

.VCL format

The VCF (Variant Call Format) has been developed to standardize the storage of genetic variation, as employed in the [1000 Genomes Project](#). Here we describe an adaptation of the VCF for describing variation in splice sites, we call VCL (VCF-like). The basic concept of our VCF adaptation to describe splice site heterogeneity is to collect all information on variants that affect a certain splice site within a single line, similar to multi-variant descriptions. We propose to extend the standard VCF characterization by specific splice site attributes:

Attribute	Description
Modality	constitutive or alternative splice site
Sequence	splice site sequence (and possibly derived sequences that as altered by variants)
Score	thermodynamic affinity of the splice site, as expressed by its agreement with the species consensus; in our implementation we are applying logarithmic measures of splice site calling probabilities as proposed by the gene finding approach GeneID