Analysis Summary

GATC Biotech AG

April, 2015

1 Analysis Summary

Single nucleotide polymorphisms (SNPs) and insertion & deletions (InDels) are detected by aligning Sanger reads to the reference sequence. From the best alignment, SNP and InDel candidates are screened, taking into account the quality value of the bases with variation as well as the quality values in the neighboring bases, using neighborhood quality standard. Ssaha2[1] is used for the alignment and ssaha2SNP[2] is used for detecting polymorphisms.

Table 1. Example SNP table, delivered as .csv file, exemplarily in Microsoft Office Excel 2010

REFERENCE	POSITION	REF.BASE	ALT.BASE	QUALITY	QUERY.NAME	QUERY POSITION	QUERY LENGTH
reference	73	G	C	51	HT-GATC-26_A01	108	497
reference	82	G	A	51	HT-GATC-26_A01	117	497
reference	91	G	A	56	HT-GATC-26_A01	126	497
reference	145	A	T	51	HT-GATC-26_A01	180	497
reference	154	C	T	56	HT-GATC-26_A01	189	497
reference	237	G	A	56	HT-GATC-26_A01	272	497
reference	394	A	G	51	HT-GATC-26_A01	432	497
reference	481	A	G	51	HT-GATC-26_A01	519	497
reference	487	A	G	56	HT-GATC-26_A01	525	497

The description is as follows

REFERENCE	Name of reference contig or chromosome where the variant occurs
POSITION	Position of reference contig or chromosome where the variant occurs
REF.BASE	The reference base at the variant site
ALT.BASE	Alternative (observed) base in the samples in general [VARIANT]
QUALITY	The Phred scaled probability of OBSERVED BASE is correct at this site given sequencing data. Since the Phred scale is 10 * log(1p), a value of 10 indicates a 1 in 10 chance of error, while a 100 indicates a 1 in 10ˆ10 chance. The higher the value the more accurate is the variant call.
QUERY.NAME	Name of the query (Sanger read)
QUERY.POSITION	Position of the query where the variant is observed
QUERY.LENGTH	Length of the query sequence in bp
REFERENCE	Name of reference contig or chromosome where the variant occurs

Table 2. Example InDel table, delivered as .csv file, exemplarily in Microsoft Office Excel 2010

REFERENCE	POSITION	INS.DEL	REF.BASE	ALT.BASE	INDEL LENGTH	QUERY.NAME	QUERY POSITION	QUERY LENGTH
reference	123	INS	---	TCC	3	HT-GATC-26_A01	123	497
reference	245	DEL	A	-	1	HT-GATC-26_A02	248	497
reference	364	INS	---	G	1	HT-GATC-26_A03	367	497
reference	370	DEL	ACT	---	3	HT-GATC-26_A04	373	497

The description is as follows

REFERENCE	Name of reference contig or chromosome where the variant occurs
POSITION	Position of reference contig or chromosome where the variant occurs
INS.DEL	Type of InDel - Insertion (INS) or Deletions (DEL)
REF.BASE	The reference base at the variant site
ALT.BASE	Alternative (observed) base in the samples in general [VARIANT]
INDEL.LENGTH	Length of the Insertion or Deletion observed in bp
QUERY.NAME	Name of the query (Sanger read)
QUERY.POSITION	Position of the query where the variant is observed
QUERY.LENGTH	Length of the query sequence in bp

1.1 References

[1] http://www.sanger.ac.uk/resources/software/ssaha2/

[2] http://www.sanger.ac.uk/resources/software/ssahasnp/

GATC Biotech AG (Hauptsitz)
European Genome and
Diagnostics Centre
Jakob-Stadler-Platz 7
78467 Konstanz
Germany
Tel. +49 (0) 7531 - 81 60 0
Fax +49 (0) 7531 - 81 60 81

GATC Biotech AG
European Custom Sequencing
Centre
Gottfried-Hagen-Straße 20
51105 Köln
Germany