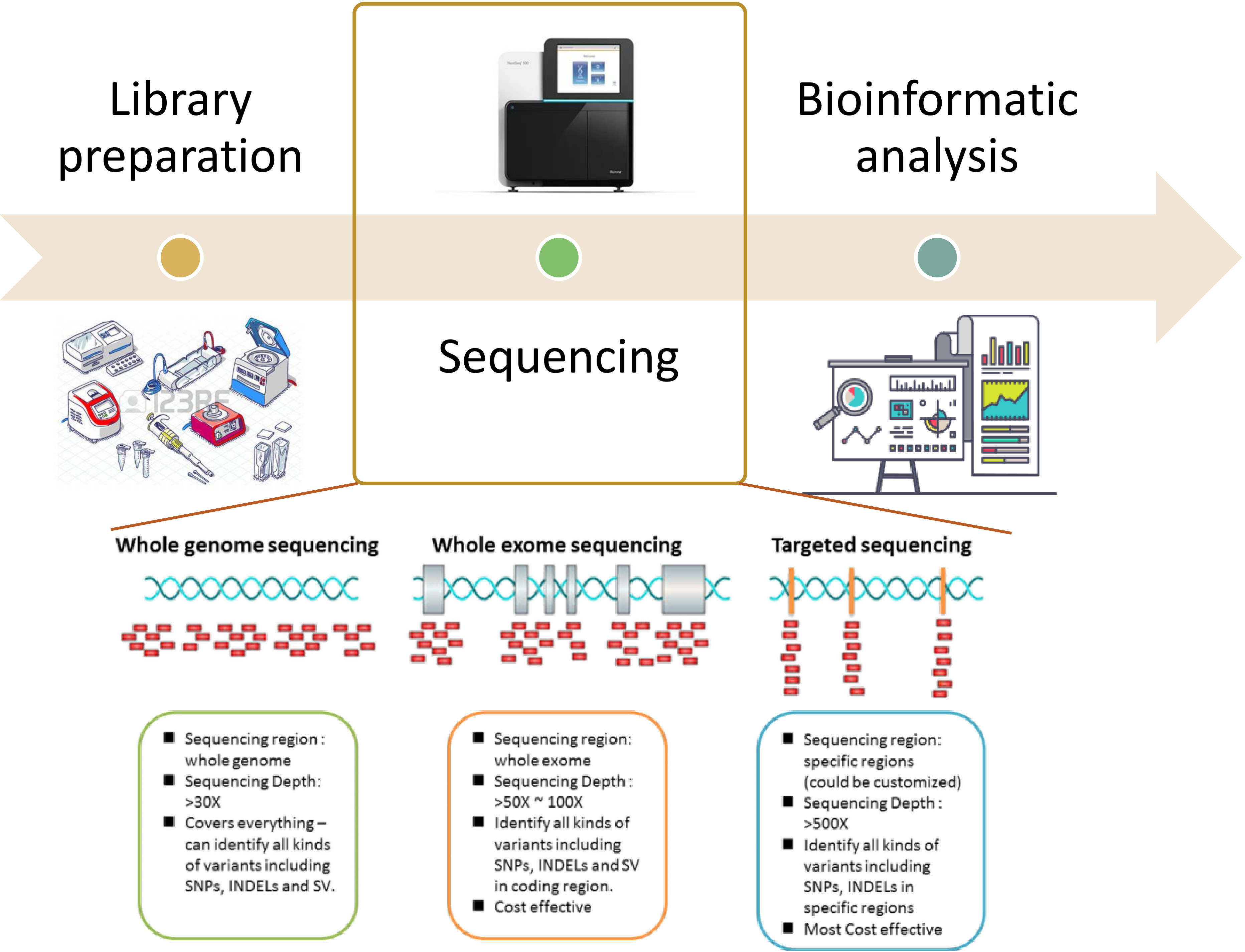
Human Genome Sequencing and Interpretation

Lesson 1 - 20/01/2020
Lesson 2 - 21/01/2020
Lesson 3 - 27/01/2020
(Lesson 4 - 28/01/2020)

Prof. Massimo Delledonne
Functional Genomics lab



Whole genome sequencing

- Sequencing region : whole genome
- Sequencing Depth: >30X
- Covers everything – can identify all kinds of variants including SNPs, INDELs and SV.

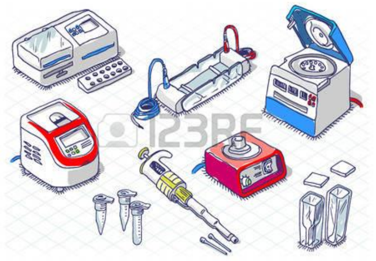
Whole exome sequencing

- Sequencing region: whole exome
- Sequencing Depth : >50X ~ 100X
- Identify all kinds of variants including SNPs, INDELs and SV in coding region.
- Cost effective

Targeted sequencing

- Sequencing region: specific regions (could be customized)
- Sequencing Depth : >500X
- Identify all kinds of variants including SNPs, INDELs in specific regions
- Most Cost effective

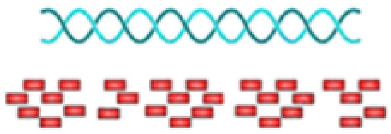
Library preparation



Sequencing

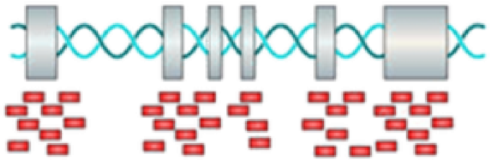
Bioinformatic analysis

Whole genome sequencing



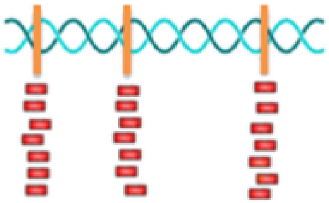
- Sequencing region : whole genome
- Sequencing Depth: >30X
- Covers everything – can identify all kinds of variants including SNPs, INDELs and SV.

Whole exome sequencing

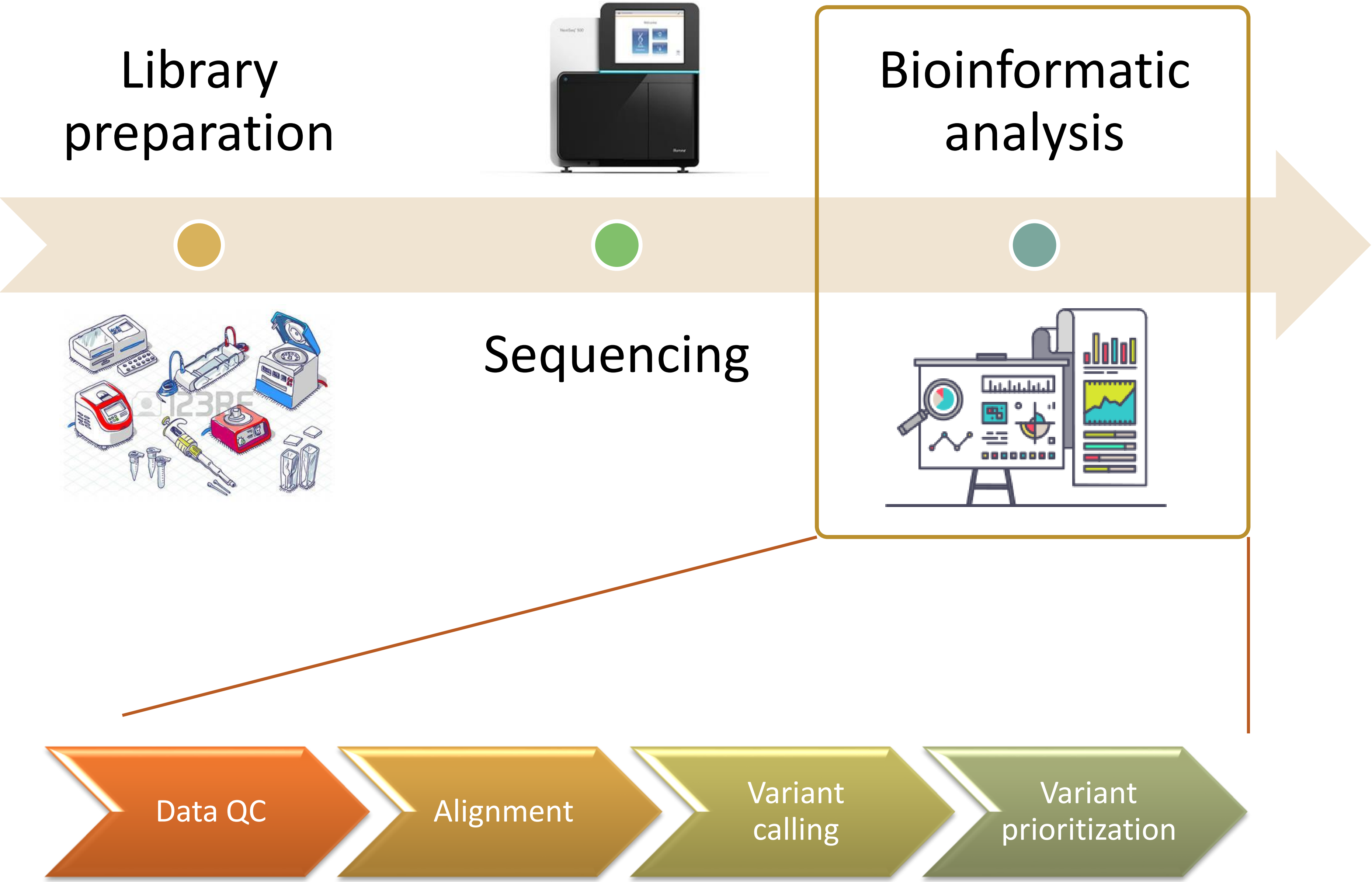


- Sequencing region: whole exome
- Sequencing Depth : >50X ~ 100X
- Identify all kinds of variants including SNPs, INDELs and SV in coding region.
- Cost effective

Targeted sequencing



- Sequencing region: specific regions (could be customized)
- Sequencing Depth : >500X
- Identify all kinds of variants including SNPs, INDELs in specific regions
- Most Cost effective



Library preparation

Sequencing

Bioinformatic analysis

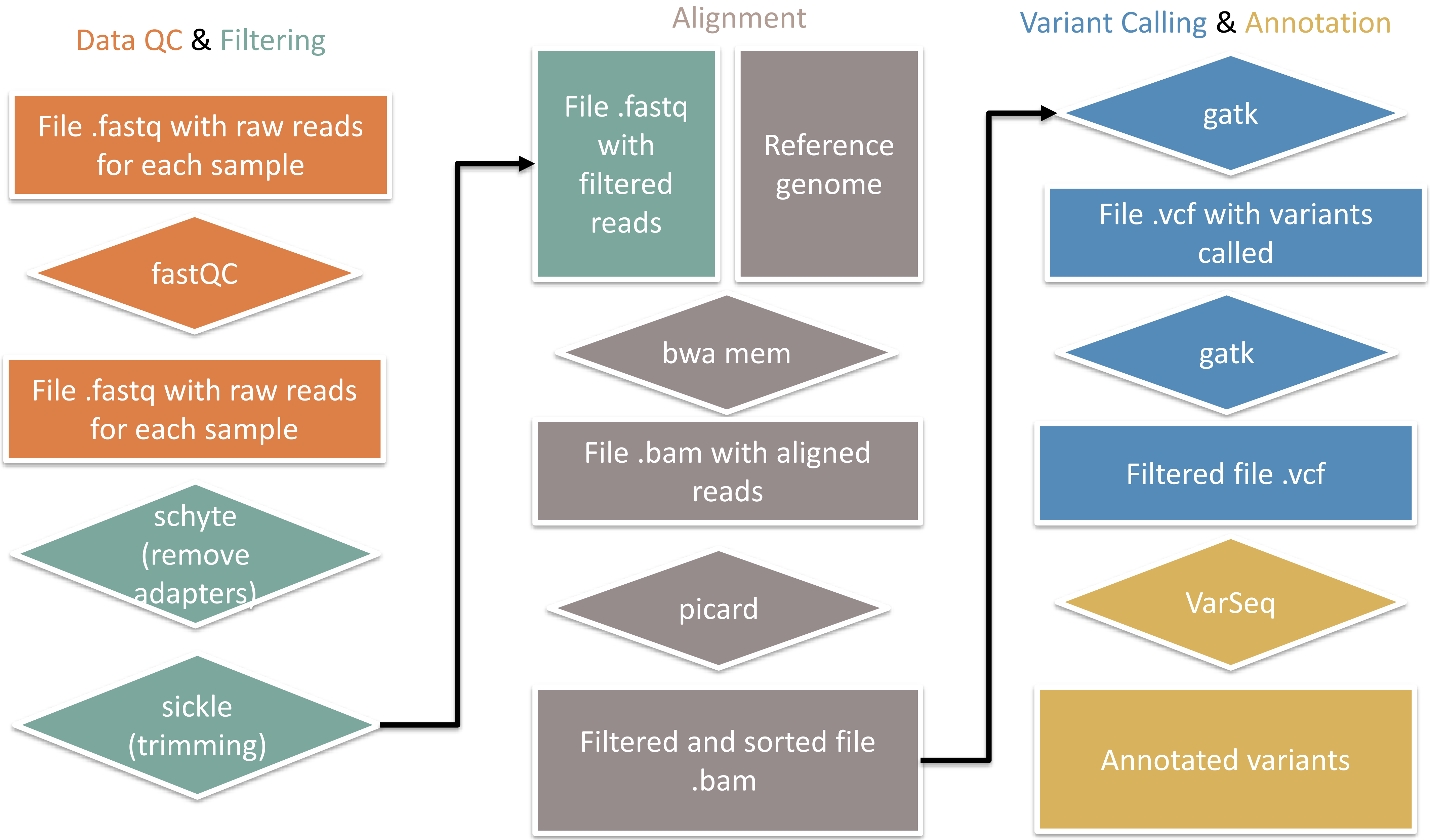
Data QC

Alignment

Variant calling

Variant prioritization

Pipeline



Data QC & Filtering

File .fastq with raw reads for each sample

fastQC

File .fastq with raw reads for each sample

schyte (remove adapters)

sickle (trimming)

Alignment

File .fastq with filtered reads

Reference genome

bwa mem

File .bam with aligned reads

picard

Filtered and sorted file .bam

Variant Calling & Annotation

gatk

File .vcf with variants called

gatk

Filtered file .vcf

VarSeq

Annotated variants

Shell

Windows:

- https://mobaxterm.mobatek.net/download.html

MAC & Linux:

- Open terminal

Connect to the server

1. Enter in the server:
    a. ssh HGSI2020@157.27.26.214
    b. Password: hgsi2020
2. Enter in the folder: cd /attachedvolume/HGSI2020
3. Create your folder: mkdir your_name
4. Enter in the created folder: cd your_name

Work on the server

1. Create a symbolic link of the files in your folder:

   ```
   ln -s ../example/samples/1351S/R*.fastq.gz .
   ```

2. Check you have copied the files: `ls`

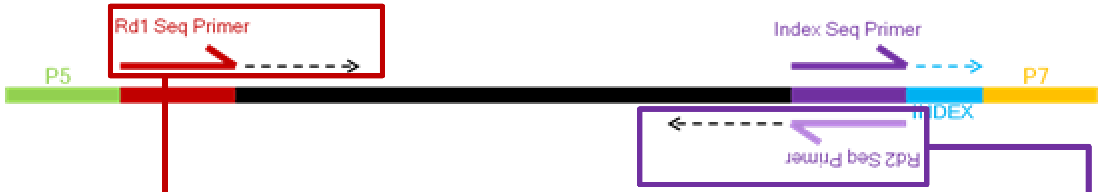
3. Open the file to see what is inside:

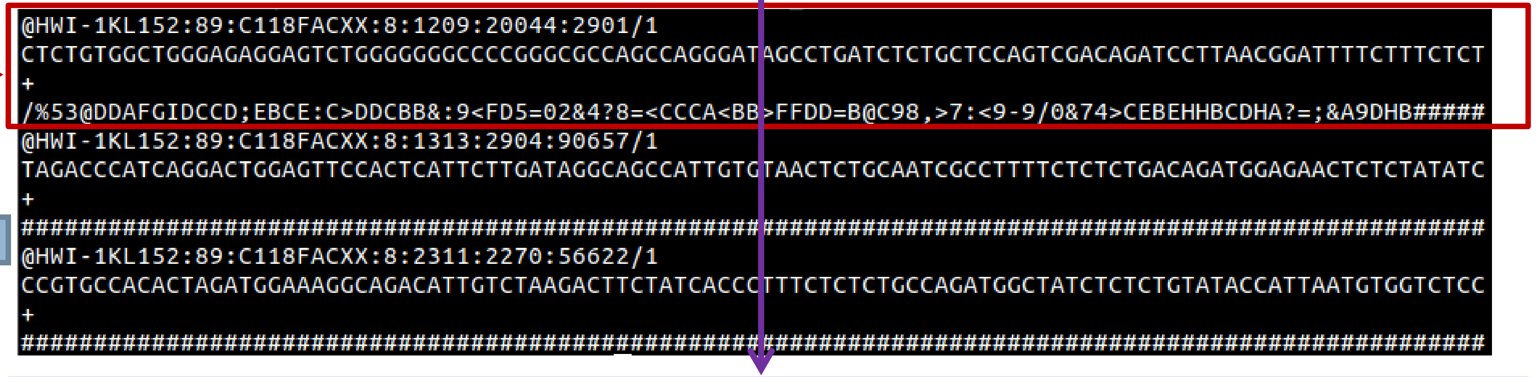
   ```
   less R1.fastq.gz
   ```

4. Close the visualization: `q`

.fq / .fastq file

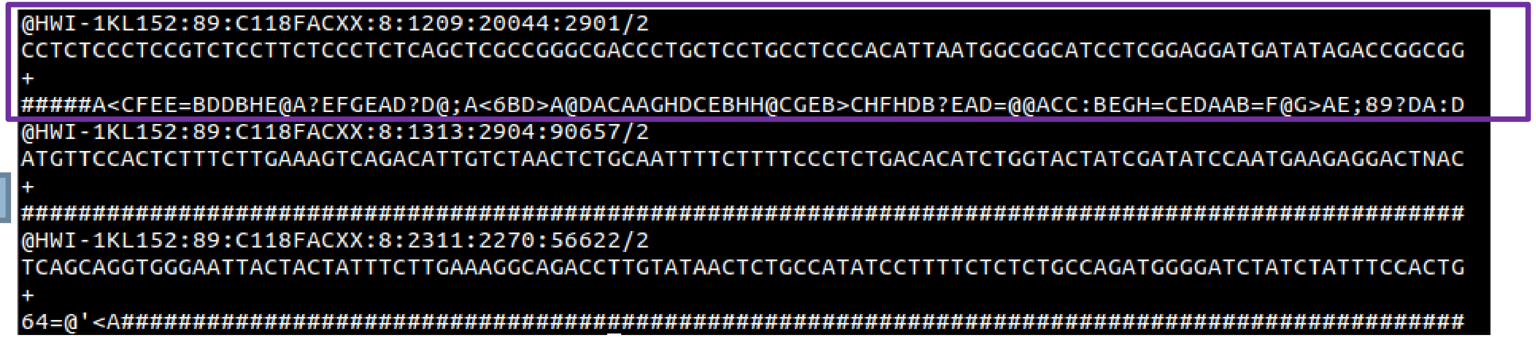
For each sample we obtain 2 fastq files containing all the sequences generated





```
@HWI-1KL152:89:C118FACXX:8:1209:20044:2901/1
CTCTGTGGCTGGGAGAGGAGTCTGGGGGGCCCCGGGCGCCAGCCAGGGGATAGCCTGATCTCTGCTCCAGTCGACAGATCCTTAACGGATTTTCTTTCTCT
+
/%53@DDAFGIDCCD;EBCE:C>DDCBB&:9<FD5=02&4?8=<CCCA<BB>FFDD=B@C98,>7:<9-9/0&74>CEBEHHBCDHA?=;&A9DHB#####
@HWI-1KL152:89:C118FACXX:8:1313:2904:90657/1
TAGACCCATCAGGACTGGAGTTCCACTCATTCTTGATAGGCAGCCATTGTGTAACTCTGCAATCGCCTTTTCTCTCTGACAGATGGAGAACTCTCTATATC
+
#####################################################################################################
@HWI-1KL152:89:C118FACXX:8:2311:2270:56622/1
CCGTGCCACACTAGATGGAAAGGCAGACATTGTCTAAGACTTCTATCACCCTTTCTCTCTGCCAGATGGCTATCTCTCTGTATACCATTAATGTGGTCTCC
+
#####################################################################################################
```

Sequence1.fastq.gz

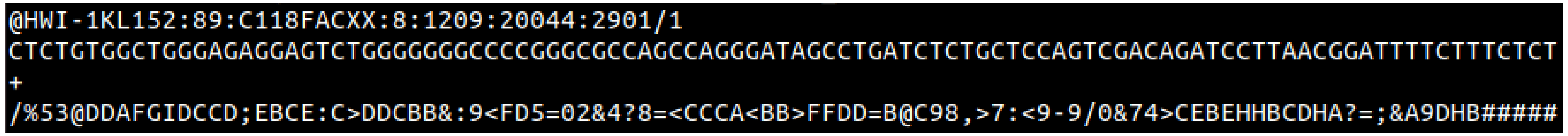


```
@HWI-1KL152:89:C118FACXX:8:1209:20044:2901/2
CCTCTCCCTCCGTCTCCTTCTCCCTCTCAGCTCGCCGGGCGACCCTGCTCCTGCCTCCCACATTAATGGCGGCATCCTCGGAGGATGATATAGACCGGCGG
+
#####A<CFEE=BDDBHE@A?EFGEAD?D@;A<6BD>A@DACAAGHDCEBHH@CGEB>CHFHDB?EAD=@@ACC:BEGH=CEDAAB=F@G>AE;89?DA:D
@HWI-1KL152:89:C118FACXX:8:1313:2904:90657/2
ATGTTCCACTCTTTCTTGAAAGTCAGACATTGTCTAACTCTGCAATTTTCTTTTCCCTCTGACACATCTGGTACTATCGATATCCAATGAAGAGGACTNAC
+
#####################################################################################################
@HWI-1KL152:89:C118FACXX:8:2311:2270:56622/2
TCAGCAGGTGGGAATTACTACTATTTCTTGAAAGGCAGACCTTGTATAACTCTGCCATATCCTTTTCTCTCTGCCAGATGGGGATCTATCTATTTCCACTG
+
64=@'<A##############################################################################################
```

Sequence2.fastq.gz

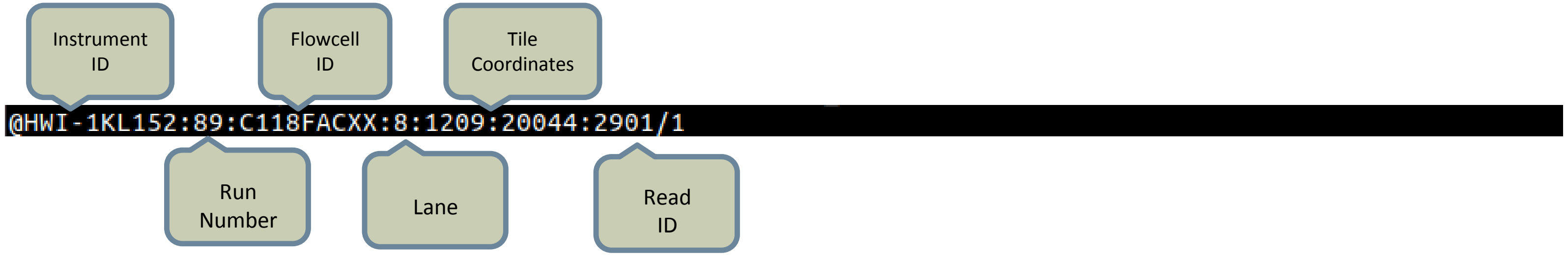
.fq / .fastq file

Each reads is represented by four rows...



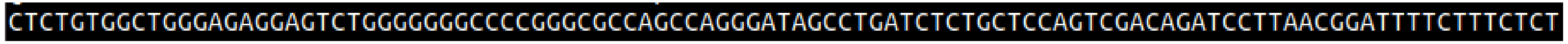
```
@HWI-1KL152:89:C118FACXX:8:1209:20044:2901/1
CTCTGTGGCTGGGAGAGGAGTCTGGGGGGGCCCCGGGCGCCAGCCAGGGATAGCCTGATCTCTGCTCCAGTCGACAGATCCTTAACGGATTTTCTTTCTCT
+
/%53@DDAFGIDCCD;EBCE:C>DDCBB&:9<FD5=02&4?8=<CCCA<BB>FFDD=B@C98,>7:<9-9/0&74>CEBEHHBCDHA?=;&A9DHB#####
```

First row identify the sequence:



```
@HWI-1KL152:89:C118FACXX:8:1209:20044:2901/1
```

Second row contains the sequence:

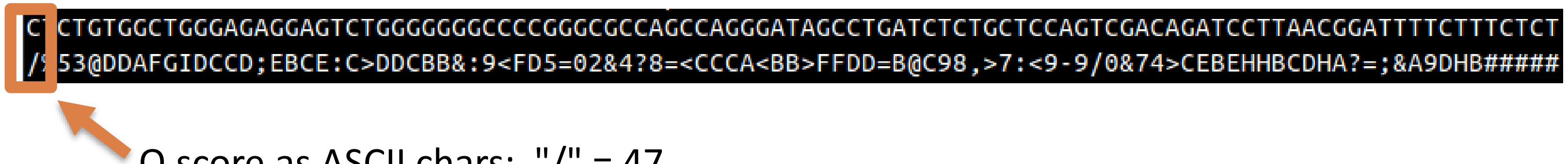


```
CTCTGTGGCTGGGAGAGGAGTCTGGGGGGGCCCCGGGCGCCAGCCAGGGATAGCCTGATCTCTGCTCCAGTCGACAGATCCTTAACGGATTTTCTTTCTCT
```

Thirs row contains a delimiter:

```
+
```

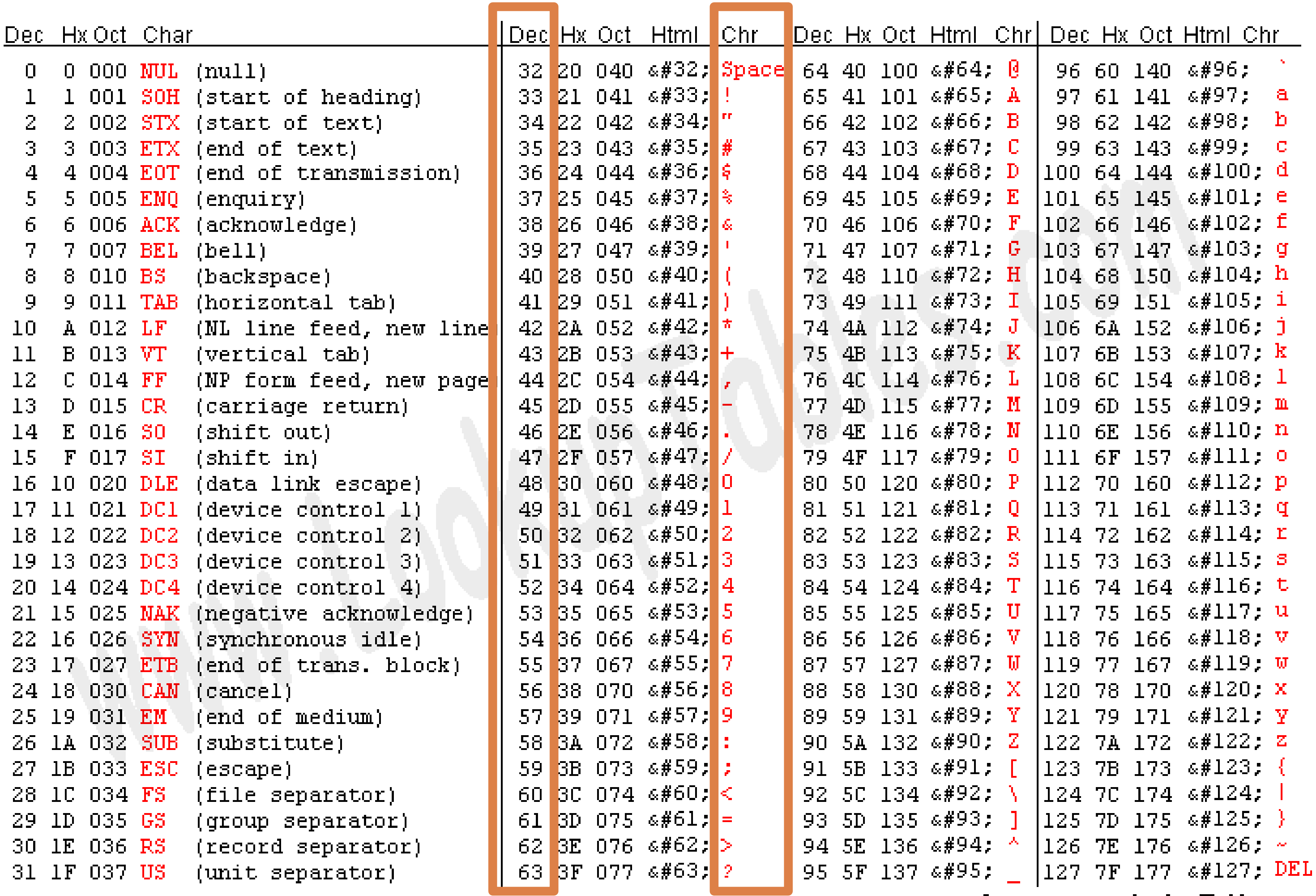
Fourth row indicate the quality of each base:



```
CTCTGTGGCTGGGAGAGGAGTCTGGGGGGGCCCCGGGCGCCAGCCAGGGATAGCCTGATCTCTGCTCCAGTCGACAGATCCTTAACGGATTTTCTTTCTCT
/%53@DDAFGIDCCD;EBCE:C>DDCBB&:9<FD5=02&4?8=<CCCA<BB>FFDD=B@C98,>7:<9-9/0&74>CEBEHHBCDHA?=;&A9DHB#####
```

Q score as ASCII chars: "/" = 47

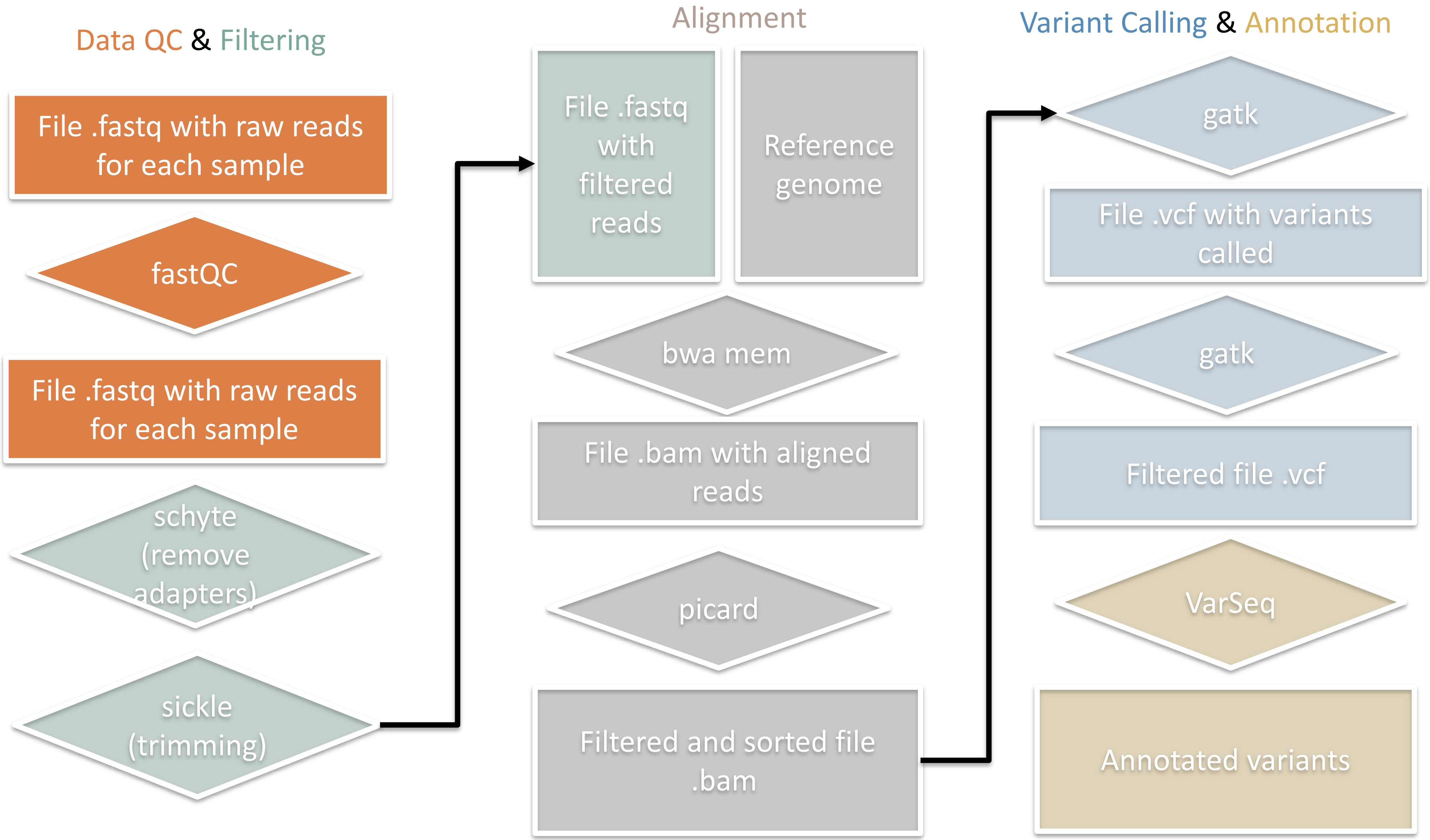
ASCII CODE



Dec	Hx	Oct	Char	Dec	Hx	Oct	Html	Chr	Dec	Hx	Oct	Html	Chr	Dec	Hx	Oct	Html	Chr
0	0	000	NUL (null)	32	20	040	&#32;	Space	64	40	100	&#64;	@	96	60	140	&#96;	`
1	1	001	SOH (start of heading)	33	21	041	&#33;	!	65	41	101	&#65;	A	97	61	141	&#97;	a
2	2	002	STX (start of text)	34	22	042	&#34;	"	66	42	102	&#66;	B	98	62	142	&#98;	b
3	3	003	ETX (end of text)	35	23	043	&#35;	#	67	43	103	&#67;	C	99	63	143	&#99;	c
4	4	004	EOT (end of transmission)	36	24	044	&#36;	$	68	44	104	&#68;	D	100	64	144	&#100;	d
5	5	005	ENQ (enquiry)	37	25	045	&#37;	%	69	45	105	&#69;	E	101	65	145	&#101;	e
6	6	006	ACK (acknowledge)	38	26	046	&#38;	&	70	46	106	&#70;	F	102	66	146	&#102;	f
7	7	007	BEL (bell)	39	27	047	&#39;	'	71	47	107	&#71;	G	103	67	147	&#103;	g
8	8	010	BS (backspace)	40	28	050	&#40;	(	72	48	110	&#72;	H	104	68	150	&#104;	h
9	9	011	TAB (horizontal tab)	41	29	051	&#41;	)	73	49	111	&#73;	I	105	69	151	&#105;	i
10	A	012	LF (NL line feed, new line	42	2A	052	&#42;	*	74	4A	112	&#74;	J	106	6A	152	&#106;	j
11	B	013	VT (vertical tab)	43	2B	053	&#43;	+	75	4B	113	&#75;	K	107	6B	153	&#107;	k
12	C	014	FF (NP form feed, new page	44	2C	054	&#44;	,	76	4C	114	&#76;	L	108	6C	154	&#108;	l
13	D	015	CR (carriage return)	45	2D	055	&#45;	-	77	4D	115	&#77;	M	109	6D	155	&#109;	m
14	E	016	SO (shift out)	46	2E	056	&#46;	.	78	4E	116	&#78;	N	110	6E	156	&#110;	n
15	F	017	SI (shift in)	47	2F	057	&#47;	/	79	4F	117	&#79;	O	111	6F	157	&#111;	o
16	10	020	DLE (data link escape)	48	30	060	&#48;	0	80	50	120	&#80;	P	112	70	160	&#112;	p
17	11	021	DC1 (device control 1)	49	31	061	&#49;	1	81	51	121	&#81;	Q	113	71	161	&#113;	q
18	12	022	DC2 (device control 2)	50	32	062	&#50;	2	82	52	122	&#82;	R	114	72	162	&#114;	r
19	13	023	DC3 (device control 3)	51	33	063	&#51;	3	83	53	123	&#83;	S	115	73	163	&#115;	s
20	14	024	DC4 (device control 4)	52	34	064	&#52;	4	84	54	124	&#84;	T	116	74	164	&#116;	t
21	15	025	NAK (negative acknowledge)	53	35	065	&#53;	5	85	55	125	&#85;	U	117	75	165	&#117;	u
22	16	026	SYN (synchronous idle)	54	36	066	&#54;	6	86	56	126	&#86;	V	118	76	166	&#118;	v
23	17	027	ETB (end of trans. block)	55	37	067	&#55;	7	87	57	127	&#87;	W	119	77	167	&#119;	w
24	18	030	CAN (cancel)	56	38	070	&#56;	8	88	58	130	&#88;	X	120	78	170	&#120;	x
25	19	031	EM (end of medium)	57	39	071	&#57;	9	89	59	131	&#89;	Y	121	79	171	&#121;	y
26	1A	032	SUB (substitute)	58	3A	072	&#58;	:	90	5A	132	&#90;	Z	122	7A	172	&#122;	z
27	1B	033	ESC (escape)	59	3B	073	&#59;	;	91	5B	133	&#91;	[	123	7B	173	&#123;	{
28	1C	034	FS (file separator)	60	3C	074	&#60;	<	92	5C	134	&#92;	\	124	7C	174	&#124;	|
29	1D	035	GS (group separator)	61	3D	075	&#61;	=	93	5D	135	&#93;	]	125	7D	175	&#125;	}
30	1E	036	RS (record separator)	62	3E	076	&#62;	>	94	5E	136	&#94;	^	126	7E	176	&#126;	~
31	1F	037	US (unit separator)	63	3F	077	&#63;	?	95	5F	137	&#95;	_	127	7F	177	&#127;	DEL

Source: www.LookupTables.com

Pipeline


Data QC & Filtering
File .fastq with raw reads for each sample
fastQC
File .fastq with raw reads for each sample
schyte (remove adapters)
sickle (trimming)
Alignment
File .fastq with filtered reads
Reference genome
bwa mem
File .bam with aligned reads
picard
Filtered and sorted file .bam
Variant Calling & Annotation
gatk
File .vcf with variants called
gatk
Filtered file .vcf
VarSeq
Annotated variants

Check quality of your fastq files with fastQC

Fastqc command

1. In your folder, create a folder for fastqc output:

```
mkdir fastqc
```

2. Launch fastQC on both files:

```
fastqc R*.fastq.gz -o fastqc
```

Fastqc command

On the server, we don't have a graphical vision, so..

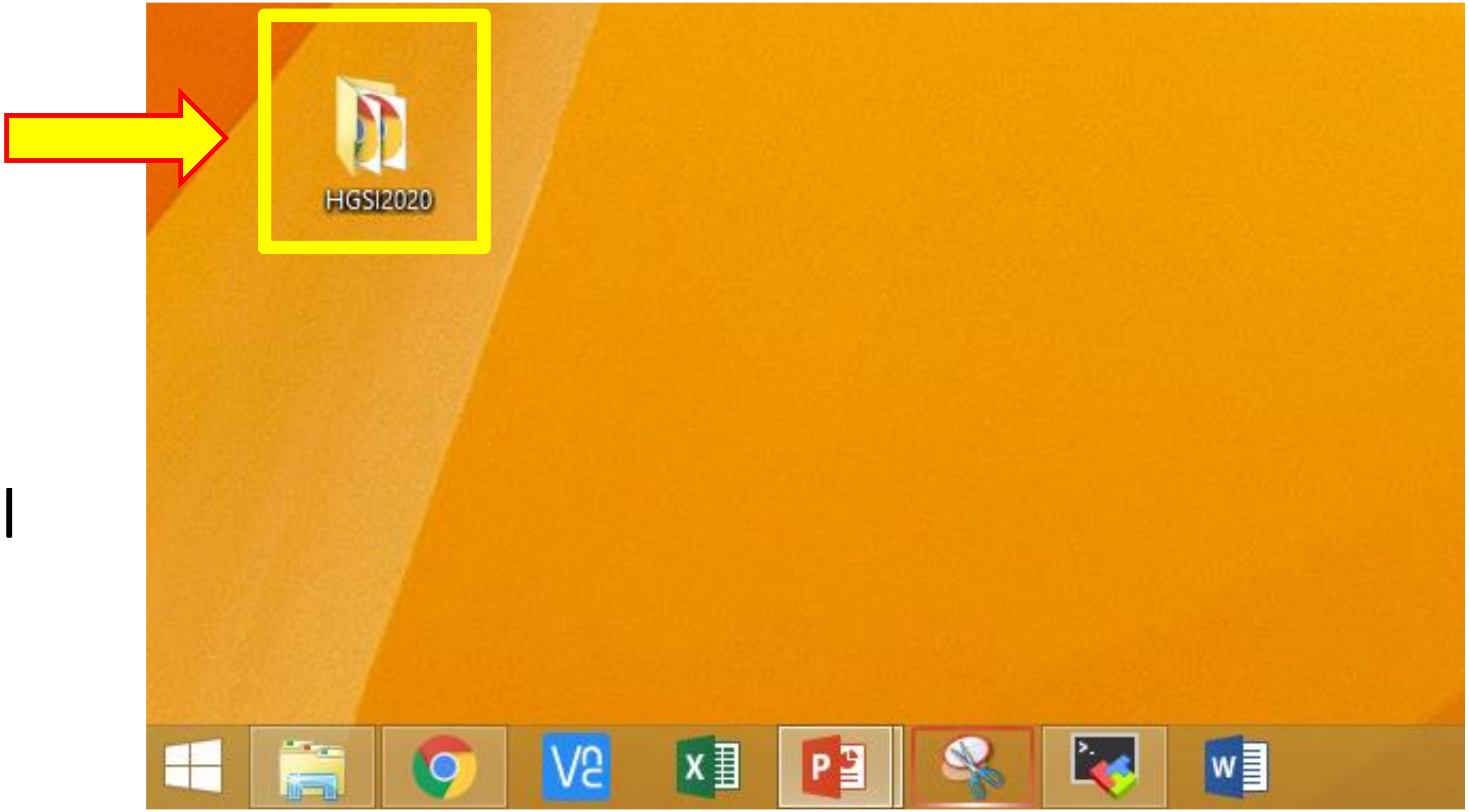
1. Open a new terminal
2. Create a folder on your PC for the course: mkdir Desktop/HGSI2020
3. Enter in the folder: cd Desktop/HGSI2020
4. Download the results here:

   ```
   rsync -auv HGSI2020@157.27.26.214:/attachedvolume/HGSI2020/Denise/fastqc/R*_fastqc.html .
   ```

   Pass: hgsi2020
5. Check you have downloaded: ls
6. Close the shell

Download the files

1. On your desktop, open the file «HGSI2020»

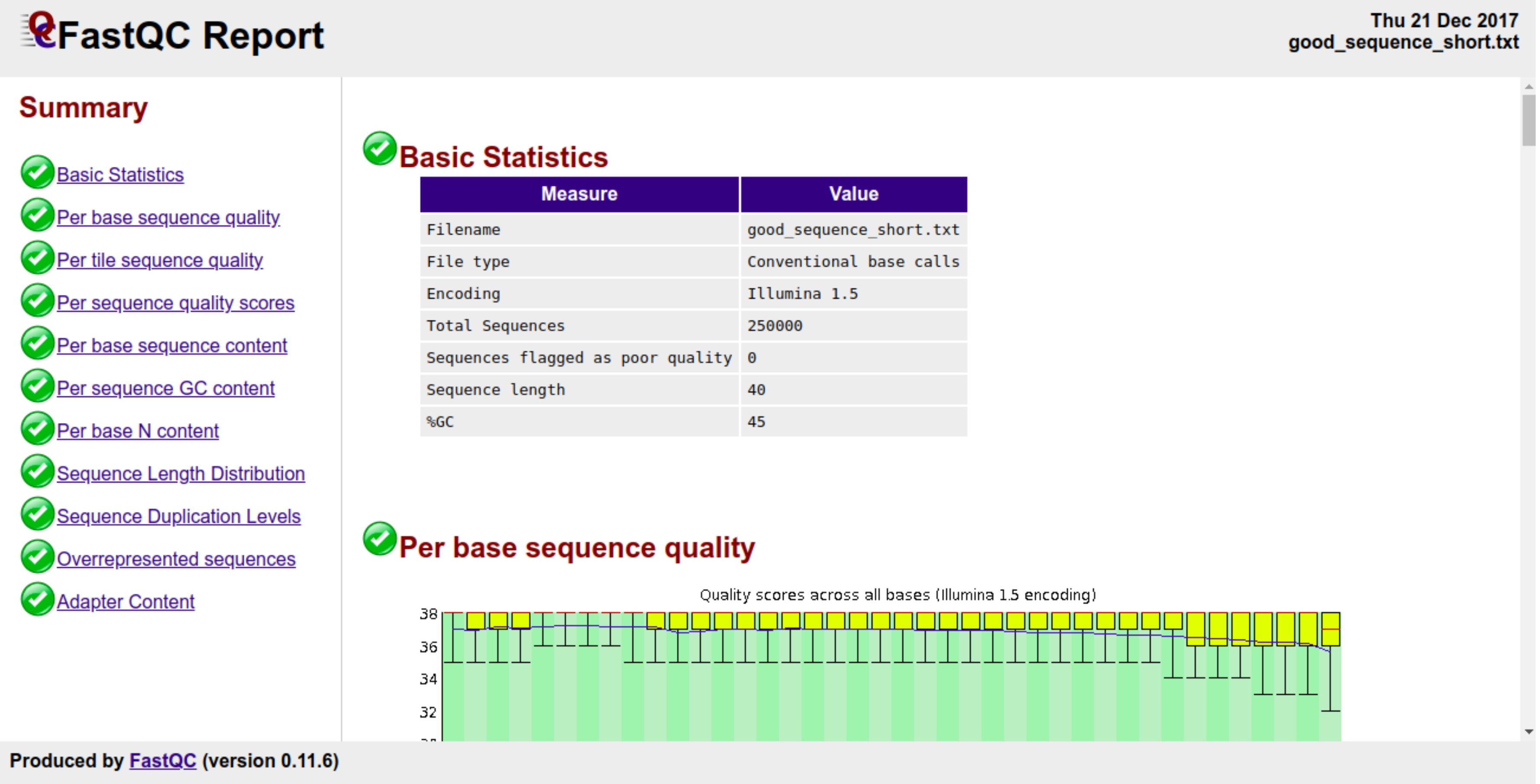


2. Open the file html

FastQC

FastQC software allows to do some quality control checks on raw sequence data coming from high throughput sequencing.

HTML Report



FastQC Report

Thu 21 Dec 2017
good_sequence_short.txt

Summary

- ✓ Basic Statistics
- ✓ Per base sequence quality
- ✓ Per tile sequence quality
- ✓ Per sequence quality scores
- ✓ Per base sequence content
- ✓ Per sequence GC content
- ✓ Per base N content
- ✓ Sequence Length Distribution
- ✓ Sequence Duplication Levels
- ✓ Overrepresented sequences
- ✓ Adapter Content

✓ **Basic Statistics**

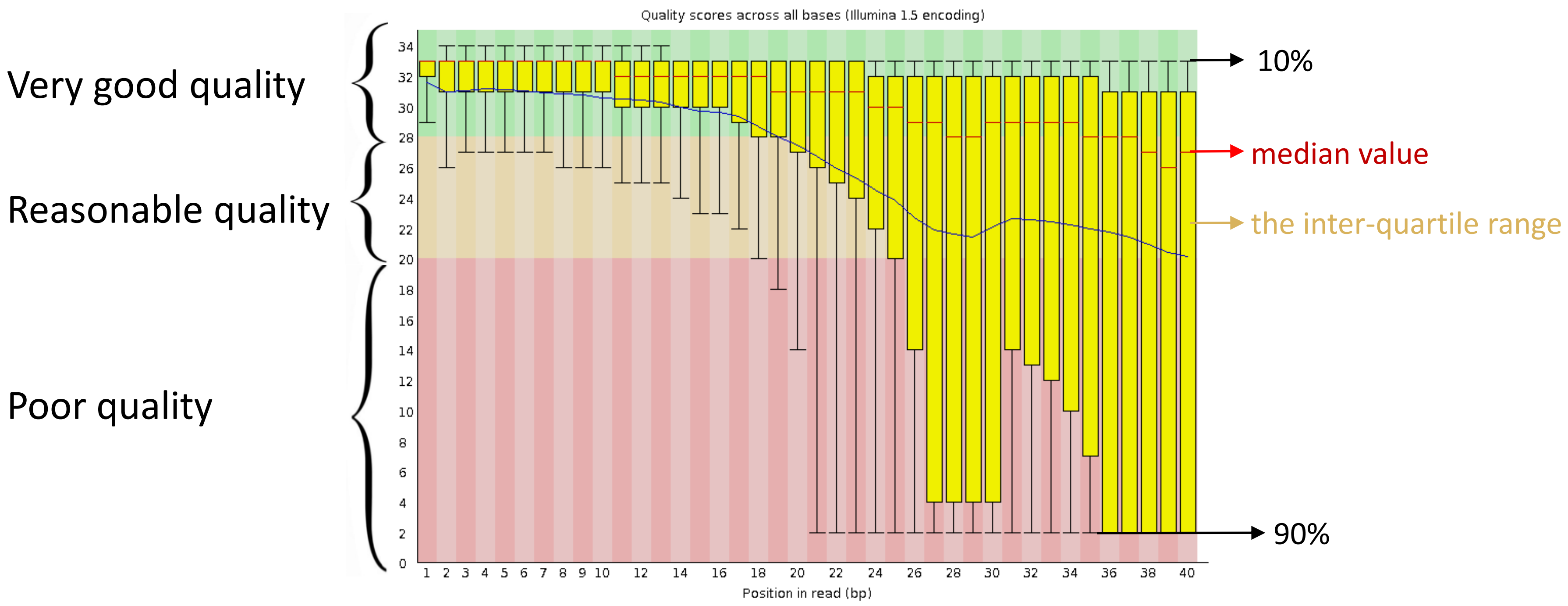
Measure	Value
Filename	good_sequence_short.txt
File type	Conventional base calls
Encoding	Illumina 1.5
Total Sequences	250000
Sequences flagged as poor quality	0
Sequence length	40
%GC	45

✓ **Per base sequence quality**

Quality scores across all bases (Illumina 1.5 encoding)

Produced by FastQC (version 0.11.6)

FastQC – Per base sequence quality



Very good quality

Reasonable quality

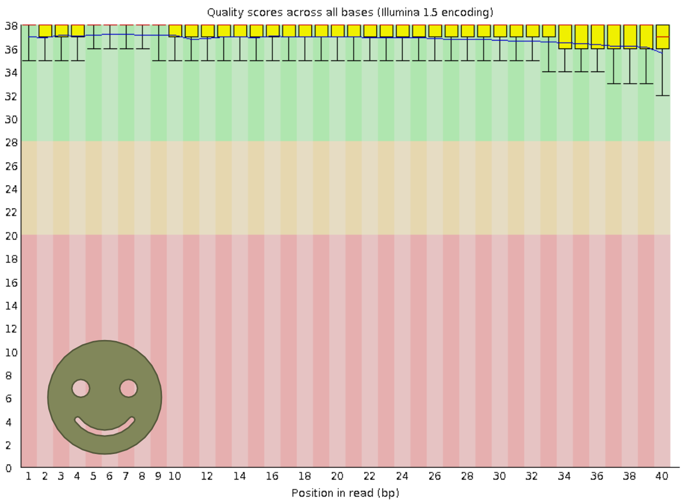
Poor quality

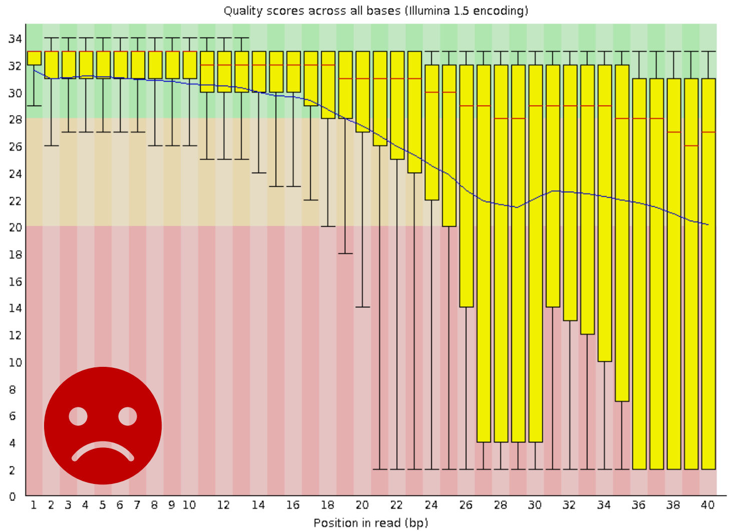
10%

median value

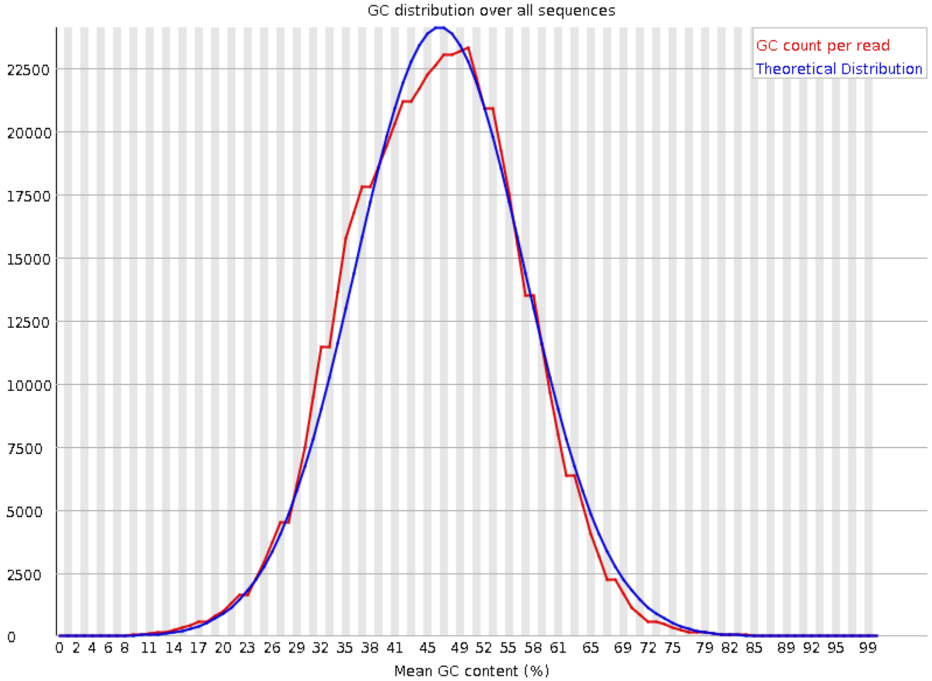
the inter-quartile range

90%

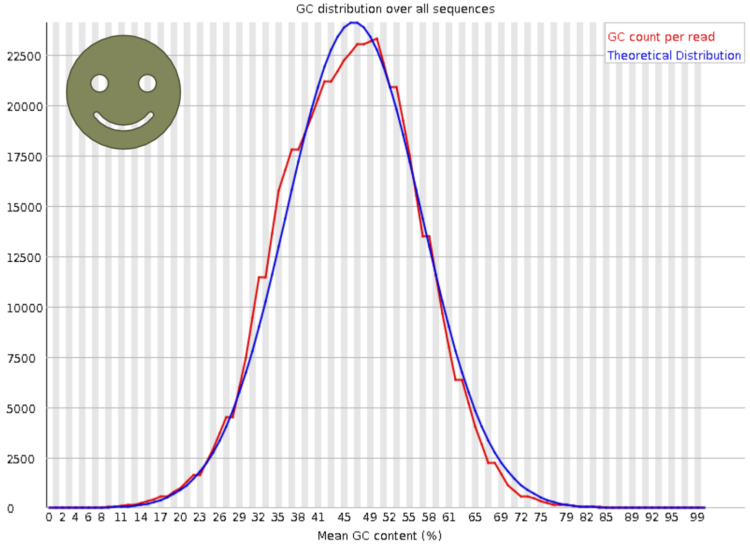


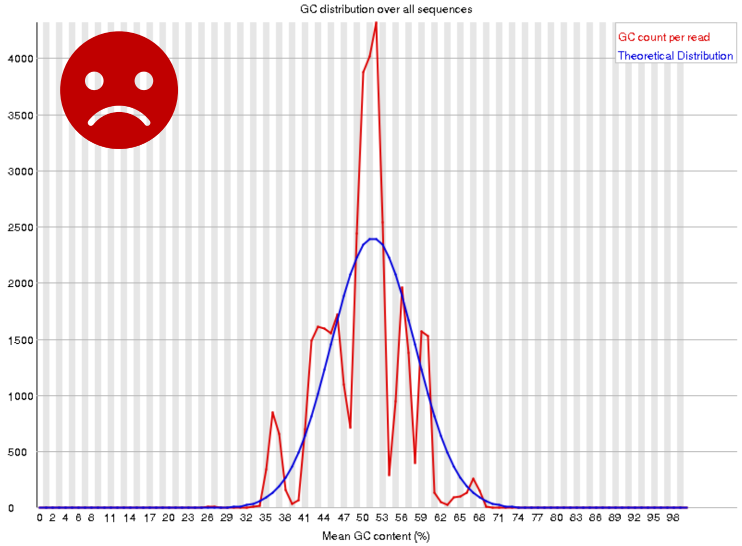


FastQC – Per sequence GC content

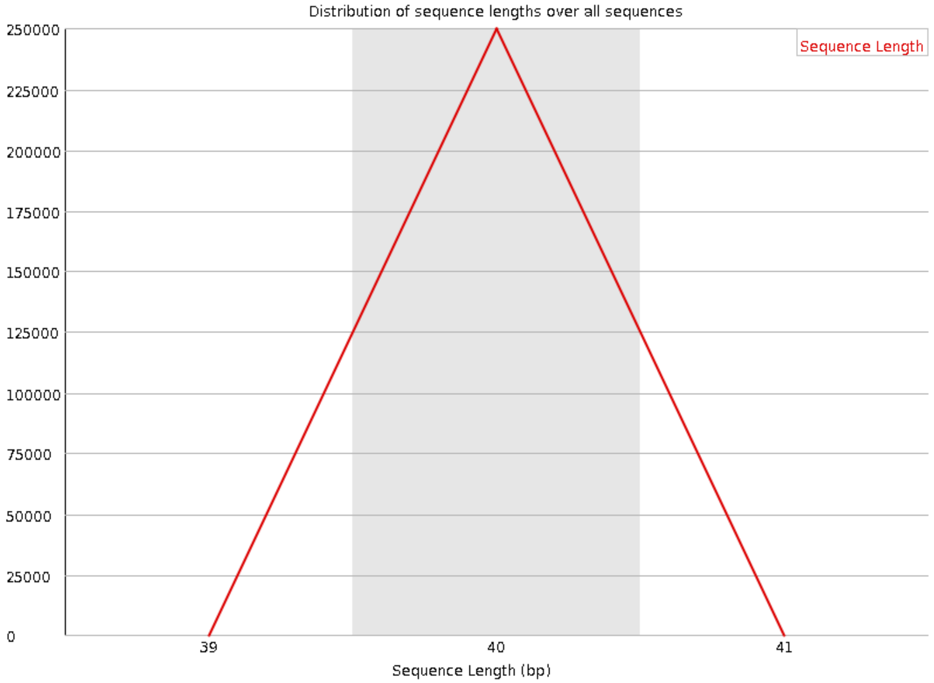


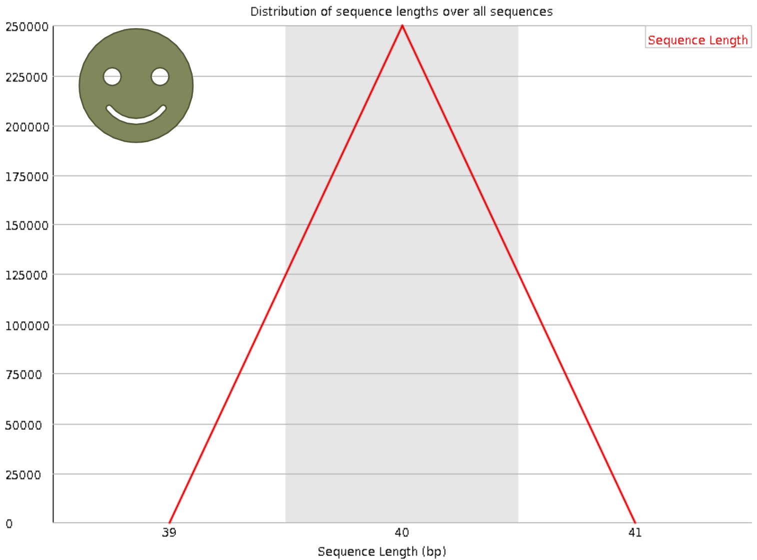
Measure of the GC content across the whole length of each sequence and compares it to a modelled normal distribution of GC content

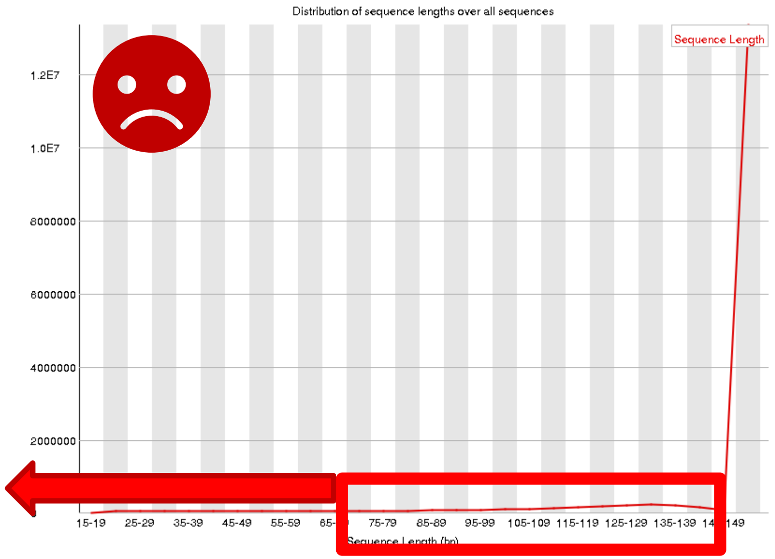




FastQC – Sequence Length Distribution

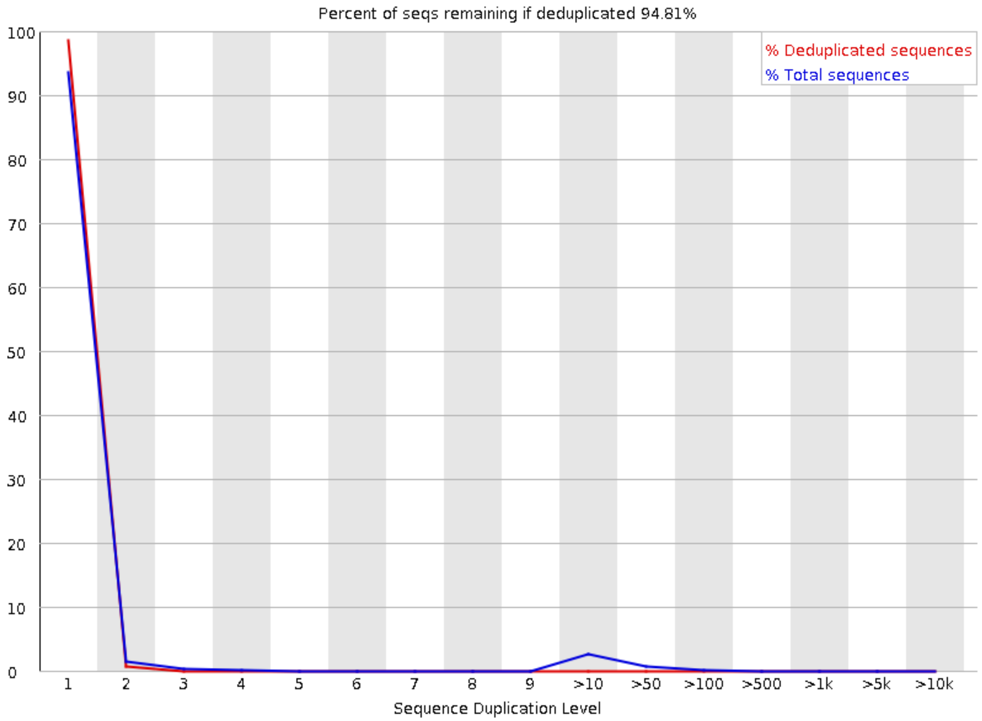


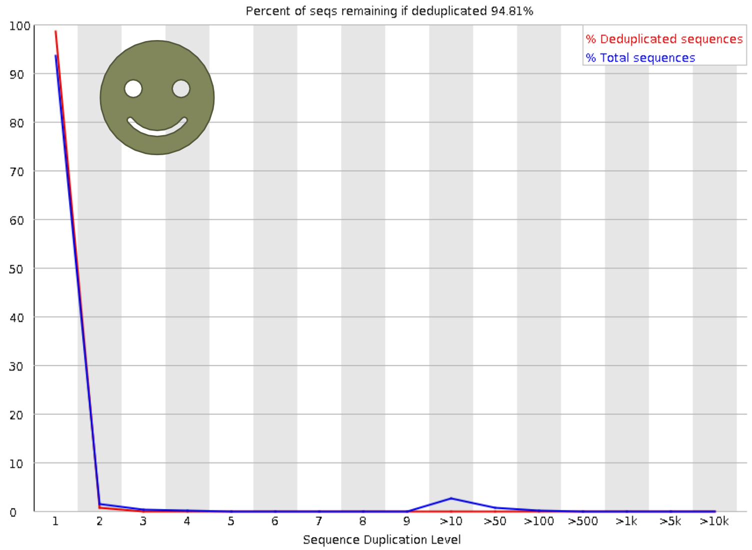




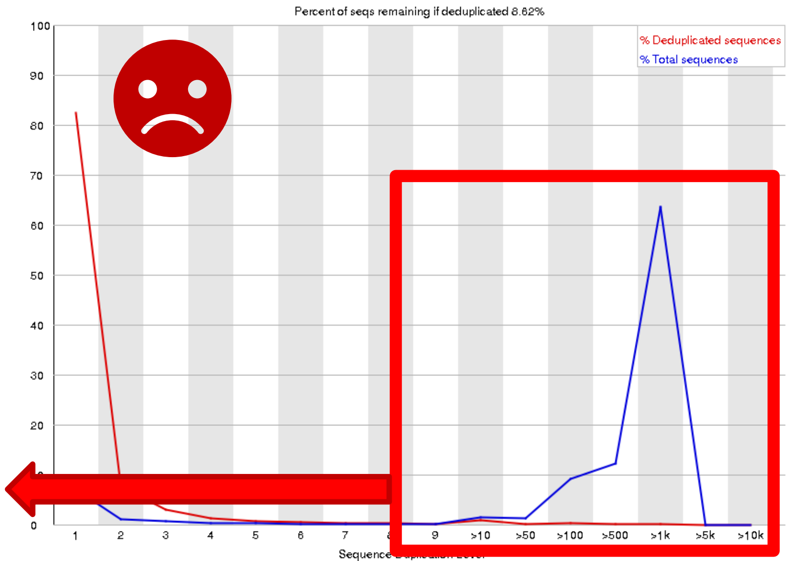
Some sequences have different length or zero length

FastQC – Sequence Duplication Levels

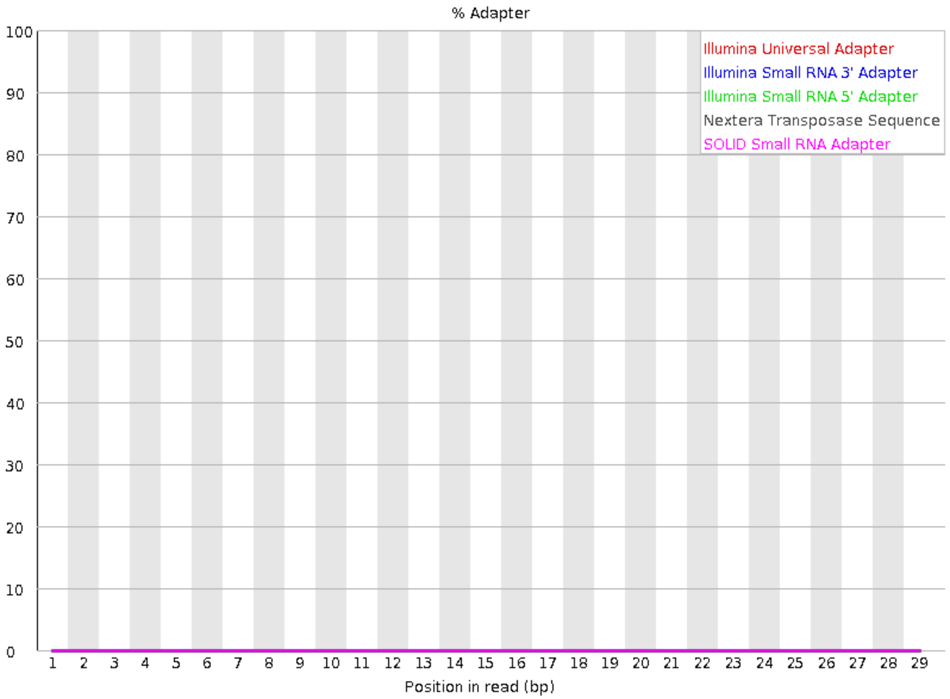


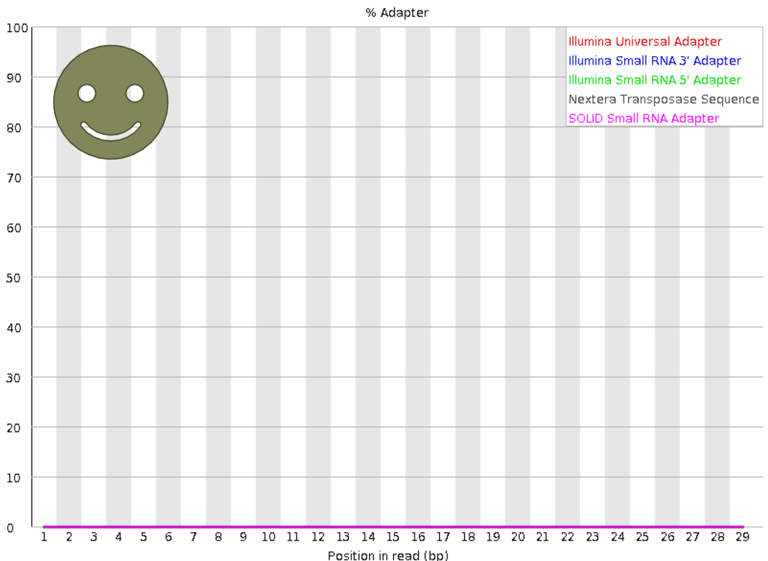


Very large number of sequences with high levels of duplication

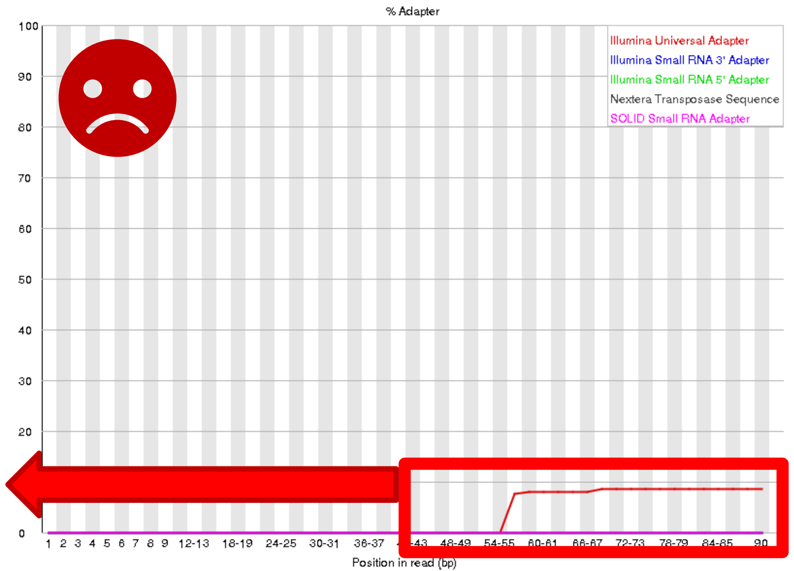


FastQC – Adapter Content

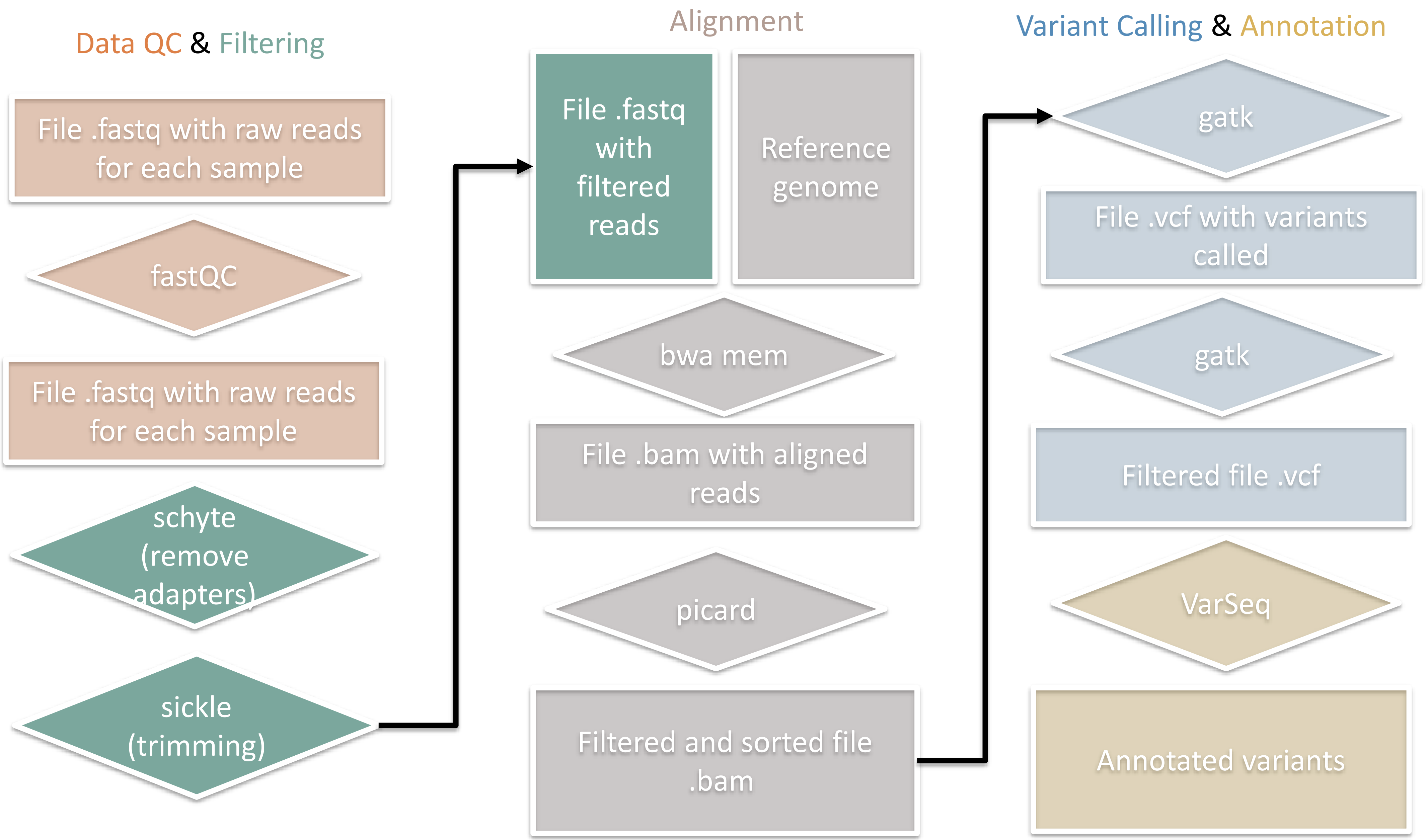




Some sequences contains adapters



Pipeline



Data QC & Filtering

File .fastq with raw reads for each sample

fastQC

File .fastq with raw reads for each sample

schyte (remove adapters)

sickle (trimming)

Alignment

File .fastq with filtered reads

Reference genome

bwa mem

File .bam with aligned reads

picard

Filtered and sorted file .bam

Variant Calling & Annotation

gatk

File .vcf with variants called

gatk

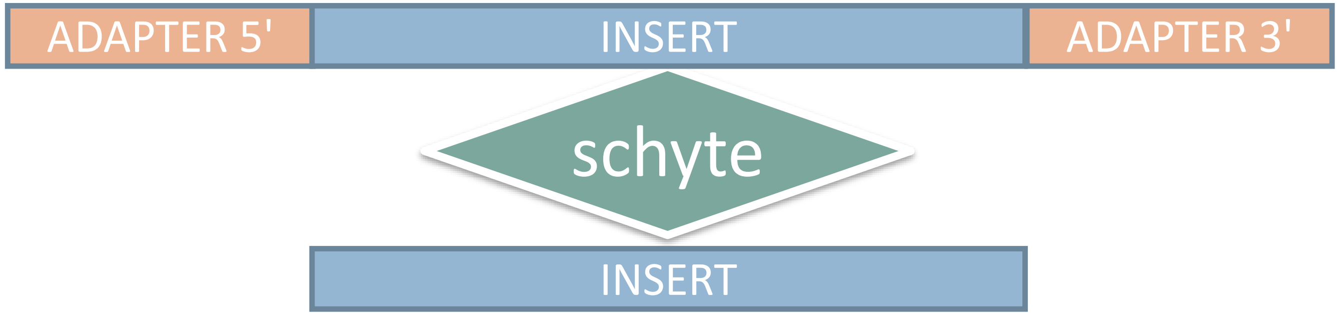
Filtered file .vcf

VarSeq

Annotated variants

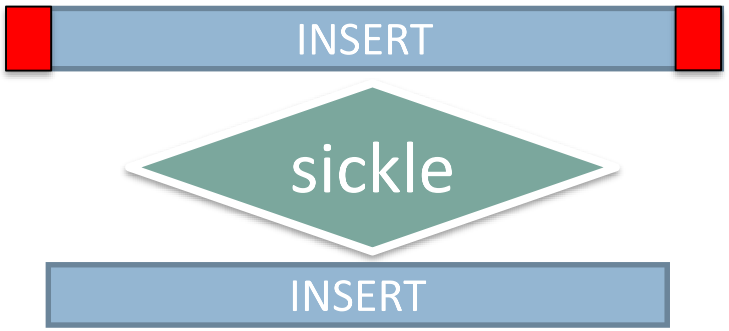
Remove adapters

Input: RAW fastQ read



Reads trimming

Input: RAW fastQ read without adapters

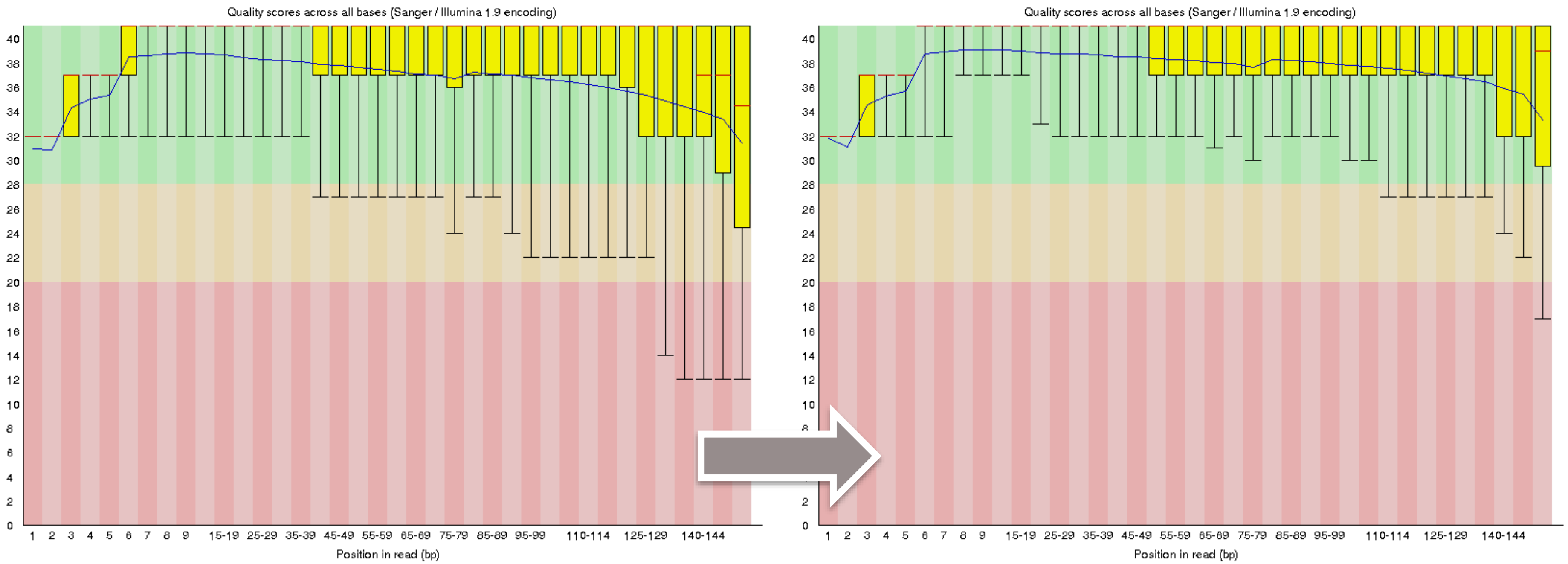


INSERT

sickle

INSERT

🟥 Low quality bases



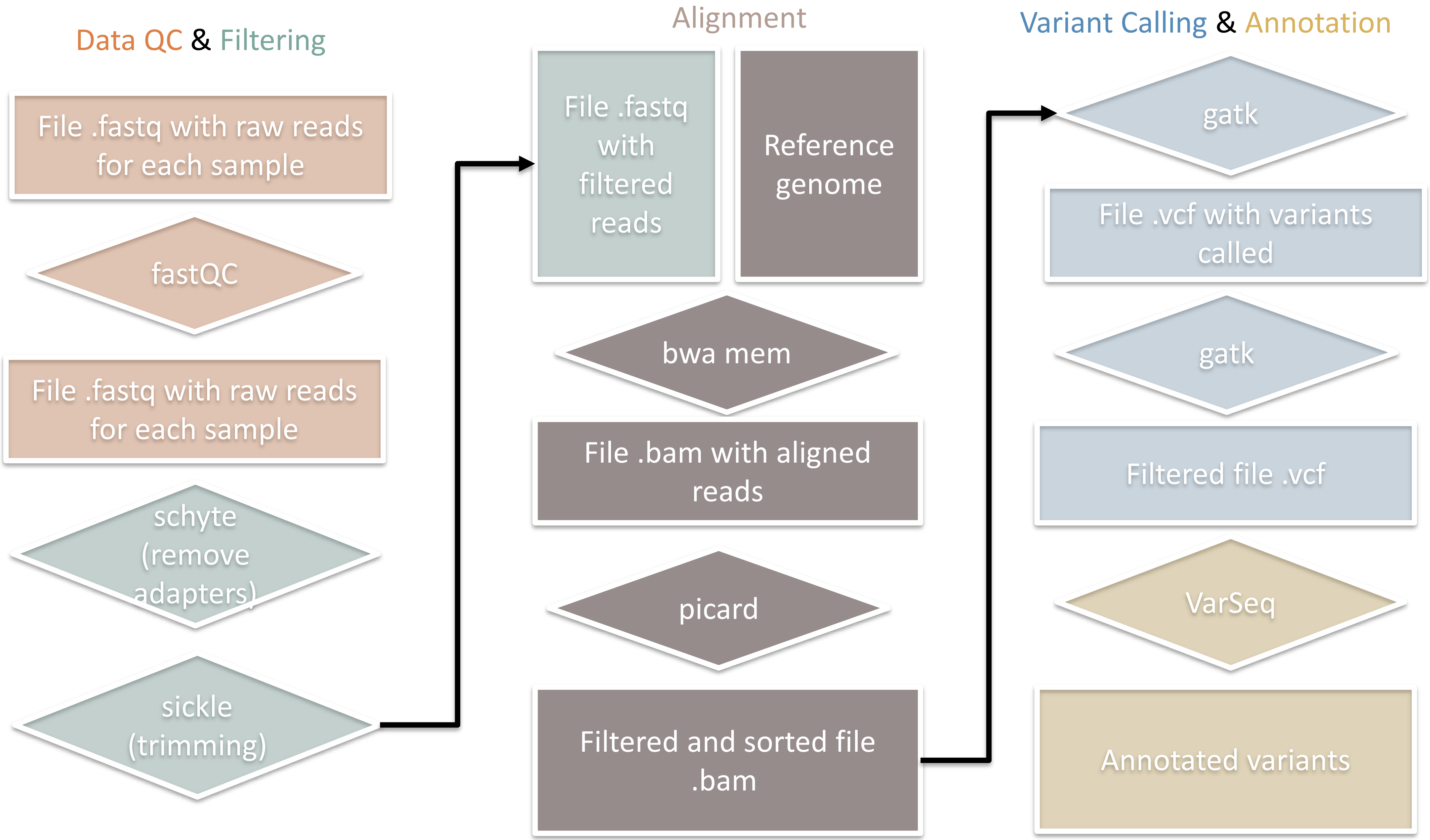
Filter command

Remove adapters from both reads and trimm reads:

```
sickle pe -g -t sanger

-f <( scythe -a ../example/reference/illumina_adapters.fa -q sanger R1.fastq.gz)

-r <( scythe -a ../example/reference/illumina_adapters.fa -q sanger R2.fastq.gz )

-o trimmed1.fastq.gz -p trimmed2.fastq.gz -s /dev/null
```

Execute the script: `bash /attachedvolume/HGSI2020/example/scripts/step0.trimming.sh`

Pipeline



Data QC & Filtering

File .fastq with raw reads for each sample

fastQC

File .fastq with raw reads for each sample

schyte (remove adapters)

sickle (trimming)

Alignment

File .fastq with filtered reads

Reference genome

bwa mem

File .bam with aligned reads

picard

Filtered and sorted file .bam

Variant Calling & Annotation

gatk

File .vcf with variants called

gatk

Filtered file .vcf

VarSeq

Annotated variants

Alignment

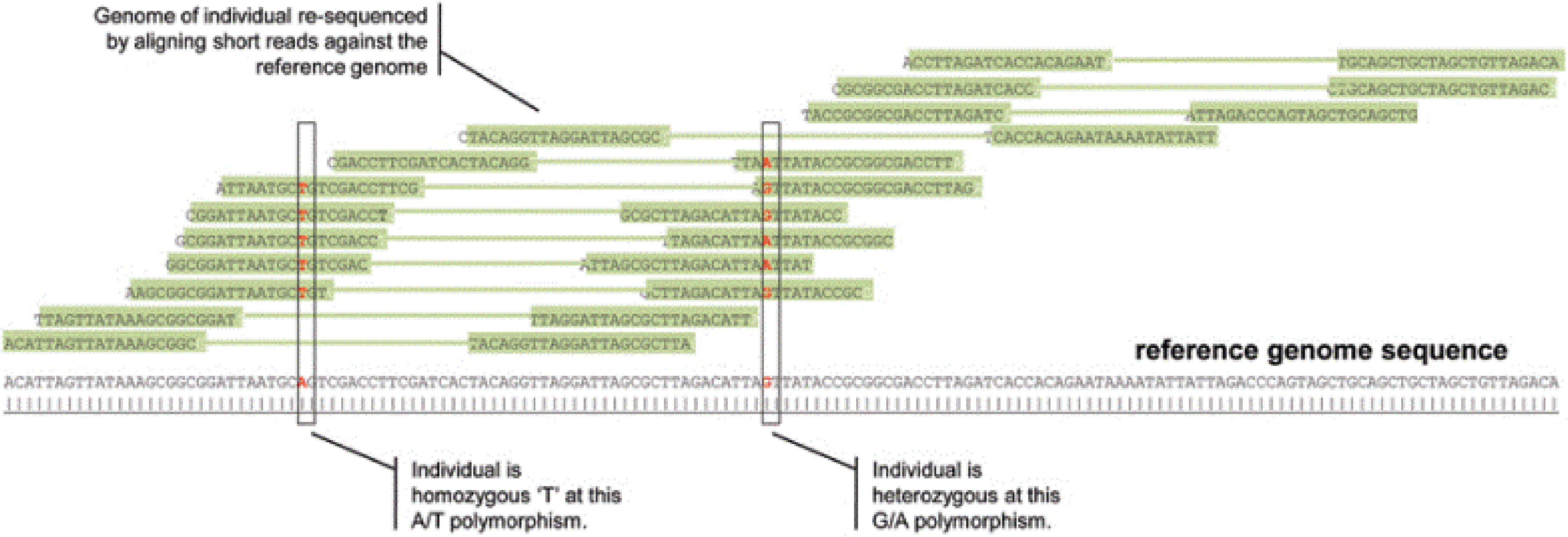
Many of the next-generation sequencing projects begin with a known, or so-called ‘reference’, genome. In this case, to make sense of the reads, their positions within the reference sequence must be determined. This process is known as aligning or ‘mapping’ the read to the reference.

Computationally difficult

- Short Reads
- Lots of repeats
- Presence of mismatch

Different algorithm solution:

- Bowtie
- BWA
- ISAAC Aligner


Genome of individual re-sequenced by aligning short reads against the reference genome
reference genome sequence
ACATTAGTTATAAAGCGGCGGATTAATGCAGTCGACCTTCGATCACTACAGGTTAGGATTAGCGCTTAGACATTAGTTATACCGCGGCGACCTTAGATCACCACAGAATAAAATATTATTAGACCCAGTAGCTGCAGCTGCTAGCTGTTAGACA
Individual is homozygous ‘T’ at this A/T polymorphism.
Individual is heterozygous at this G/A polymorphism.

Alignment command

- Align your reads to the reference genome (chr6 hg38):

```
/attachedvolume/HGSI2020/example/bin/bwa-0.7.12 mem
/attachedvolume/HGSI2020/example/reference/chr6.hg38.fa
trimmed1.fastq.gz trimmed2.fastq.gz > sample.sam
```

- Turn your file sam into file bam:

```
samtools view -bT /attachedvolume/HGSI2020/example/reference/chr6.hg38.fa -o
sample.bam sample.sam
```

Alignment command

- Sort your file:

```
samtools sort sample.bam -o sample.sorted.bam
```

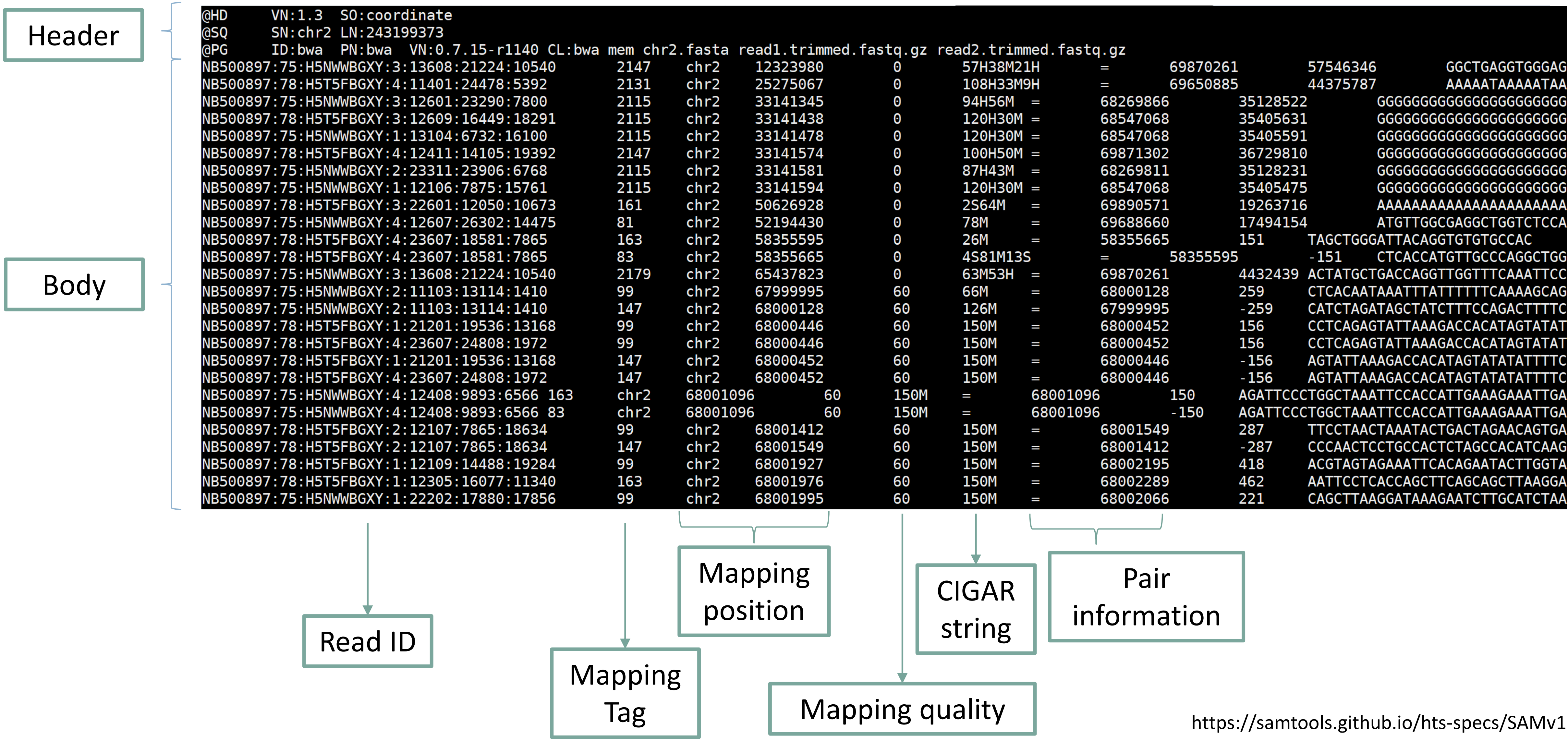
- Create index for your bam file:

```
samtools index sample.sorted.bam
```

- Open the file:

```
samtools view sample.sorted.bam | less -S
```

Alignment output – BAM file



```
@HD     VN:1.3  SO:coordinate
@SQ     SN:chr2 LN:243199373
@PG     ID:bwa  PN:bwa  VN:0.7.15-r1140 CL:bwa mem chr2.fasta read1.trimmed.fastq.gz read2.trimmed.fastq.gz
NB500897:75:H5NWWBGXY:3:13608:21224:10540       2147    chr2    12323980        0       57H38M21H       =       69870261        57546346        GGCTGAGGTGGGAG
NB500897:78:H5T5FBGXY:4:11401:24478:5392        2131    chr2    25275067        0       108H33M9H       =       69650885        44375787        AAAAATAAAAATAA
NB500897:75:H5NWWBGXY:3:12601:23290:7800        2115    chr2    33141345        0       94H56M  =       68269866        35128522        GGGGGGGGGGGGGGGGGGGGG
NB500897:78:H5T5FBGXY:3:12609:16449:18291       2115    chr2    33141438        0       120H30M =       68547068        35405631        GGGGGGGGGGGGGGGGGGGGG
NB500897:75:H5NWWBGXY:1:13104:6732:16100        2115    chr2    33141478        0       120H30M =       68547068        35405591        GGGGGGGGGGGGGGGGGGGGG
NB500897:78:H5T5FBGXY:4:12411:14105:19392       2147    chr2    33141574        0       100H50M =       69871302        36729810        GGGGGGGGGGGGGGGGGGGGG
NB500897:75:H5NWWBGXY:2:23311:23906:6768        2115    chr2    33141581        0       87H43M  =       68269811        35128231        GGGGGGGGGGGGGGGGGGGGG
NB500897:75:H5NWWBGXY:1:12106:7875:15761        2115    chr2    33141594        0       120H30M =       68547068        35405475        GGGGGGGGGGGGGGGGGGGGG
NB500897:78:H5T5FBGXY:3:22601:12050:10673       161     chr2    50626928        0       2S64M   =       69890571        19263716        AAAAAAAAAAAAAAAAAAAAA
NB500897:75:H5NWWBGXY:4:12607:26302:14475       81      chr2    52194430        0       78M     =       69688660        17494154        ATGTTGGCGAGGCTGGTCTCCA
NB500897:78:H5T5FBGXY:4:23607:18581:7865        163     chr2    58355595        0       26M     =       58355665        151     TAGCTGGGATTACAGGTGTGTGCCAC
NB500897:78:H5T5FBGXY:4:23607:18581:7865        83      chr2    58355665        0       4S81M13S        =       58355595        -151    CTCACCATGTTGCCCAGGCTGG
NB500897:75:H5NWWBGXY:3:13608:21224:10540       2179    chr2    65437823        0       63M53H  =       69870261        4432439 ACTATGCTGACCAGGTTGGTTTCAAATTCC
NB500897:75:H5NWWBGXY:2:11103:13114:1410        99      chr2    67999995        60      66M     =       68000128        259     CTCACAATAAATTTATTTTTTCAAAAGCAG
NB500897:75:H5NWWBGXY:2:11103:13114:1410        147     chr2    68000128        60      126M    =       67999995        -259    CATCTAGATAGCTATCTTTCCAGACTTTTC
NB500897:78:H5T5FBGXY:1:21201:19536:13168       99      chr2    68000446        60      150M    =       68000452        156     CCTCAGAGTATTAAAGACCACATAGTATAT
NB500897:78:H5T5FBGXY:4:23607:24808:1972        99      chr2    68000446        60      150M    =       68000452        156     CCTCAGAGTATTAAAGACCACATAGTATAT
NB500897:78:H5T5FBGXY:1:21201:19536:13168       147     chr2    68000452        60      150M    =       68000446        -156    AGTATTAAAGACCACATAGTATATATTTTC
NB500897:78:H5T5FBGXY:4:23607:24808:1972        147     chr2    68000452        60      150M    =       68000446        -156    AGTATTAAAGACCACATAGTATATATTTTC
NB500897:75:H5NWWBGXY:4:12408:9893:6566 163     chr2    68001096        60      150M    =       68001096        150     AGATTCCCTGGCTAAATTCCACCATTGAAAGAAATTGA
NB500897:75:H5NWWBGXY:4:12408:9893:6566 83      chr2    68001096        60      150M    =       68001096        -150    AGATTCCCTGGCTAAATTCCACCATTGAAAGAAATTGA
NB500897:78:H5T5FBGXY:2:12107:7865:18634        99      chr2    68001412        60      150M    =       68001549        287     TTCCTAACTAAATACTGACTAGAACAGTGA
NB500897:78:H5T5FBGXY:2:12107:7865:18634        147     chr2    68001549        60      150M    =       68001412        -287    CCCAACTCCTGCCACTCTAGCCACATCAAG
NB500897:78:H5T5FBGXY:1:12109:14488:19284       99      chr2    68001927        60      150M    =       68002195        418     ACGTAGTAGAAATTCACAGAATACTTGGTA
NB500897:78:H5T5FBGXY:1:12305:16077:11340       163     chr2    68001976        60      150M    =       68002289        462     AATTCCTCACCAGCTTCAGCAGCTTAAGGA
NB500897:75:H5NWWBGXY:1:22202:17880:17856       99      chr2    68001995        60      150M    =       68002066        221     CAGCTTAAGGATAAAGAATCTTGCATCTAA
```

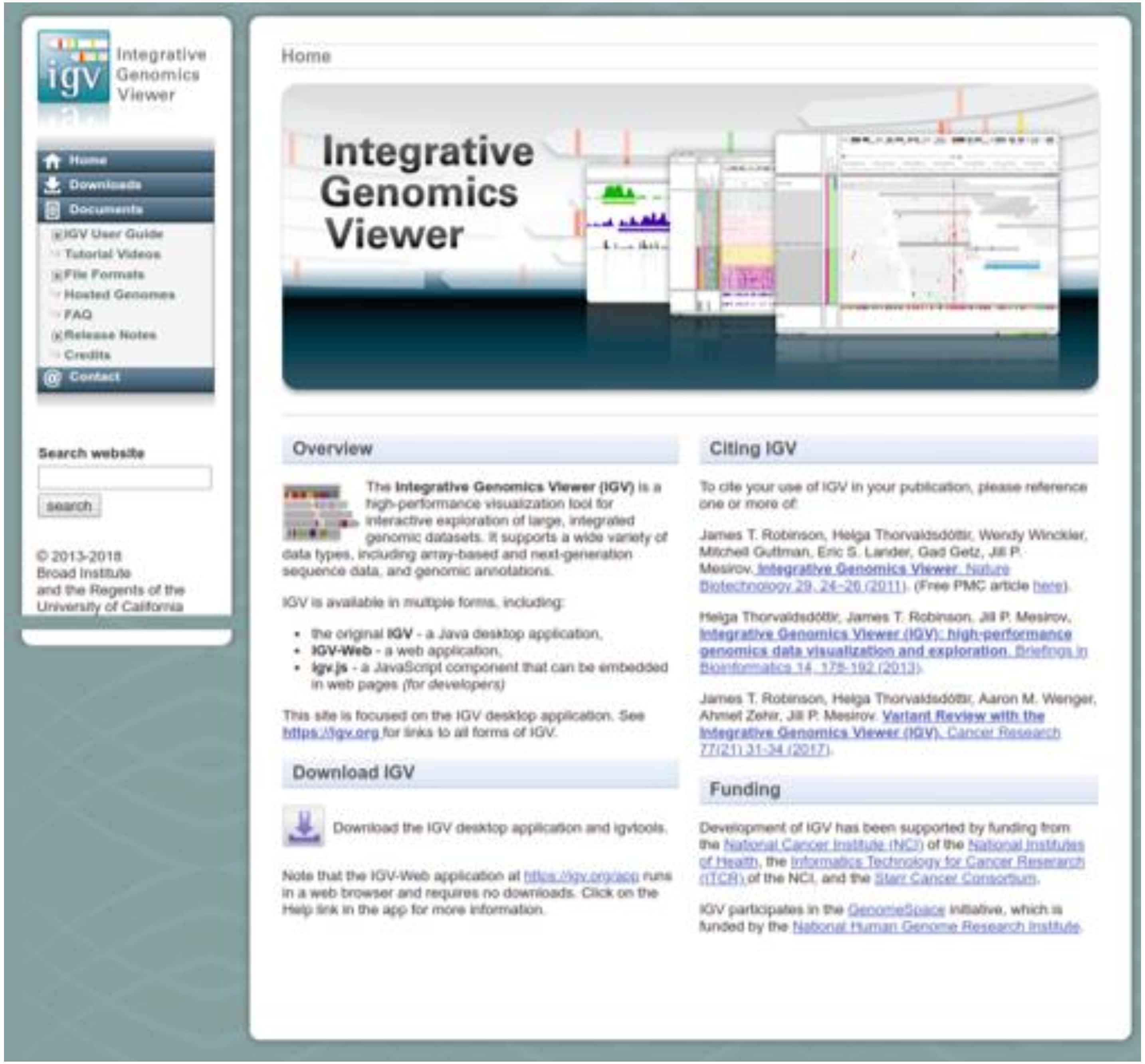
https://samtools.github.io/hts-specs/SAMv1.pdf

Cigar

Op	Description
M	alignment match (can be a sequence match or mismatch)
I	insertion to the reference
D	deletion from the reference
N	skipped region from the reference
S	soft clipping (clipped sequences present in SEQ)
H	hard clipping (clipped sequences NOT present in SEQ)
P	padding (silent deletion from padded reference)
=	sequence match
X	sequence mismatch

IGV

https://software.broadinstitute.org/software/igv/



© 2013-2018
Broad Institute
and the Regents of the
University of California

Overview

The **Integrative Genomics Viewer (IGV)** is a high-performance visualization tool for interactive exploration of large, integrated genomic datasets. It supports a wide variety of data types, including array-based and next-generation sequence data, and genomic annotations.

IGV is available in multiple forms, including:

- the original **IGV** - a Java desktop application,
- **IGV-Web** - a web application,
- **igv.js** - a JavaScript component that can be embedded in web pages *(for developers)*

This site is focused on the IGV desktop application. See https://igv.org for links to all forms of IGV.

Download IGV

Download the IGV desktop application and igvtools.

Note that the IGV-Web application at https://igv.org/app runs in a web browser and requires no downloads. Click on the Help link in the app for more information.

Citing IGV

To cite your use of IGV in your publication, please reference one or more of:

James T. Robinson, Helga Thorvaldsdóttir, Wendy Winckler, Mitchell Guttman, Eric S. Lander, Gad Getz, Jill P. Mesirov. **Integrative Genomics Viewer.** Nature Biotechnology 29, 24–26 (2011). (Free PMC article here).

Helga Thorvaldsdóttir, James T. Robinson, Jill P. Mesirov. **Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration.** Briefings in Bioinformatics 14, 178-192 (2013).

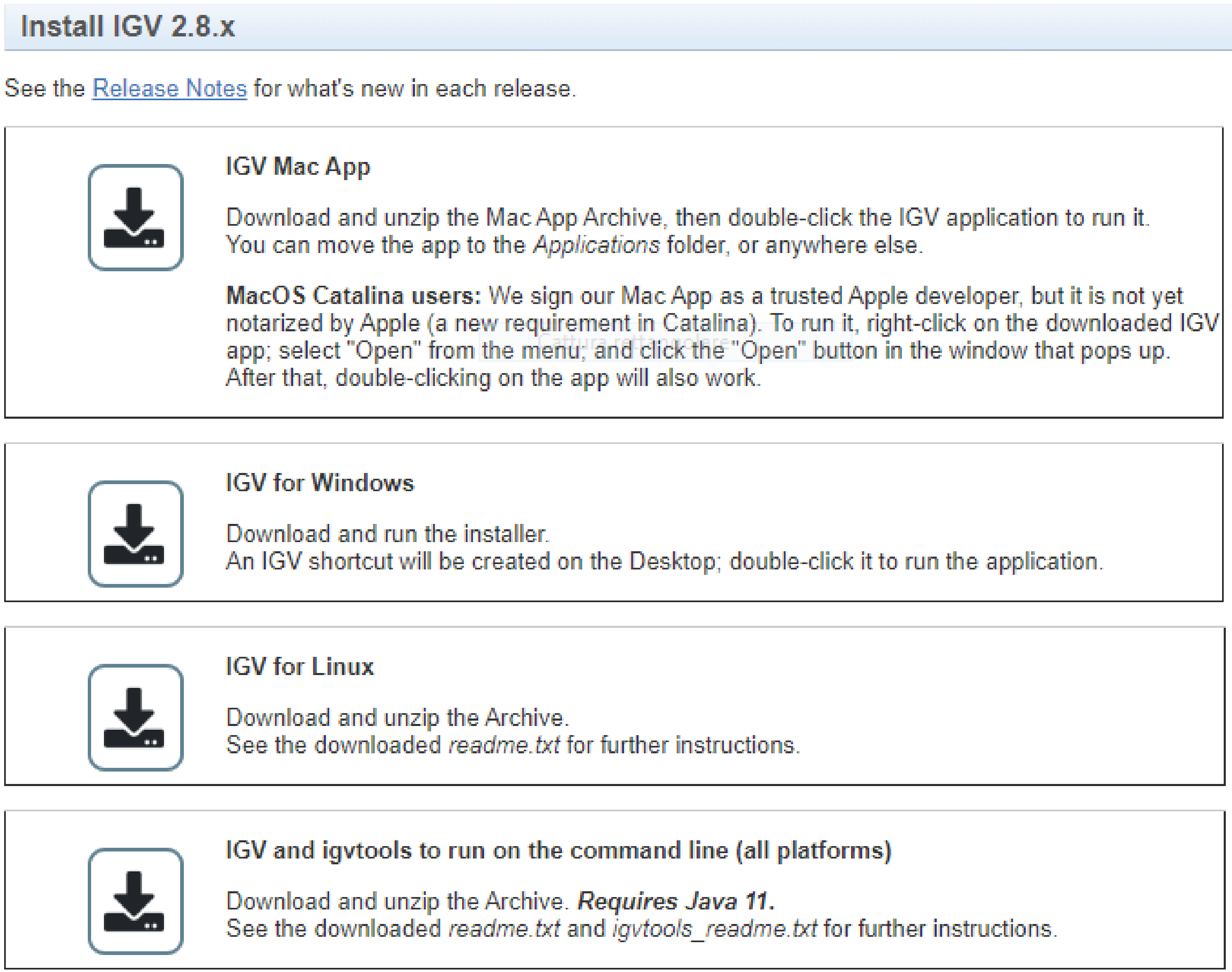
James T. Robinson, Helga Thorvaldsdóttir, Aaron M. Wenger, Ahmet Zehir, Jill P. Mesirov. **Variant Review with the Integrative Genomics Viewer (IGV).** Cancer Research 77(21) 31-34 (2017).

Funding

Development of IGV has been supported by funding from the National Cancer Institute (NCI) of the National Institutes of Health, the Informatics Technology for Cancer Research (ITCR) of the NCI, and the Starr Cancer Consortium.

IGV participates in the GenomeSpace initiative, which is funded by the National Human Genome Research Institute.

Download IGV



Install IGV 2.8.x

See the Release Notes for what's new in each release.

IGV Mac App

Download and unzip the Mac App Archive, then double-click the IGV application to run it.
You can move the app to the *Applications* folder, or anywhere else.

MacOS Catalina users: We sign our Mac App as a trusted Apple developer, but it is not yet notarized by Apple (a new requirement in Catalina). To run it, right-click on the downloaded IGV app; select "Open" from the menu; and click the "Open" button in the window that pops up. After that, double-clicking on the app will also work.

IGV for Windows

Download and run the installer.
An IGV shortcut will be created on the Desktop; double-click it to run the application.

IGV for Linux

Download and unzip the Archive.
See the downloaded *readme.txt* for further instructions.

IGV and igvtools to run on the command line (all platforms)

Download and unzip the Archive. ***Requires Java 11.***
See the downloaded *readme.txt* and *igvtools_readme.txt* for further instructions.

Download the bam and the bai

- Download the bam file and the index file on your pc:
- Open new terminal:

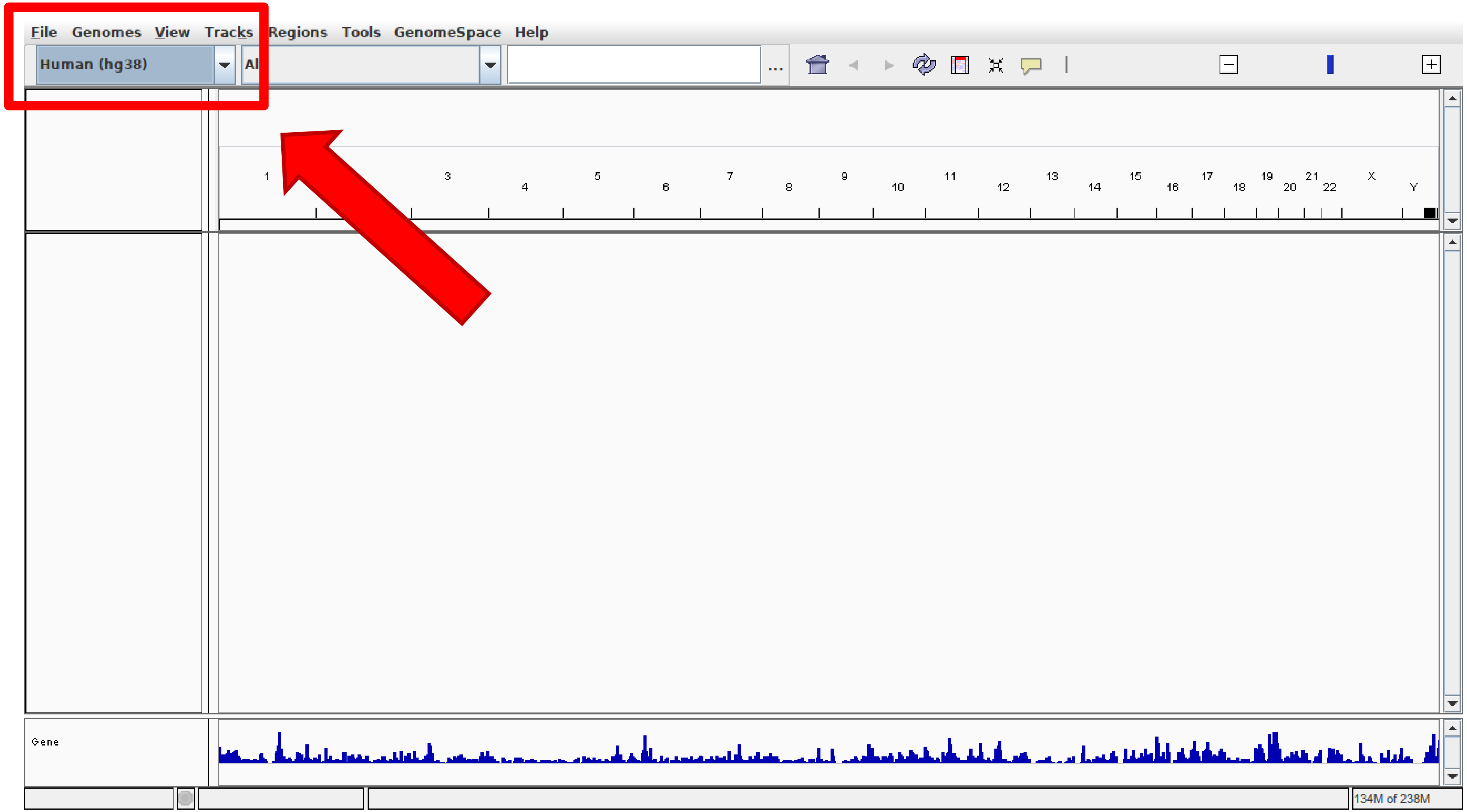
```
cd Desktop/HGSI2020
rsync -auv HGSI2020@157.27.26.214:/attachedvolume/HGSI2020/Denise/sample.sorted.bam* .
```

- Check if you have downloaded: `ls`

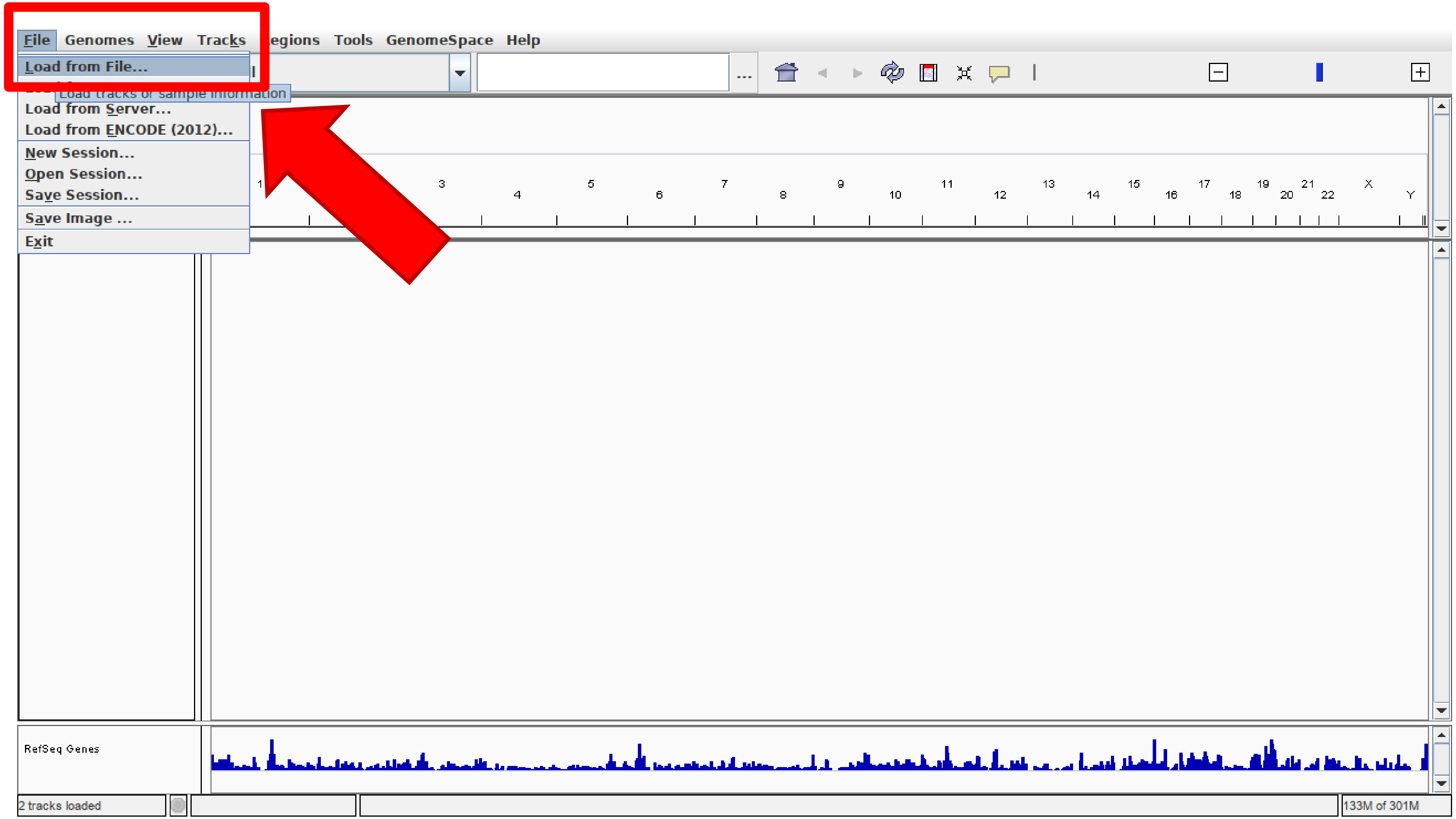
- Open IGV

```
./igv.sh for Ubuntu
```

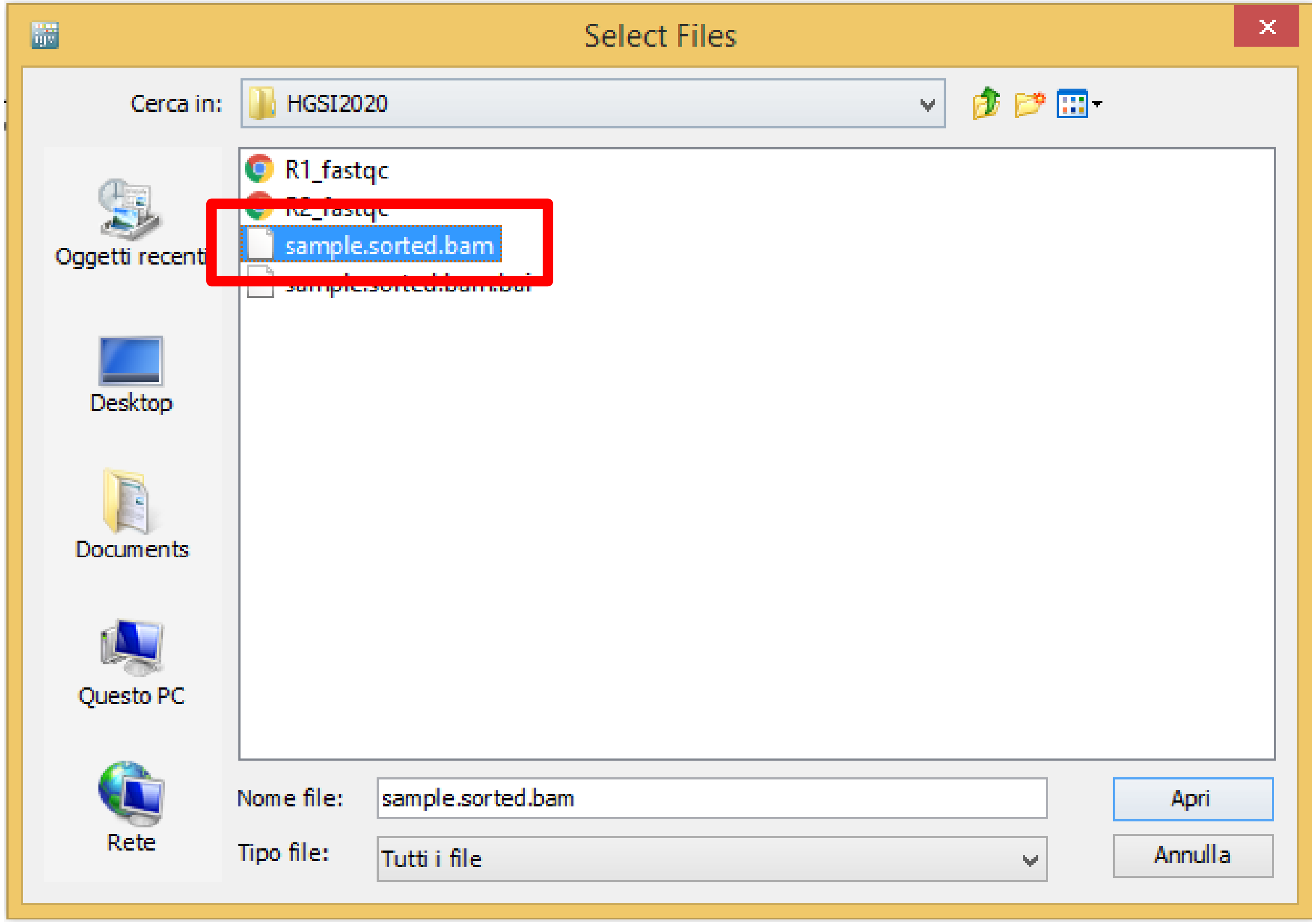
Choose the right genome



Upload the bam



Go into the folder "HGSI2020", choose the file bam and open it



Select Files

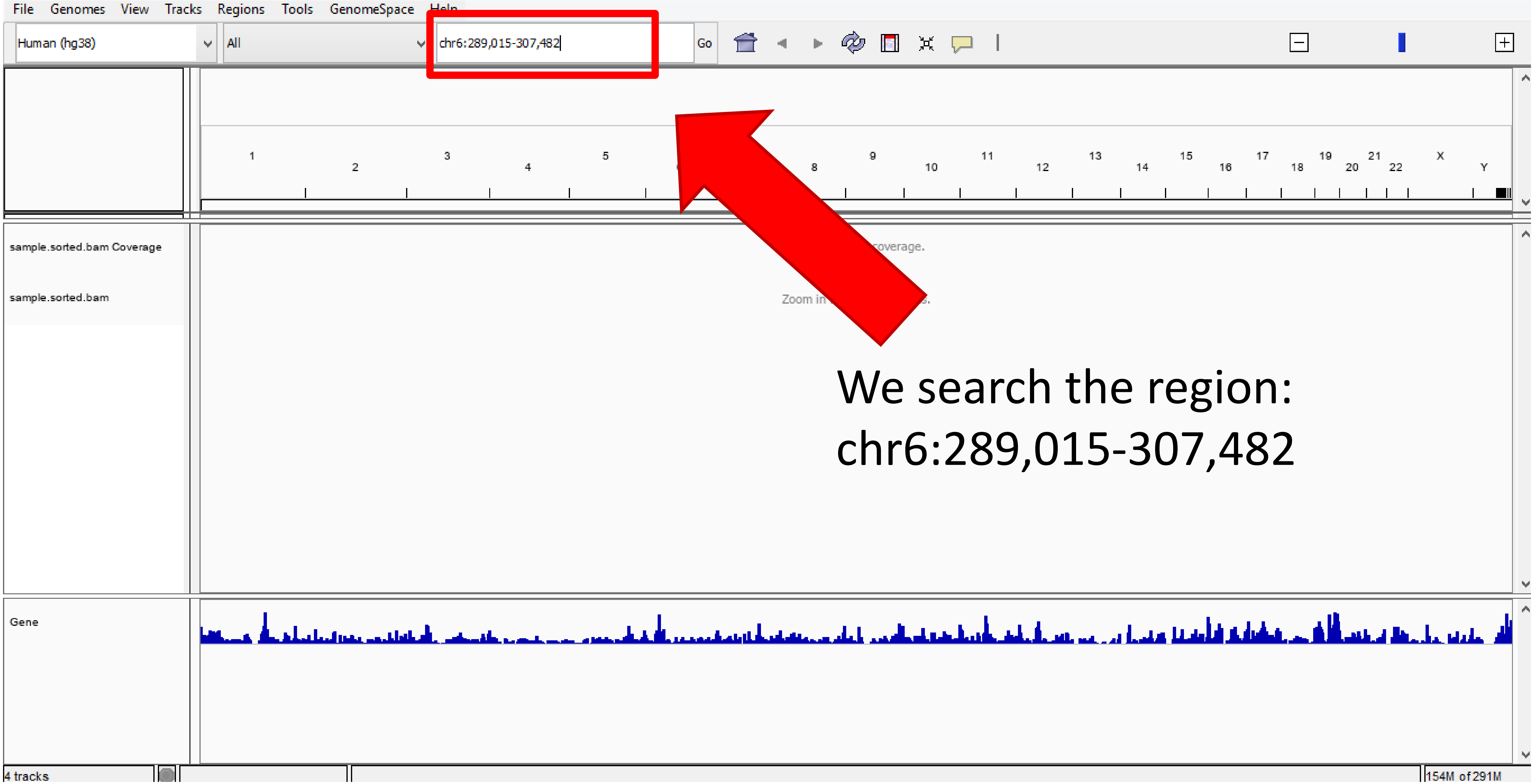
Cerca in: HGSI2020

R1_fastqc
R2_fastqc
sample.sorted.bam
sample.sorted.bam.bai

Nome file: sample.sorted.bam

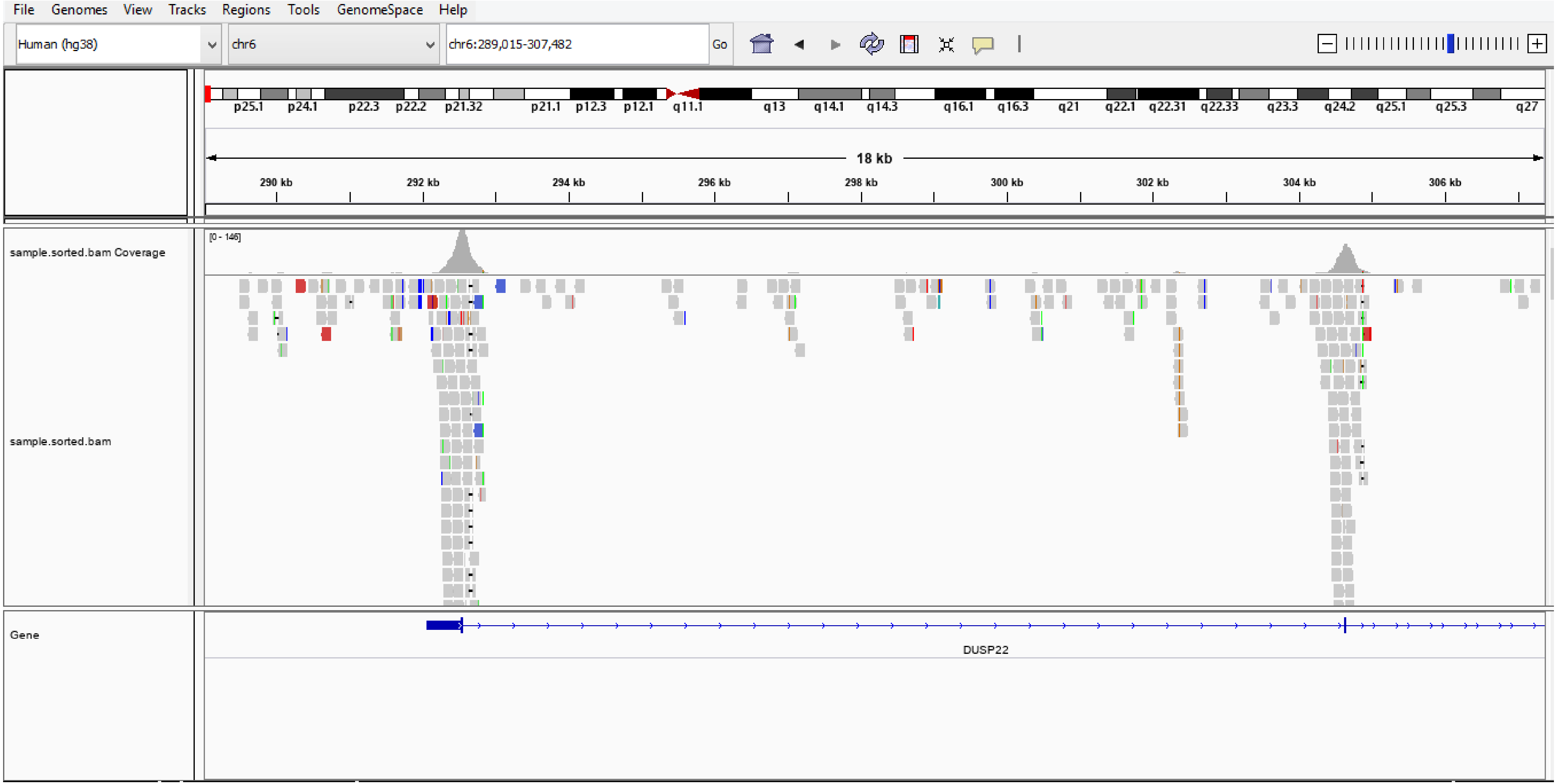
Tipo file: Tutti i file

Search a specific region



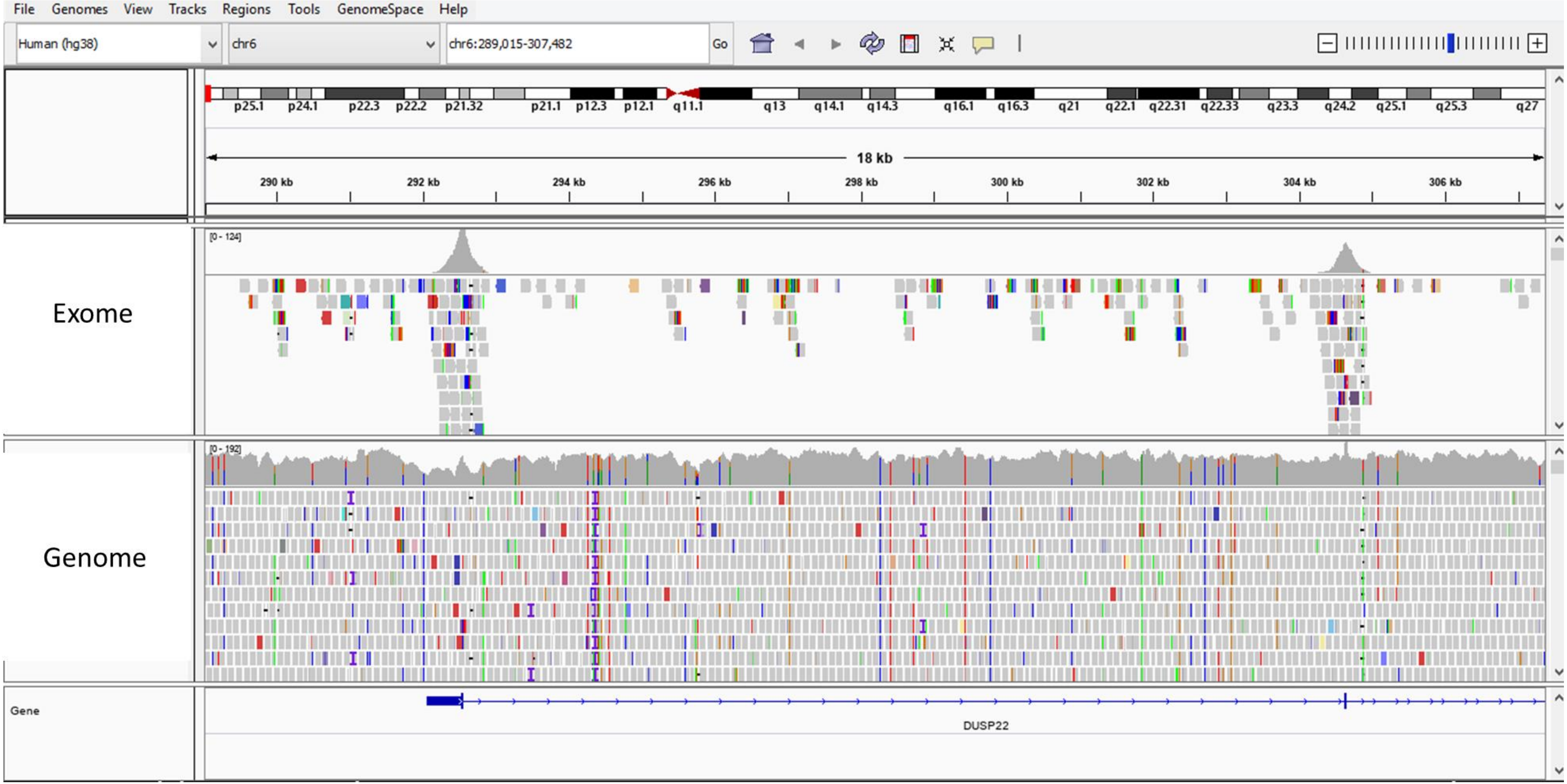
We search the region:
chr6:289,015-307,482

Results



File
Genomes
View
Tracks
Regions
Tools
GenomeSpace
Help
Human (hg38)
chr6
chr6:289,015-307,482
Go
p25.1
p24.1
p22.3
p22.2
p21.32
p21.1
p12.3
p12.1
q11.1
q13
q14.1
q14.3
q16.1
q16.3
q21
q22.1
q22.31
q22.33
q23.3
q24.2
q25.1
q25.3
q27
18 kb
290 kb
292 kb
294 kb
296 kb
298 kb
300 kb
302 kb
304 kb
306 kb
sample.sorted.bam Coverage
[0 - 146]
sample.sorted.bam
Gene
DUSP22

Difference between genome and exome sequencing


File
Genomes
View
Tracks
Regions
Tools
GenomeSpace
Help
Human (hg38)
chr6
chr6:289,015-307,482
Go
p25.1
p24.1
p22.3
p22.2
p21.32
p21.1
p12.3
p12.1
q11.1
q13
q14.1
q14.3
q16.1
q16.3
q21
q22.1
q22.31
q22.33
q23.3
q24.2
q25.1
q25.3
q27
18 kb
290 kb
292 kb
294 kb
296 kb
298 kb
300 kb
302 kb
304 kb
306 kb
[0 - 124]
Exome
[0 - 192]
Genome
Gene
DUSP22