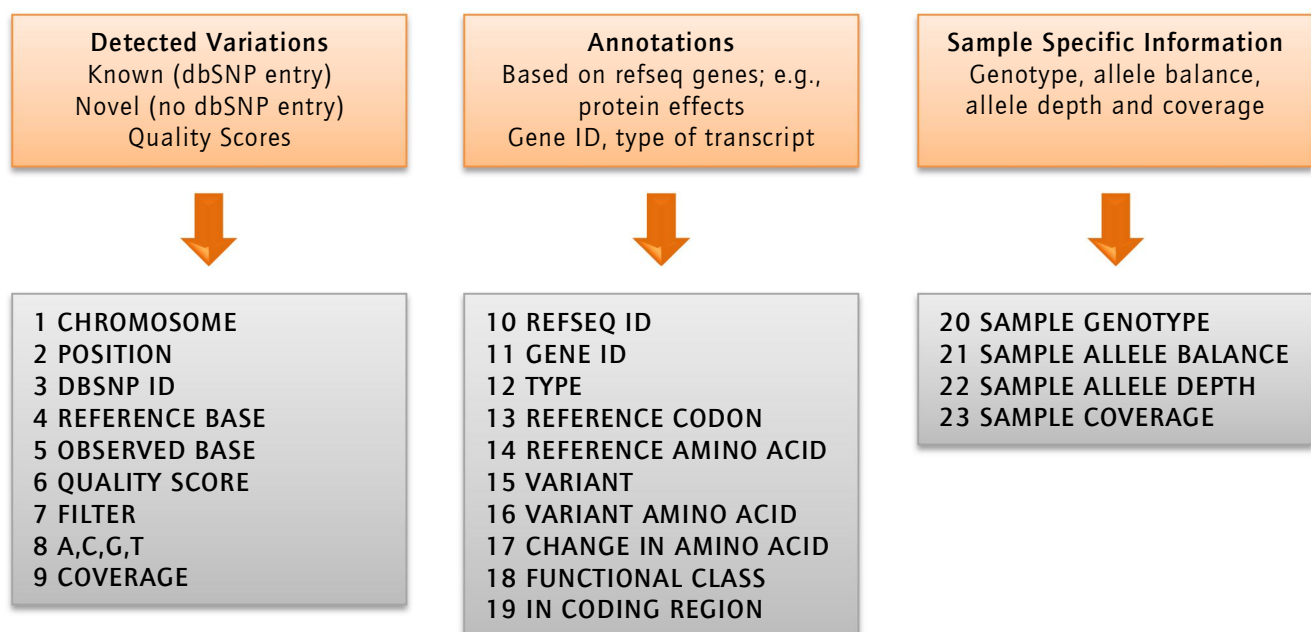
SNP and InDel table

CHR	POSITION	DBSNP ID	REF BASE	OBSERVED BASE	QUALITY SCORE	FILTER*	A,C,G,T*	COV*	REFSEQ ID	GENE ID
chr3	164777677	rs9290259	T	G	9919.08	PASS	0,0,306,0	306	NM_001041	SI
chr19	53911973	rs10425136	A	G	9853.53	PASS	0,0,252,0	252	NM_001040185	ZNF765
chr12	55945119	rs7313899	A	G	9616.99	PASS	0,0,281,0	281	NM_001005494	OR6C4
chr3	46399798	rs1799865	T	C	9551.17	PASS	0,254,0,0	254	NM_001123041	CCR2
chr12	10570965	rs2682495	C	G	9476.45	PASS	0,0,278,0	278	NM_002261	KLRC3

TYPE*	REF CODON*	REF AA*	VARIANT CODON*	VARIANT AA*	CHANGE IN AA*	FUNCTIONAL CLASS	IN CODING REGION	SAMPLE GENOTYPE	SAMPLE ALLELE BALANCE	SAMPLE ALLELE DEPTH
intron	NA	NA	NA	NA	NA	NA	no	1/1	NA	0.306
CDS	AGC	Ser	GGC	Gly	yes	missense	yes	1/1	NA	0.252
CDS	ATC	Ile	GTC	Val	yes	missense	yes	1/1	NA	0.281
CDS	AAT	Asn	AAC	Asn	no	silent	yes	1/1	NA	0.254
CDS	TGT	Cys	TCT	Ser	yes	missense	yes	1/1	NA	0.278

* not contained in InDel table

Content



Column description

Column	Description
1 CHROMOSOME	Name of chromosome (or reference contig) where the variant occurs
2 POSITION	Position of chromosome (or reference contig) where the variant occurs
3 DBSNP ID	The dbSNP rs identifier of the SNP based on the contig or chromosome position of the call. If there is an entry in the dbSNP then the respective rs id will be displayed. Dot ('.') indicates no entry in the dbSNP.
4 REFERENCE BASE	The reference base at the variant site
5 OBSERVED BASE	Alternative (observed) base in the samples in general [VARIANT]
6 QUALITY SCORE	The Phred scaled probability of REFERENCE/OBSERVED polymorphism exists at this site given sequencing data. The value is computed based on error models designed by Broad Institute. Since the Phred scale is $-10 * \log(1-p)$, a value of 10 indicates a 1 in 10 chance of error, while a 100 indicates a 1 in 10^{10} chance. The higher the value the more accurate is the variant call.
7 FILTER	In addition to quality score, several filters are defined to filter the SNPs by considering factors other than quality score alone. More than one filter is defined and applied to the variant calls. Default filters are SnpCluster (more than 2 SNPs found in cluster of size=10), LowQual (SNP with quality score < 30), LowCov (SNP with coverage < 20), Mask (SNP is at least 10 base near to indel location) and HardToValidate (Not enough evidence to validate). Variant passing the default filters are tagged "PASS"
8 A,C,G,T	All the observed bases at the variant position
9 COVERAGE	Sequencing depth or coverage at the variant position. More accurate is to see the SAMPLE:COVERAGE
10 REFSEQ ID	The refseq entry associated with the location of the variant call. If present, refSeq ID will be displayed. if not, "NA" will be displayed
11 GENE ID	The gene entry associated with the location of the variant call. If present, Gene ID will be displayed. if not, "NA" will be displayed
12 TYPE	Annotation type e.g., CDS, 3UTR, 5UTR, INTRON etc.,
13 REFERENCE CODON	Reference codon at the variant call e.g., TCG
14 REFERENCE AMINO ACID	Corresponding reference aminoacid at the variant call e.g., Ser(ine)
15 VARIANT CODON	Variant codon
16 VARIANT AMINO ACID	Corresponding variant aminoacid
17 CHANGE IN AMINO ACID	Is there a change in aminoacid? True Yes or False No
18 FUNCTIONAL CLASS	Functional class of the SNP - silent (synonymous), missense (non-synonymous), nonsense (stop-gaining), readthrough (stop-loss) and NA (unclassified)
19 IN CODING REGION	Is the observed variant in the coding region? True Yes or False No
20 SAMPLE:GENOTYPE	The genotype of the sample. For a diploid, the GENOTYPE indicates the two alleles carried by the sample, encoded by a 0 for the REFERENCE allele, 1 for the first ALTERNATIVE (OBSERVED) allele. Possible GENOTYPES are 0/0 (the sample is homozygous to reference), 0/1 (the sample is heterozygous, carrying 1 copy of each of the REFERENCE and ALTERNATIVE alleles) and 1/1 (the sample is homozygous alternate i.e., completely opposite to the REFERENCE)
21 SAMPLE:ALLELE BALANCE	Allele balance is a ratio of the REFERENCE bases to the total bases observed in the give position. This applies for only heterozygous calls and value ranges from > 0.0 to < 1.0
22 SAMPLE:ALLELE DEPTH	The allele depth, one for each REFERENCE and ALTERNATIVE (OBSERVED), is the count of all reads that carried with them the respective alleles. The read counts also include the poor mapping quality reads, unlike the COVERAGE counts.
23 SAMPLE:COVERAGE	The total depth of the reads that passed the internal quality control metrics (for e.g., mapping quality >17) from all reads present at this site.