**VCF format**

<http://gatkforums.broadinstitute.org/gatk/discussion/1268/what-is-a-vcf-and-how-should-i-interpret-it>

**6. How to extract information from a VCF in a sane, (mostly) straightforward way**

Use [VariantsToTable](https://www.broadinstitute.org/gatk/guide/tooldocs/org_broadinstitute_gatk_tools_walkers_variantutils_VariantsToTable.php).

No, really, **don't write your own parser** if you can avoid it. This is not a comment on how smart or how competent we think you are -- it's a comment on how annoyingly obtuse and convoluted the VCF format is.

Seriously. The VCF format lends itself really poorly to parsing methods like regular expressions, and we hear sob stories all the time from perfectly competent people whose home-brewed parser broke because it couldn't handle a more esoteric feature of the format. We know we broke a bunch of people's scripts when we introduced a new representation for spanning deletions in multisample callsets. OK, we ended up replacing it with a better representation a month later that was a lot less disruptive and more in line with the spirit of the specification -- but the point is, that first version was technically legal by the 4.2 spec, and that sort of thing can happen *at any time*. So yes, the VCF is a difficult format to work with, and one way to deal with that safely is to not home-brew parsers.

(Why are we sticking with it anyway? Because, as Winston Churchill famously put it, VCF is the worst variant call representation, except for all the others.)

/home/afedorov/1000GENOMES

esv2663150\_10lines.txt and Excel file

vi resultJune8 (se nowrap) example with deletions

esv2663150 deletion on chr 19

<http://www.ncbi.nlm.nih.gov/dbvar/variants/esv2663150/>

/home/afedorov/DENISOV

zcat T\_hg19\_1000g.19.mod.vcf.gz |more

rs59558746 at 544960