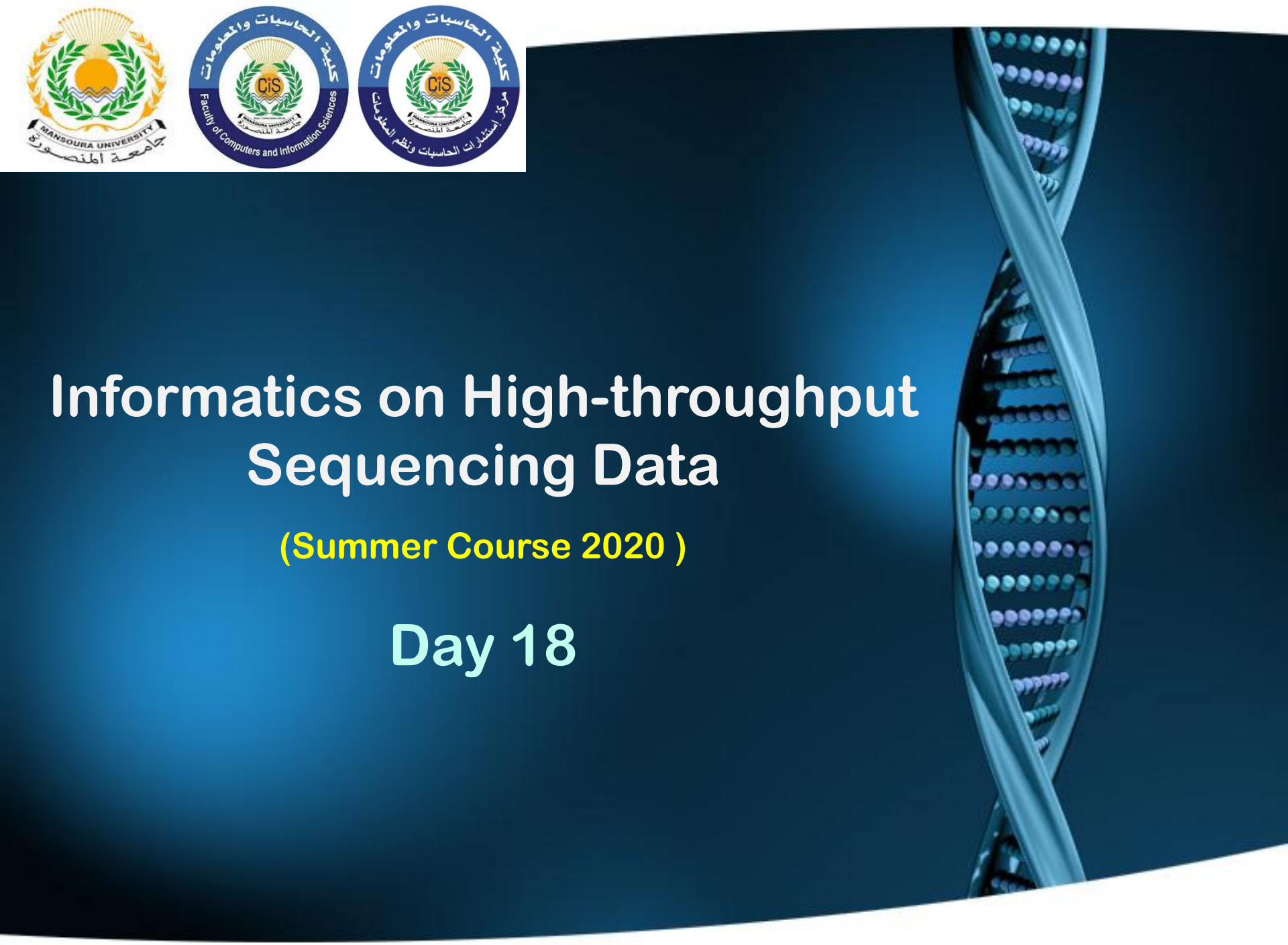
Informatics on High-throughput Sequencing Data

(Summer Course 2020)

Day 18



VCF/BCF format

- VCF is the standard file format for storing variation data.
- VCF files are tab delimited text files.

Types of variants

SNPs

Alignment	*VCF representation*			
ACGT	POS	REF	ALT
ATGT	2	C	T

Insertions

Alignment	*VCF representation*		
AC-GT	POS	REF	ALT
ACTGT	2	C	CT

Deletions

Alignment	*VCF representation*		
ACGT	POS	REF	ALT
A--T	1	ACG	A

Complex events

Alignment	*VCF representation*		
ACGT	POS	REF	ALT
A-TT	1	ACG	AT

Large structural variants

VCF representation

POS	REF	ALT	INFO
100	T	<DEL>	SVTYPE=DEL;END=300

http://vcftools.sourceforge.net/VCF-poster.pdf

http://digitheadslabnotebook.blogspot.com/2013/01/vcf-variant-call-format.html

VCF/BCF format

The variant call format and VCFtools

Petr Danecek, Adam Auton, Goncalo Abecasis, Cornelis A. Albers, Eric Banks,
Mark A. DePristo, Robert E. Handsaker, Gerton Lunter, Gabor T. Marth, Stephen T. Sherry
... Show more
Author Notes

Bioinformatics, Volume 27, Issue 15, 1 August 2011, Pages 2156–2158,
https://doi.org/10.1093/bioinformatics/btr330
Published: 07 June 2011 **Article history** ▾

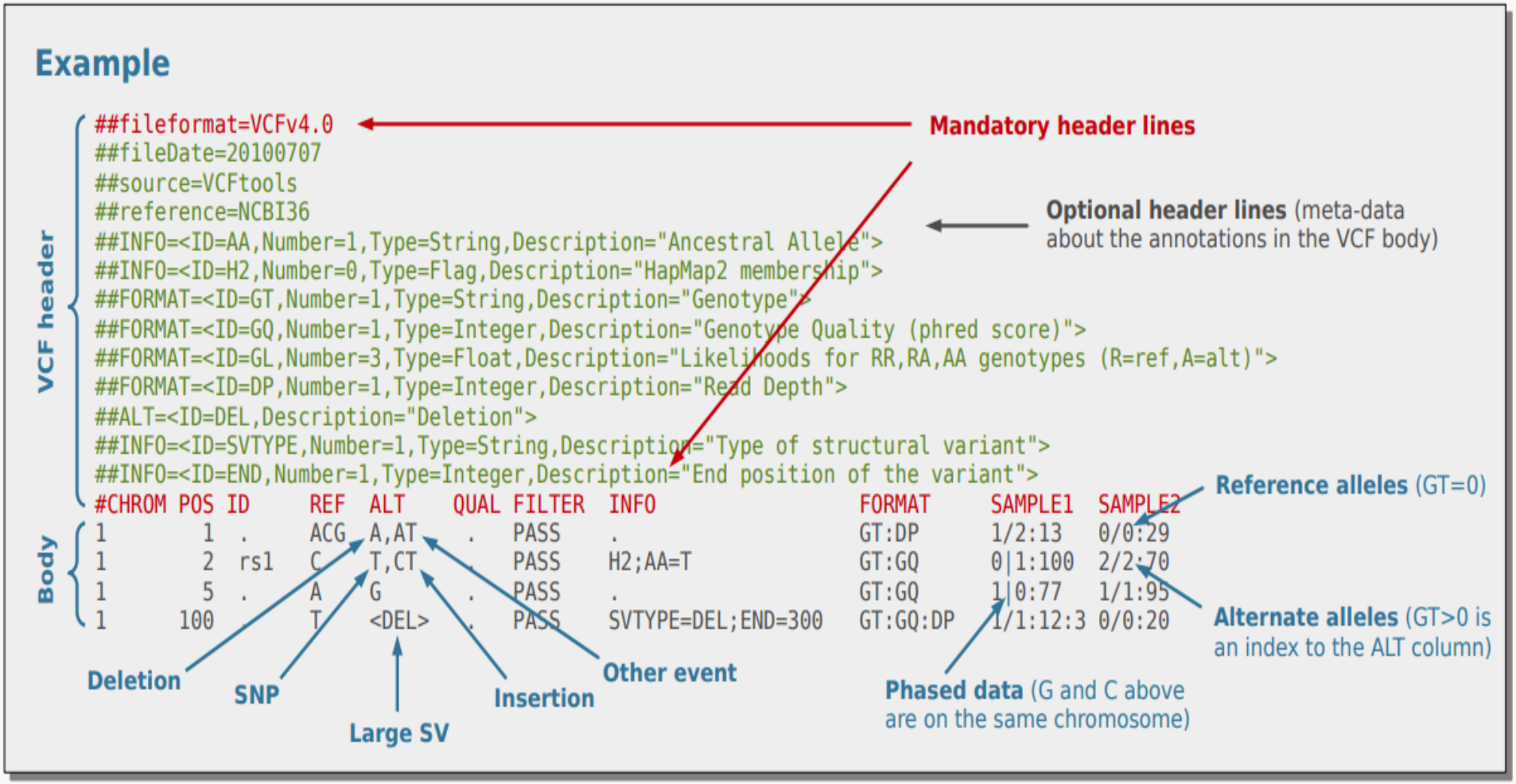
PDF Split View Cite Permissions Share ▾

Abstract

Summary: The variant call format (VCF) is a generic format for storing DNA polymorphism data such as SNPs, insertions, deletions and structural variants, together with rich annotations. VCF is usually stored in a compressed manner and can be indexed for fast data retrieval of variants from a range of positions on the reference genome. The format was developed for the 1000 Genomes Project, and has also been adopted by other projects such as UK10K, dbSNP and the NHLBI Exome Project. VCFtools is a software suite that implements various

https://academic.oup.com/bioinformatics/article/27/15/2156/402296

VCF/BCF format



https://samtools.github.io/hts-specs/VCFv4.2.pdf
http://vcftools.sourceforge.net/VCF-poster.pdf

VCF/BCF format

```
##fileformat=VCFv4.2
##FILTER=<ID=PASS,Description="All filters passed">
##samtoolsVersion=1.10+htslib-1.10
##samtoolsCommand=samtools mpileup -v -u -f /Users/sarael-metwally/Documents/Summer/bwa/wu_0.v7.f
##reference=file:///Users/sarael-metwally/Documents/Summer/bwa/wu_0.v7.fas
##contig=<ID=Chr1,length=29923332>
##contig=<ID=Chr2,length=19386101>
##contig=<ID=Chr3,length=23042017>
##contig=<ID=Chr4,length=18307997>
##contig=<ID=Chr5,length=26567293>
##contig=<ID=chloroplast,length=154478>
##contig=<ID=mitochondria,length=366924>
##ALT=<ID=*,Description="Represents allele(s) other than observed.">
##INFO=<ID=INDEL,Number=0,Type=Flag,Description="Indicates that the variant is an INDEL.">
##INFO=<ID=IDV,Number=1,Type=Integer,Description="Maximum number of reads supporting an indel">
##INFO=<ID=IMF,Number=1,Type=Float,Description="Maximum fraction of reads supporting an indel">
```

SAM tools

- `./samtools mpileup -v -u -f /Users/sarael-metwally/Documents/Summer/bwa/wu_0.v7.fas sample.sorted.bam > sample.vcf`
- `./samtools mpileup -g -f /Users/sarael-metwally/Documents/Summer/bwa/wu_0.v7.fas sample.sorted.bam > sample.bcf`

Thanks!

// | ?