alpha version

The -format described here has been applied to describe the splice site variants of the Geuvadis Project.

CHR	POS	ID	REF	ALT	QUAL	FILT	INFO	FORMAT	I1	I2
20	14370	-14370 ^20	TTGTACG TG	ttgtaGgtg, ttgtCcgtg	-10001	q-1000	MOD=ALT;ALT1=SNP1;ALT2=SNP2;VAR_SCORES=1. 5277311,3.4223458;SNPS=SNP1,SNP2	GT	0|1	0|0
...										

Description of variations of and extensions to the standard VCF

Attribute	Description
POS	the position of the splice site, provided as the first/last position included in the adjacent exon (cf. AStalavista default coordinates)
ID	string identifying the splice site uniquely, composed by strand, genomic coordinate, site type symbol (cf. AStalavista conventions) and chromosome ID
REF	the reference sequence of the splice site, as obtained by extracting the corresponding sequence stretch from the genome
ALT	comma-separated list of the corresponding splice site sequences *after* applying the corresponding genetic variant(s) to the reference sequence
QUAL	the score of the reference splice site sequence (as shown in REF)
FILT	"q-1000" when the splice site sequence has not observed in the training set (-Infinity, in practice represented by a value << -1000), "PASS" otherwise
INFO	MOD: either alternative (ALT) or constitutive (CON) splice site ALTx :(combinations of) variants that form each alternative variant (same ordering as in ALT column) VAR_SCORES: comma-separated list of scores assigned to the corresponding variant(s); the ordering corresponds to the one used for ALT SNPS: concatenation of *all* variants considered for the description of this splice site (line)