Mansoura University
Faculty of Computers and Information
Department of Computer Science
First Semester: 2020-2021

[MED121] Bioinformatics: Sequencing Technologies II

Grade: Third Year (Medical Informatics Program)

Sara El-Metwally, Ph.D.

Faculty of Computers and Information,

Mansoura University,

Egypt.

AGENDA

- Key attributes of different Sequencing Technologies.
- Sequencing Projects.
- Data Deluge
- Sequencing Services.
- Base Qualities
- FASTA/FASTQ Genomic files.
- Genome Annotations.

SEQUENCING TECHNOLOGIES

Throughput

Read length

sequencing errors

Sequencing cost

Bioinformatics tools available for data analysis and processing

TOP TECHNOLOGIES IN THE SEQUENCING MARKET.

Company	Instruments

Illumina	MiniSeq;
	NextSeq;
	MiSeq;
	HiSeq;
	NovaSeq

Pacific biosciences	RSII; Sequel
Oxford Nanopore Technologies	SmidgION (under dev); MinION; GridION; PromethION (under dev)

Table credits : **https://academic.oup.com/femsle/article/365/16/fny103/4982775**

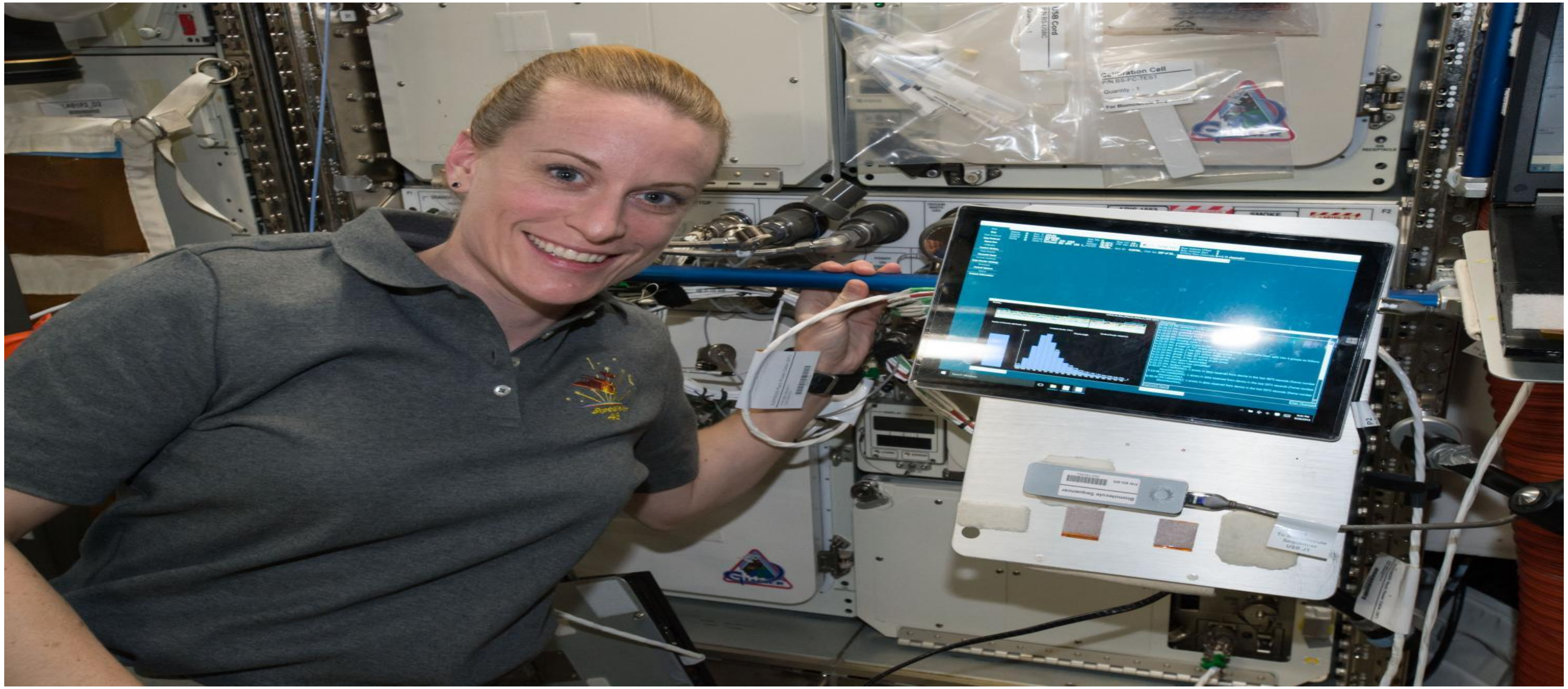
Sequencing Power for Every Scale

The broadest portfolio available

Sequencing System	iSeq™	MiniSeq™	MiSeq®	NextSeq®	HiSeq®	HiSeq® X	NovaSeq®
					4000	Five/Ten	6000
Output per run	1.2 Gb	7.5 Gb	15 Gb	120 Gb	1.5 Tb	1.8 Tb	1 Tb - 6 Tb[1]
Instrument price	$19.9K	$49.5K	$99K	$275K	$900K	$6M[2]/$10M[2]	$985K
Installed base[3]	NA	~600	~6,000	~2,400	~2,300[4]		~285

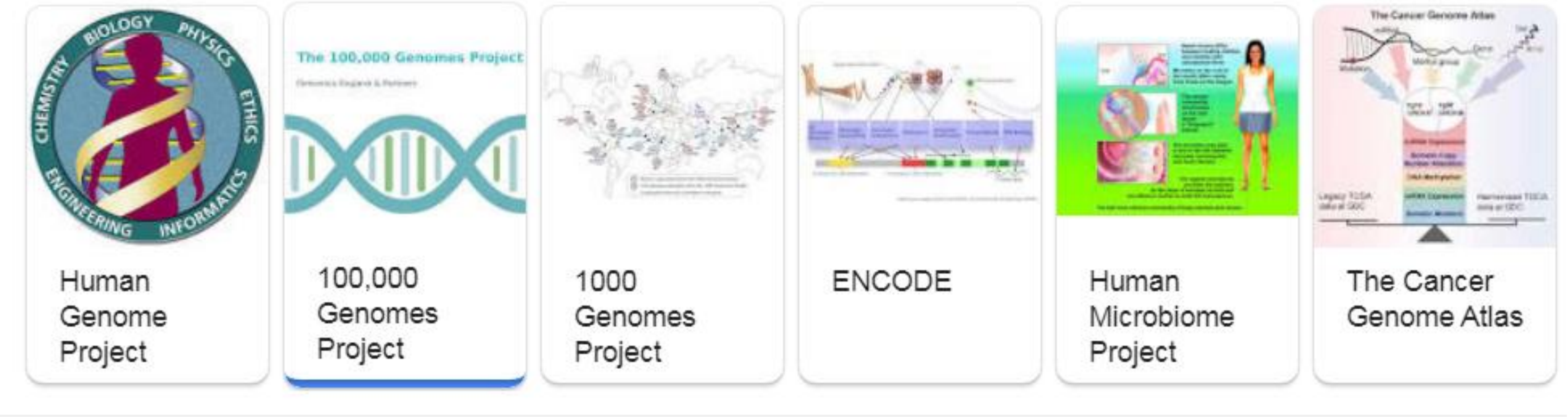
1. Output per run for the S1, S2 and S4 flow cells equal 1 Tb, 2 Tb and 6 Tb, respectively assuming two flow cells per run
2. Based on purchase of 5 and 10 units for HiSeq X Five and HiSeq X Ten, respectively
3. Based on end of fiscal year 2017
4. Combined HiSeq family

SEQUENCING BY NANOPORE (FUN!)



Kate Rubins is pictured aboard ISS with the USB MinION sequencer (lower right) that was used in the first-ever DNA sequencing in space in August 2016.

SEQUENCING PROJECTS



Human Genome Project

100,000 Genomes Project

1000 Genomes Project

ENCODE

Human Microbiome Project

The Cancer Genome Atlas

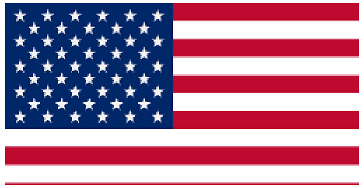
SEQUENCING PROJECTS

China—100,000 Genomes Project

United Kingdom—100,000 Genomes Project

Turkey—Turkish Genome Project

France—France Génomique (Médicine France Génomique 2025 or French Plan for Genomic Medicine 2025)



United States

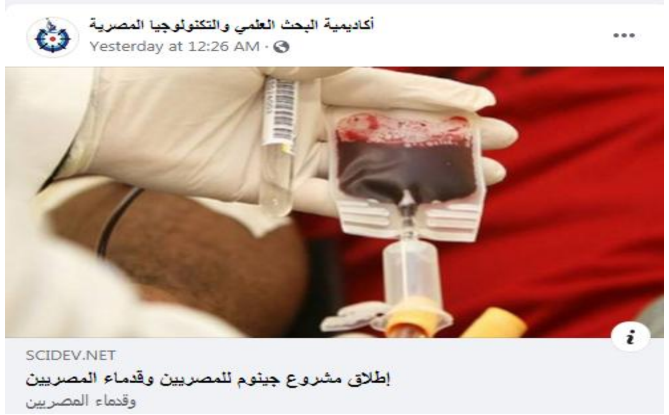
Dubai, United Arab Emirates—Dubai Genomics

Saudi Arabia—Saudi Human Genome Program

Japan—Initiative on Rare and Undiagnosed Diseases

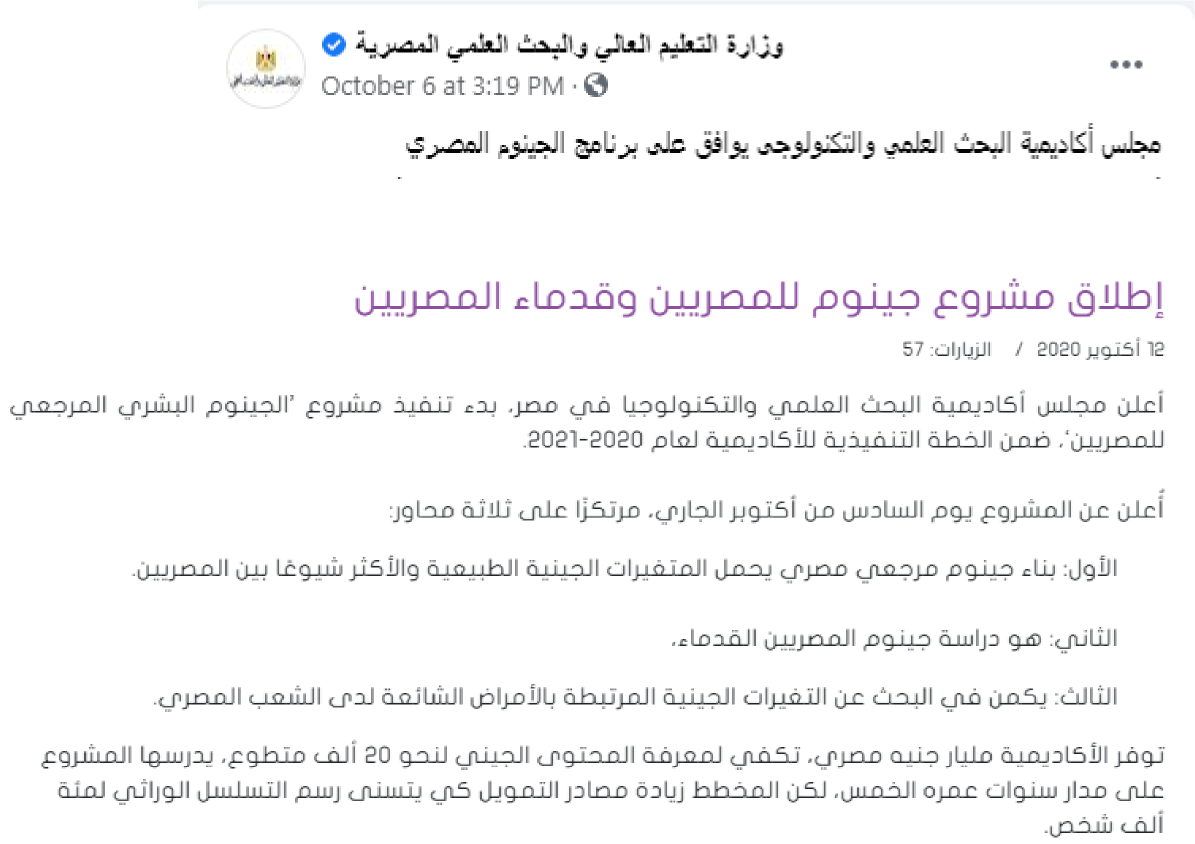
https://www.clinicalomics.com/topics/translational-research/biomarkers-topic/biobanking/10-countries-in-100k-genome-club/

SEQUENCING PROJECTS



أكاديمية البحث العلمي والتكنولوجيا المصرية
Yesterday at 12:26 AM ·

SCIDEV.NET
إطلاق مشروع جينوم للمصريين وقدماء المصريين
وقدماء المصريين



وزارة التعليم العالي والبحث العلمي المصرية
October 6 at 3:19 PM ·

مجلس أكاديمية البحث العلمي والتكنولوجى يوافق على برنامج الجينوم المصري

إطلاق مشروع جينوم للمصريين وقدماء المصريين

12 أكتوبر 2020 / الزيارات: 57

أعلن مجلس أكاديمية البحث العلمي والتكنولوجيا في مصر، بدء تنفيذ مشروع 'الجينوم البشري المرجعي للمصريين'، ضمن الخطة التنفيذية للأكاديمية لعام 2020-2021.

أُعلن عن المشروع يوم السادس من أكتوبر الجاري، مرتكزًا على ثلاثة محاور:

الأول: بناء جينوم مرجعي مصري يحمل المتغيرات الجينية الطبيعية والأكثر شيوعًا بين المصريين.

الثاني: هو دراسة جينوم المصريين القدماء،

الثالث: يكمن في البحث عن التغيرات الجينية المرتبطة بالأمراض الشائعة لدى الشعب المصري.

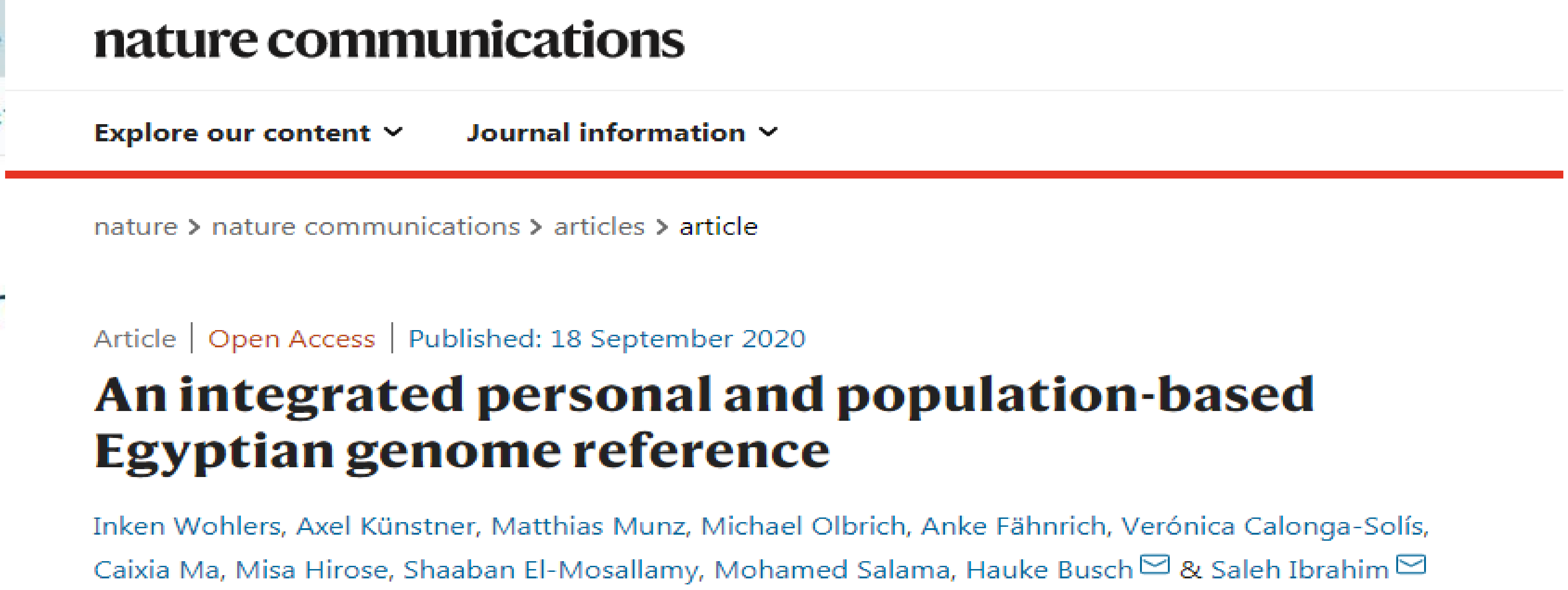
توفر الأكاديمية مليار جنيه مصري، تكفي لمعرفة المحتوى الجيني لنحو 20 ألف متطوع، يدرسها المشروع على مدار سنوات عمره الخمس، لكن المخطط زيادة مصادر التمويل كي يتسنى رسم التسلسل الوراثي لمئة ألف شخص.

EGYPTIAN GENOME

EgyptRef | Personal Genome

Home | الملخص العربي | Contac

An Egyptian Genor


nature communications

Explore our content ⌄ | Journal information ⌄

nature > nature communications > articles > article

Article | Open Access | Published: 18 September 2020

An integrated personal and population-based Egyptian genome reference

Inken Wohlers, Axel Künstner, Matthias Munz, Michael Olbrich, Anke Fähnrich, Verónica Calonga-Solís, Caixia Ma, Misa Hirose, Shaaban El-Mosallamy, Mohamed Salama, Hauke Busch & Saleh Ibrahim

We have taken advantage of these technologies (PacBio, 10X Genomics, Illumina) to sequence and de-novo assemble the genome of an Egyptian individual. We integrated the sequences of an additional 109 Egyptian individuals to generate an Egyptian Reference

DATA DELUGE

Sequencing Centers 2018

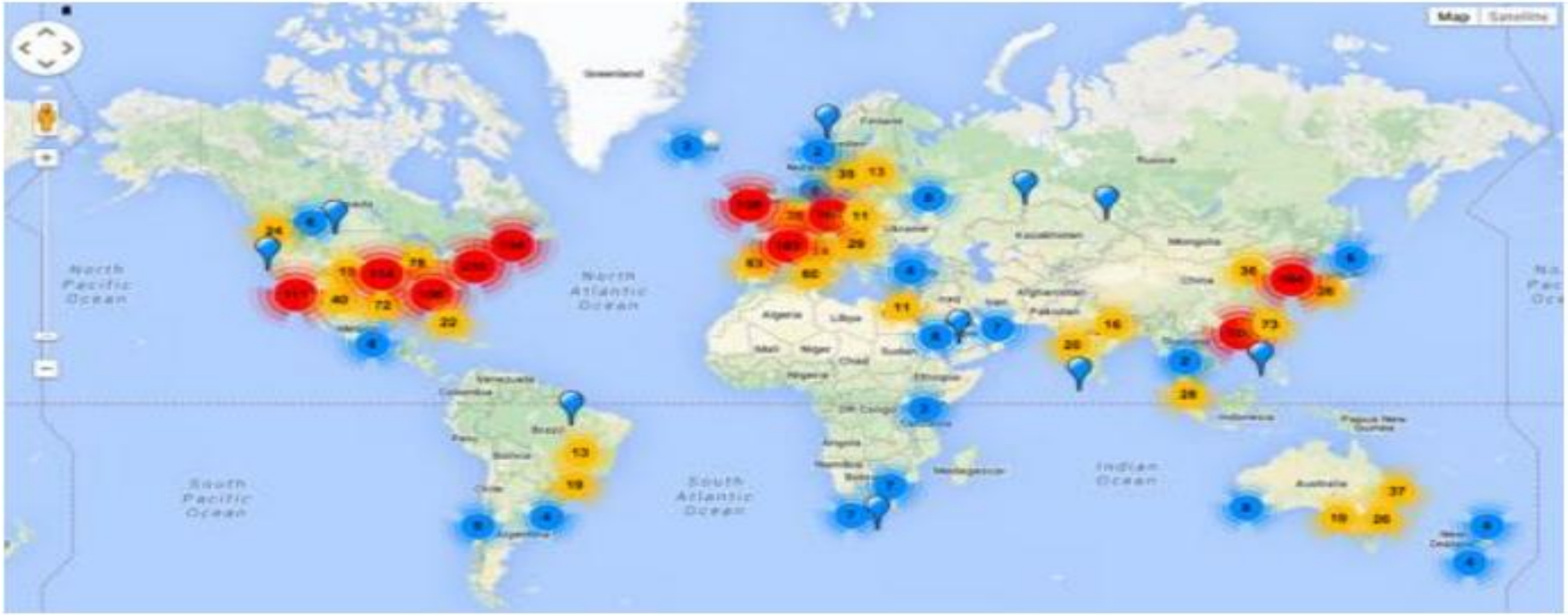


Image credit: https://pubmed.ncbi.nlm.nih.gov/24920863/

DATA DELUGE

Sequencing Centers 2028

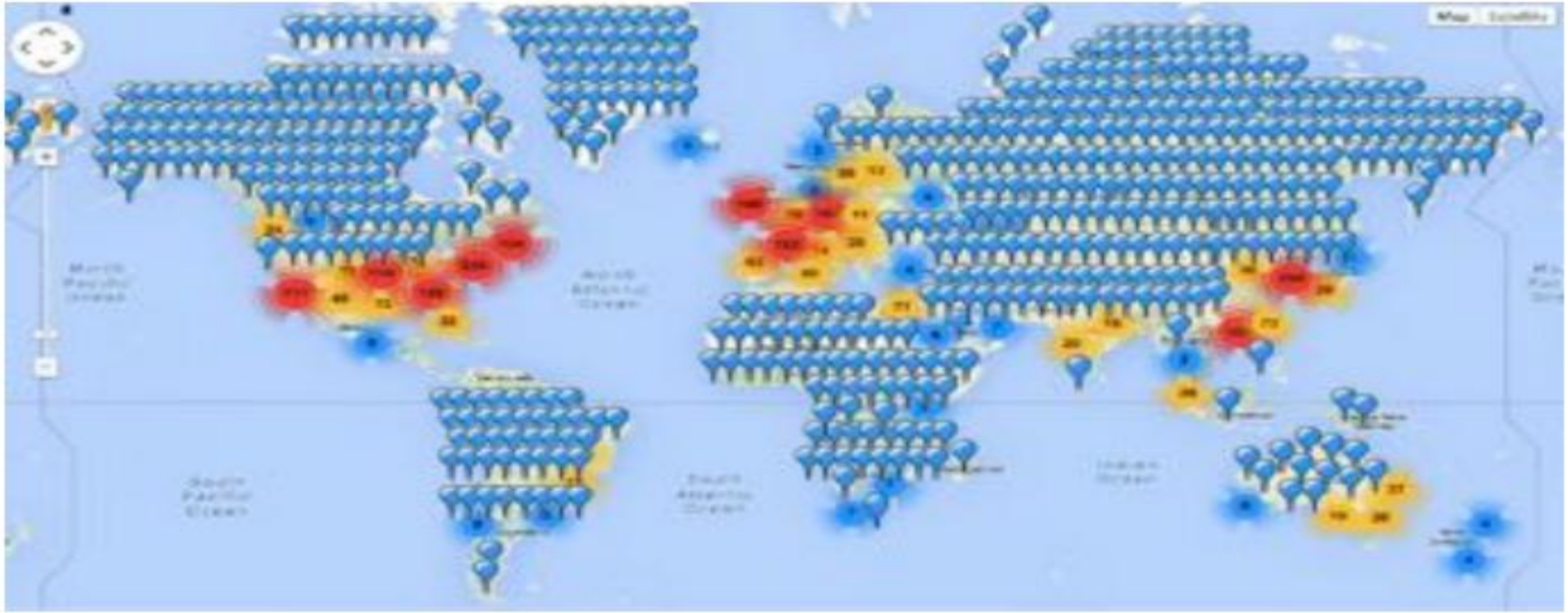


Image credit: https://pubmed.ncbi.nlm.nih.gov/24920863/

SEQUENCING SERVICES

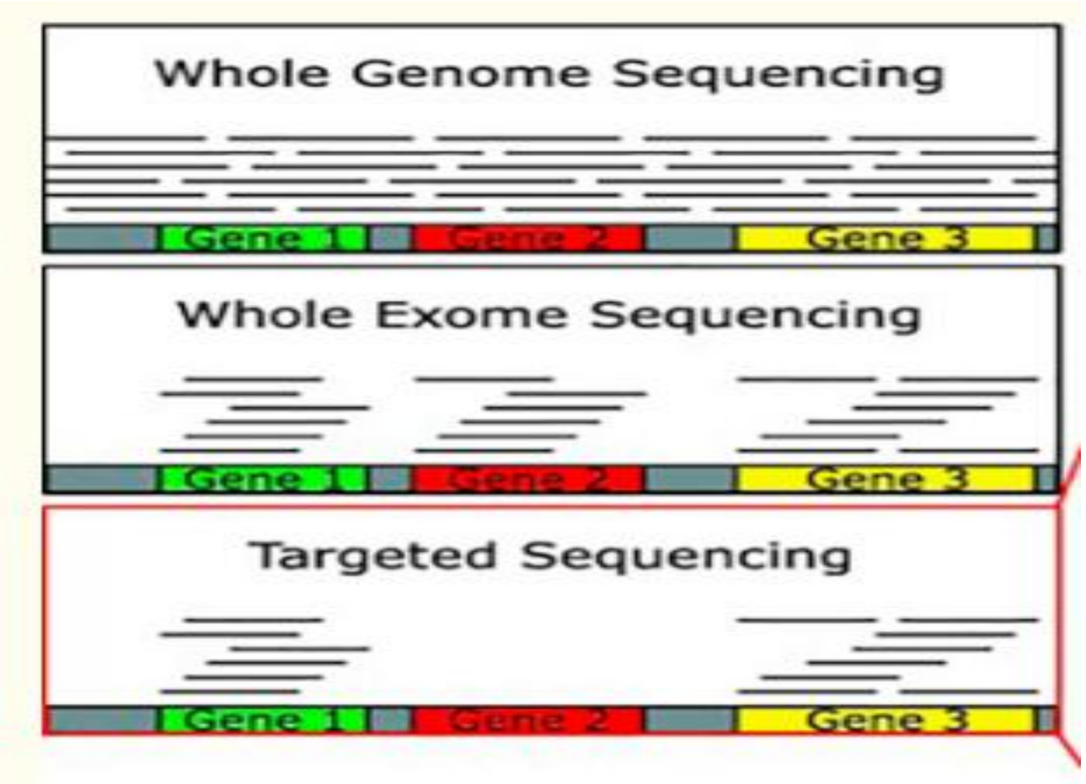
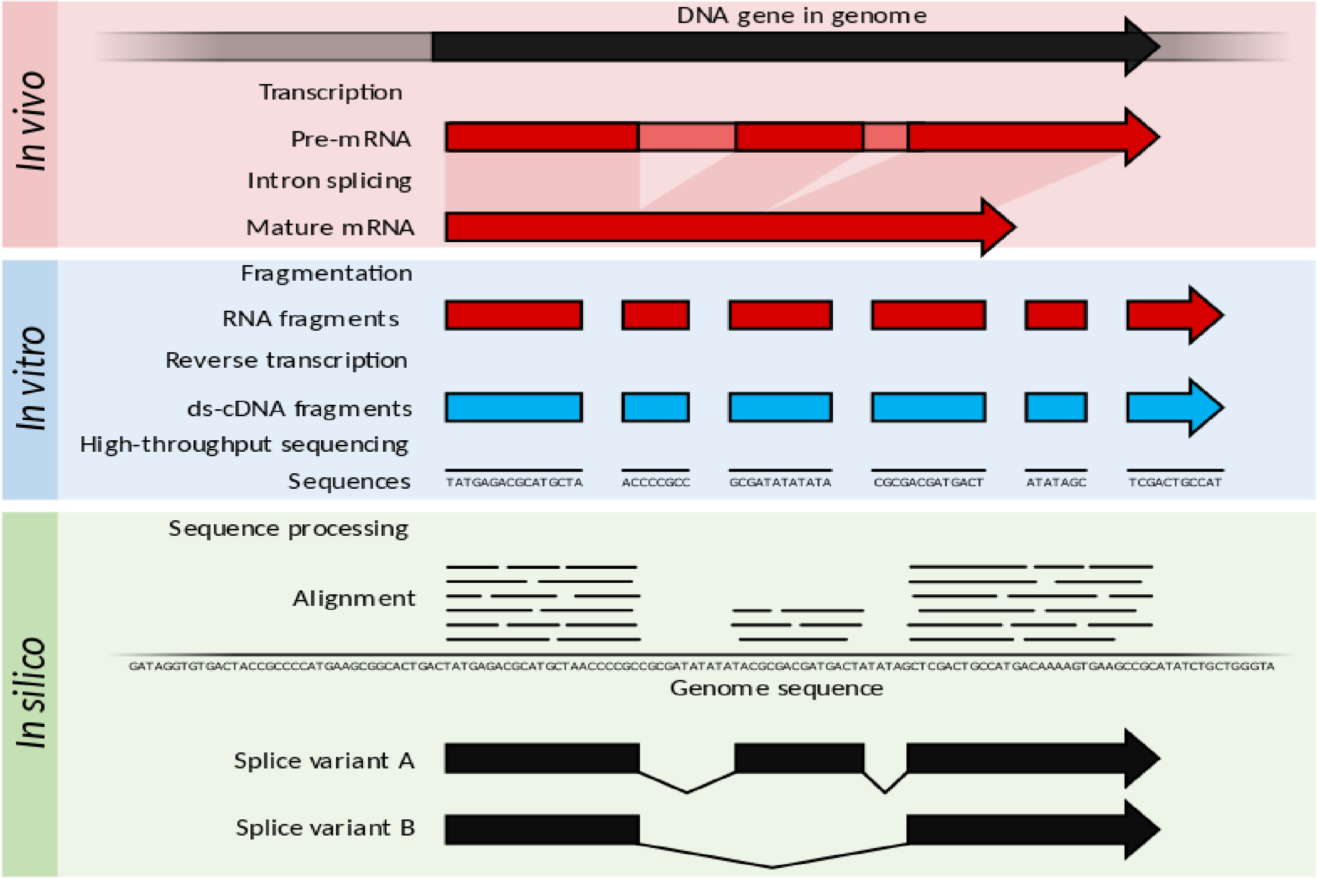


Image credits: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6861594/

https://www.abmgood.com/Whole-Genome-Sequencing-Service.html

For example:
the genes *KRAS* and *TP53* are often targeted across a range of cancer types, as they are commonly found to be mutated with a number of hotspots. *BRAF* and *EGFR* are also screened in many solid tumors, as they contain clinically relevant mutation

RNA-SEQ



- RNA-seq is a particular technology-based sequencing technique which uses next-generation sequencing (NGS) to reveal the presence and quantity of RNA in a biological sample at a given moment, analyzing the continuously changing cellular transcriptome.

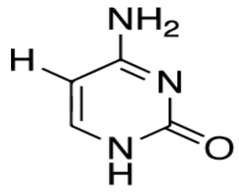
Image credit: https: https://en.wikipedia.org/wiki/RNA-Seq#/media/File:Summary_of_RNA-Seq.svg

NOTES

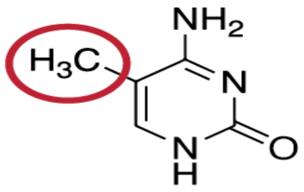
- Protein sequencing refers to methods for determining the amino acid sequence of proteins (or peptides) and analysis of the sequence, for example to infer protein conformation. Techniques include mass spectrometry and the Edman degradation reaction as well as prediction of the protein sequence from the encoding DNA or mRNA sequence.

SEQUENCING SERVICES

- **Bisulfite Sequencing: is the use of bisulfite treatment of DNA before routine sequencing to determine the pattern of methylation.**
- **DNA methylation is a biological process by which methyl groups are added to the DNA molecule. Methylation can change the activity of a DNA segment without changing the sequence. When located in a gene promoter, DNA methylation typically acts to repress gene transcription.**



Cytosine



methylated Cytosine

Image credit: https://en.wikipedia.org/wiki/DNA_methylation#/media/File:DNA_methylation.png

SEQUENCING SERVICES

- In mammals, DNA methylation is essential for normal development and is associated with a number of key processes including genomic imprinting, X-chromosome inactivation, repression of transposable elements, aging, and carcinogenesis.
- Treatment of DNA with bisulfite converts cytosine residues to uracil, but leaves 5-methylcytosine residues unaffected. Therefore, DNA that has been treated with bisulfite retains only methylated cytosines. Thus, bisulfite treatment introduces specific changes in the DNA sequence that depend on the methylation status of individual cytosine residues, yielding single-nucleotide resolution information about the methylation status of a segment of DNA.

SEQUENCING SERVICES

- ChIP-sequencing, also known as ChIP-seq, is a method used to analyze protein interactions with DNA. ChIP-seq combines chromatin immunoprecipitation (ChIP) with massively parallel DNA sequencing to identify the binding sites of DNA-associated proteins. It can be used to map global binding sites precisely for any protein of interest.
- Single cell sequencing examines the sequence information from individual cells with optimized next-generation sequencing (NGS) technologies, providing a higher resolution of cellular differences and a better understanding of the function of an individual cell .

SEQUENCING SERVICES

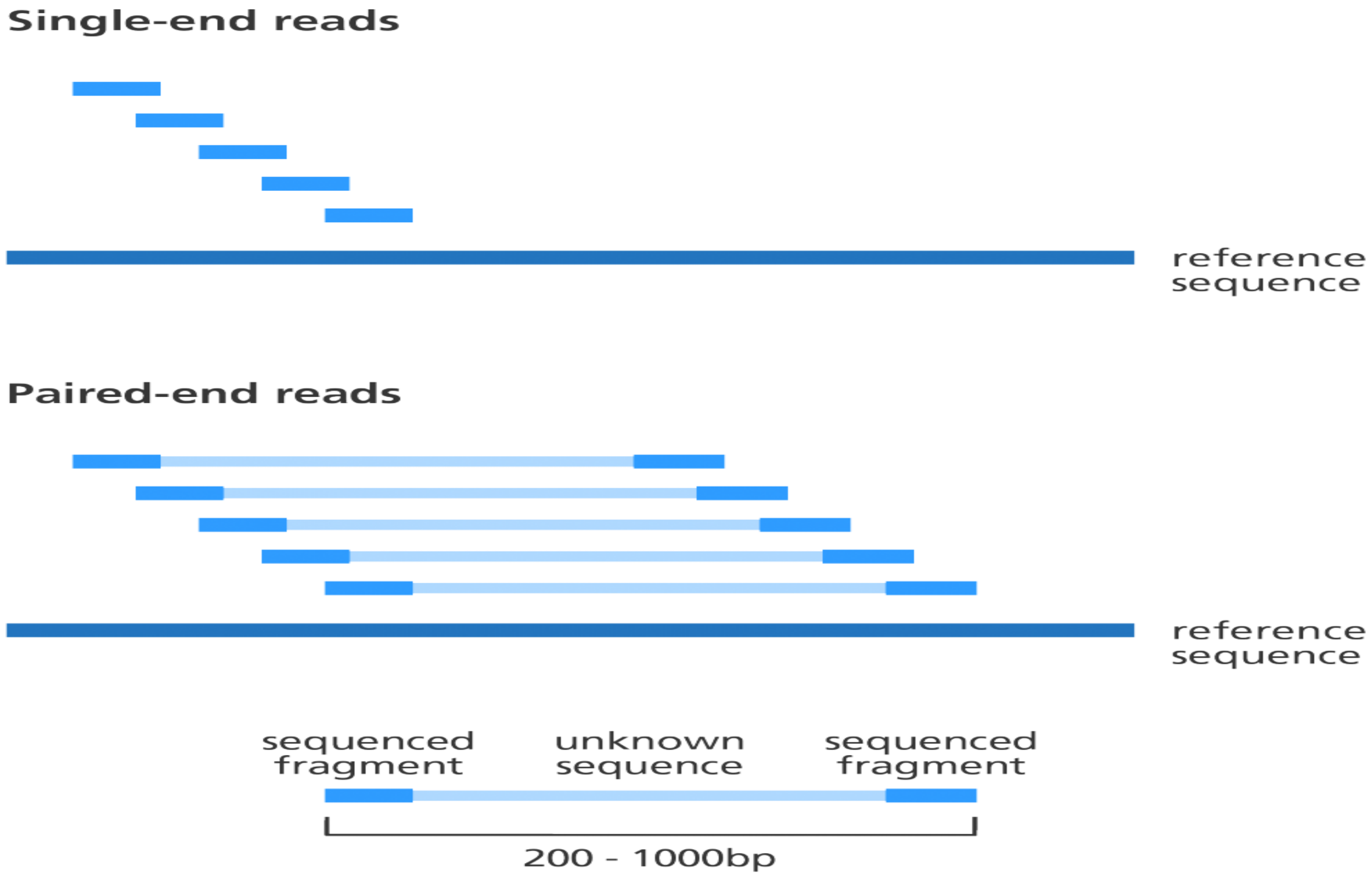


Image credit: https://www.biostars.org/p/267167/#267170

SEQUENCING SERVICES

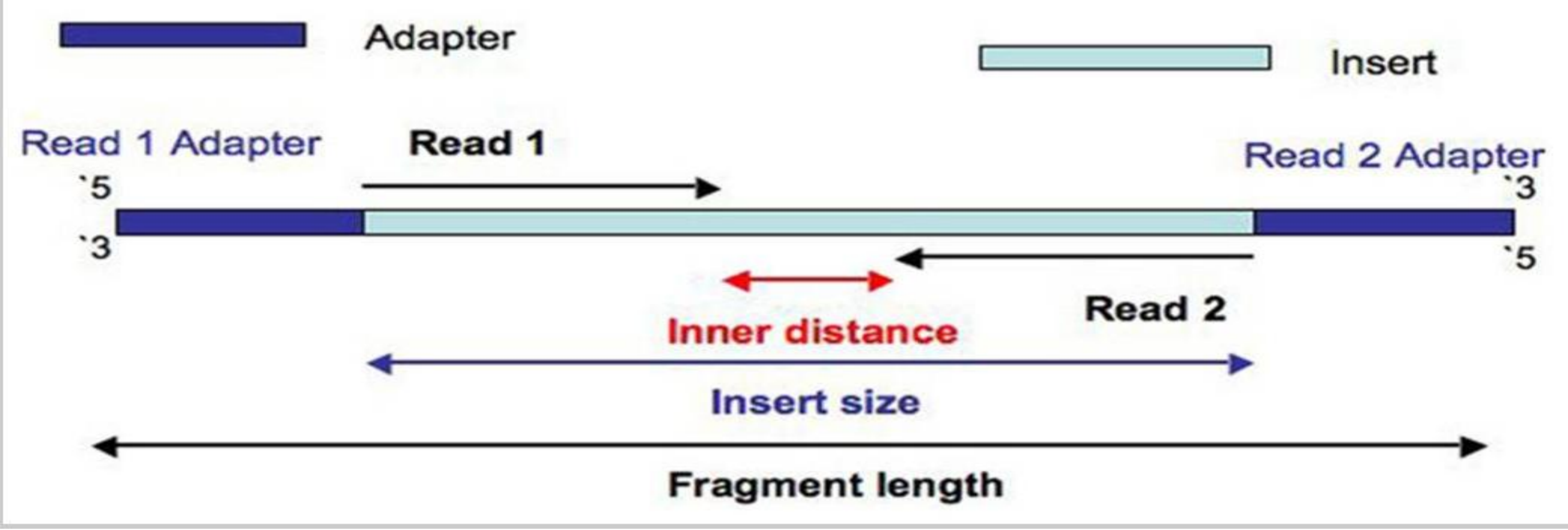


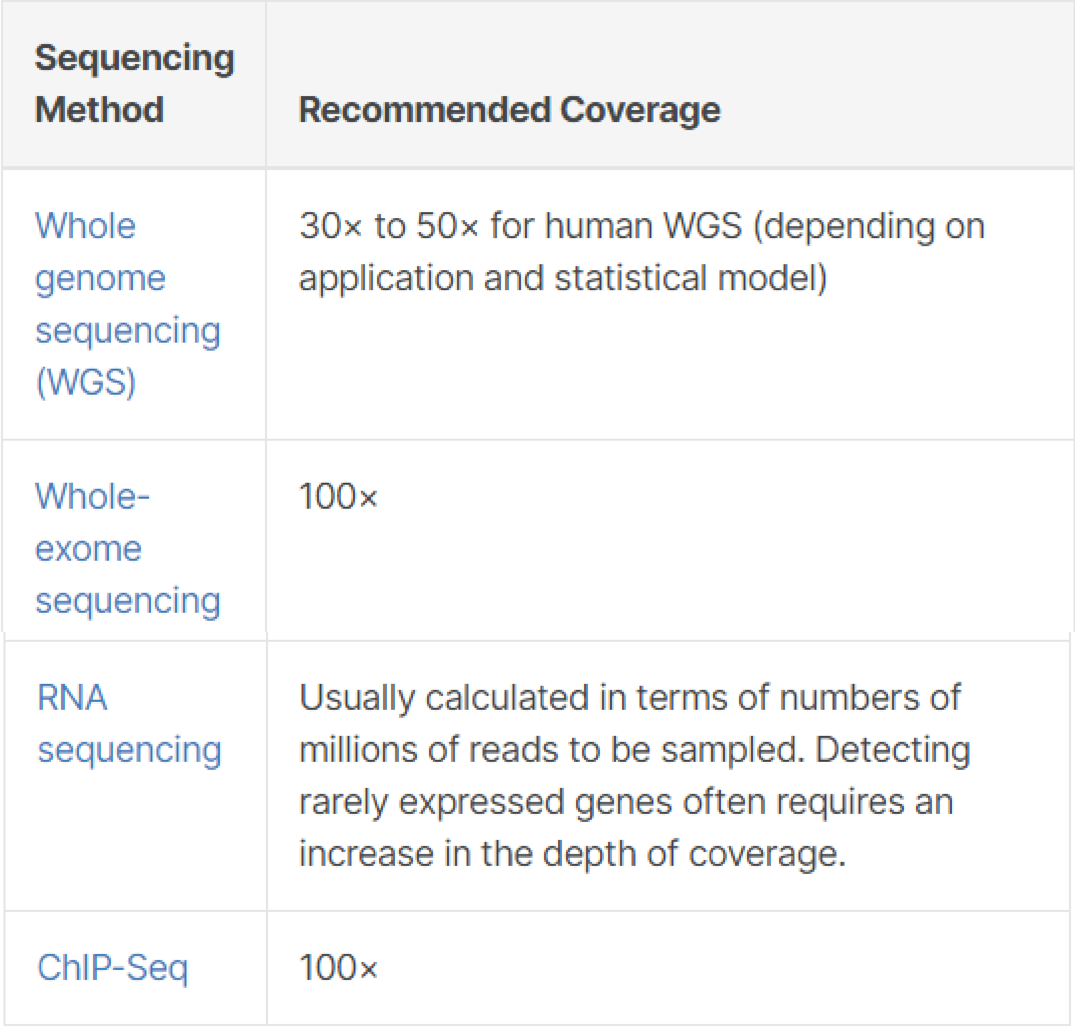
Image credit: https://thesequencingcenter.com/knowledge-base/what-are-paired-end-reads/

COVERAGE DEPTH

- Per-base coverage is the average number of times a base of a genome is sequenced. The coverage depth of a genome is calculated as the number of bases of all short reads that match a genome divided by the length of this genome. It is often expressed as 1X, 2X, 3X,… (1, 2, or, 3 times coverage).
- coverage describes the average number of reads that align to, or "cover," known reference bases. The sequencing coverage level often determines whether variant discovery can be made with a certain degree of confidence at particular base positions.
- At higher levels of coverage, each base is covered by a greater number of aligned sequence reads, so base calls can be made with a higher degree of confidence.

https://www.ecseq.com/support/ngs/how-to-calculate-the-coverage-for-a-sequencing-experiment

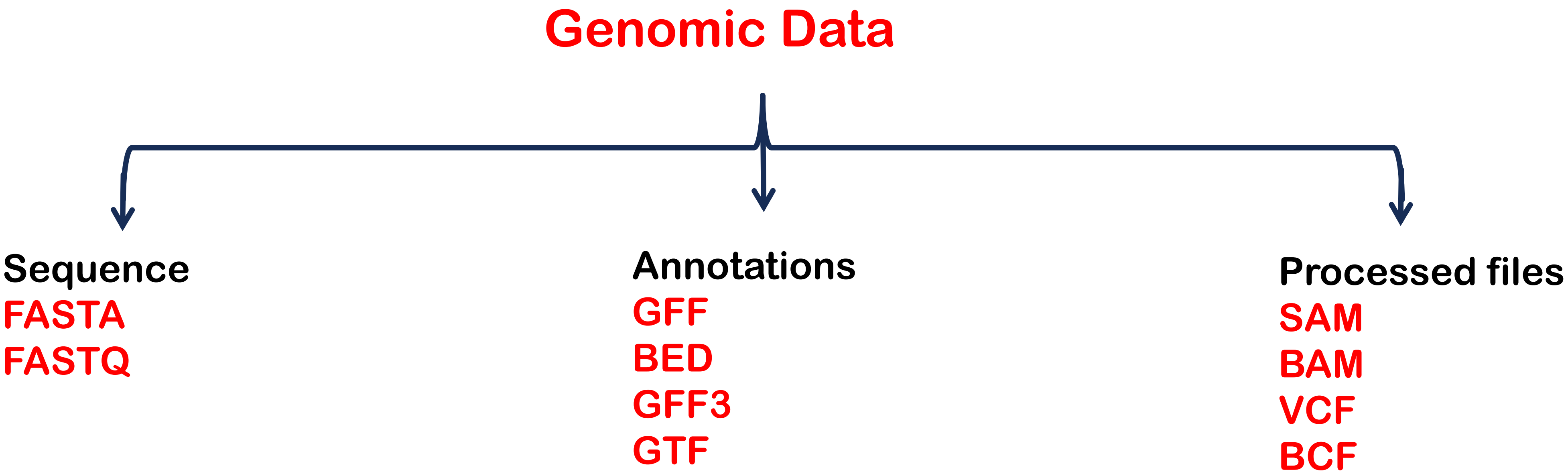
COVERAGE DEPTH



Sequencing Method	Recommended Coverage
Whole genome sequencing (WGS)	30× to 50× for human WGS (depending on application and statistical model)
Whole-exome sequencing	100×
RNA sequencing	Usually calculated in terms of numbers of millions of reads to be sampled. Detecting rarely expressed genes often requires an increase in the depth of coverage.
ChIP-Seq	100×

Table Credit:
https://www.illumina.com/science/technology/next-generation-sequencing/plan-experiments/coverage.html

GENOMIC DATA



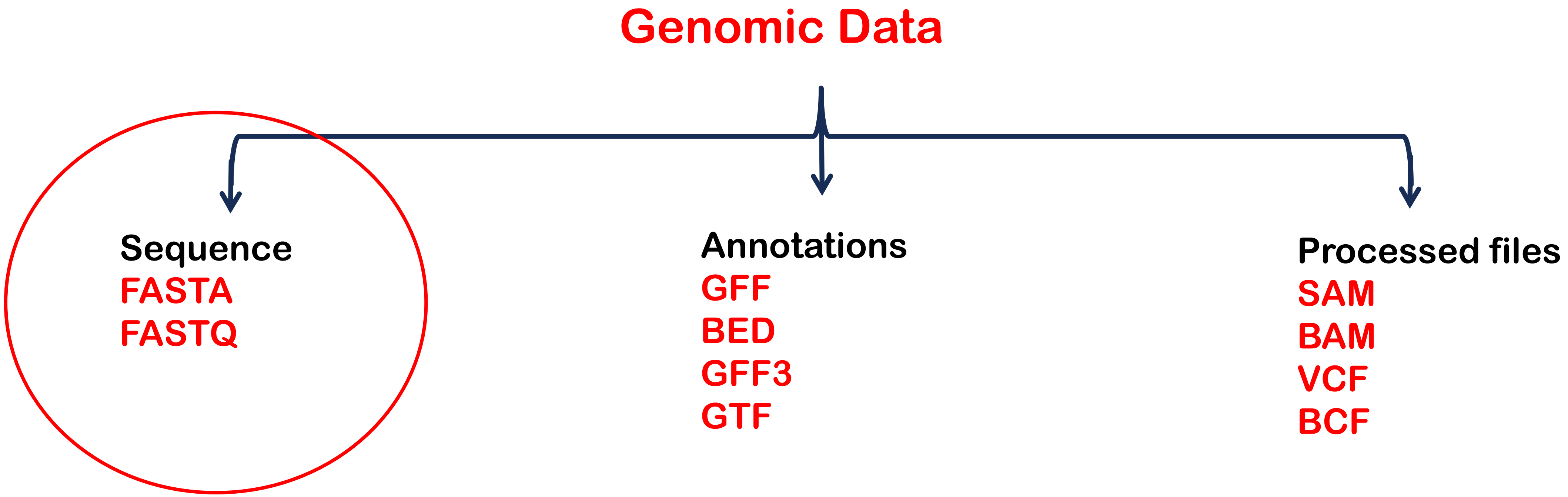
Genomic Data

Sequence
FASTA
FASTQ

Annotations
GFF
BED
GFF3
GTF

Processed files
SAM
BAM
VCF
BCF

GENOMIC DATA



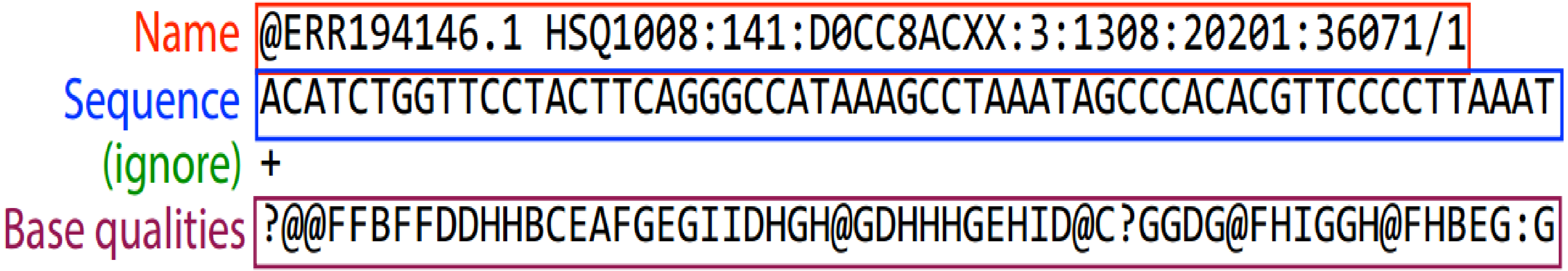
Genomic Data

Sequence
FASTA
FASTQ

Annotations
GFF
BED
GFF3
GTF

Processed files
SAM
BAM
VCF
BCF

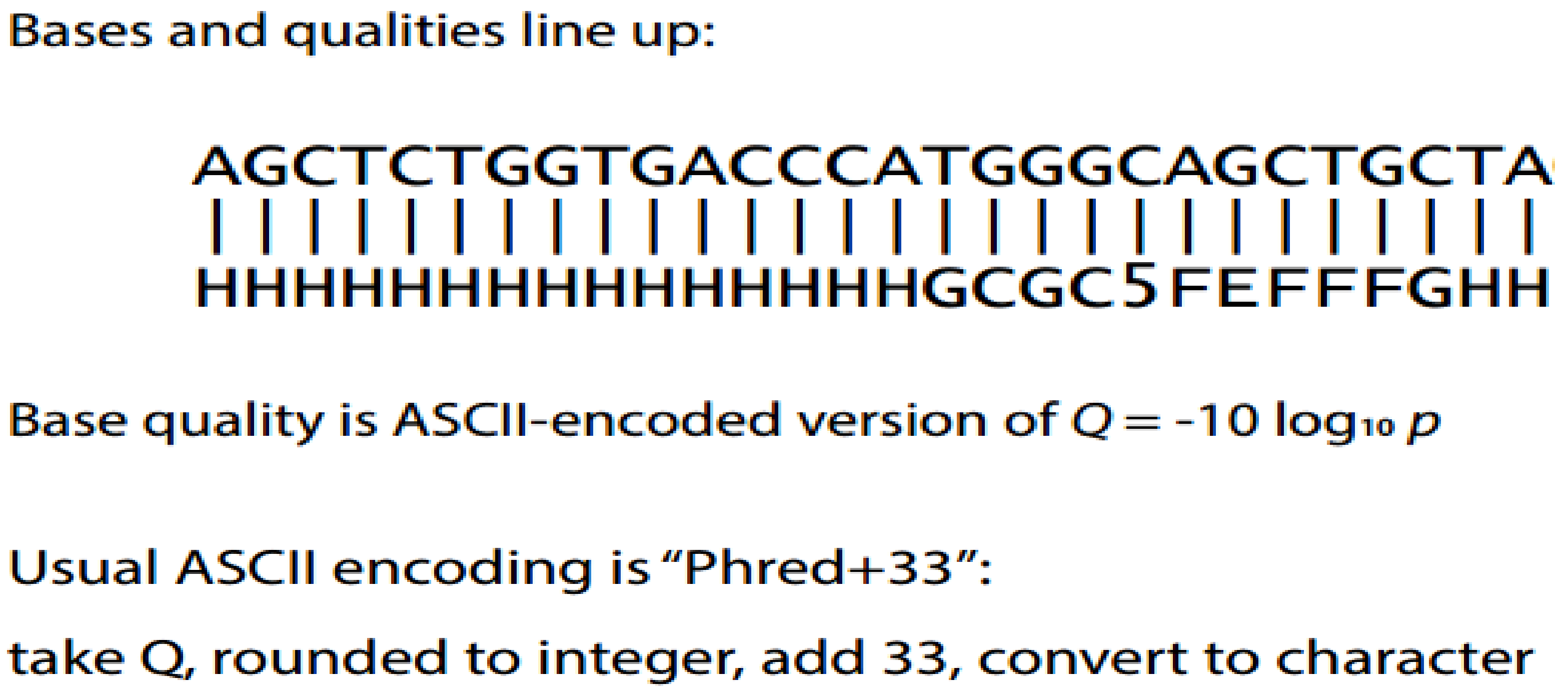
GENOMIC DATA (A DATA IN FASTQ)


Name
Sequence
(ignore)
Base qualities

```
@ERR194146.1 HSQ1008:141:D0CC8ACXX:3:1308:20201:36071/1
ACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAAT
+
?@@FFBFFDDHHBCEAFGEGIIDHGH@GDHHHGEHID@C?GGDG@FHIGGH@FHBEG:G
```

Slide Credit: Ben Langmead course of Algorithms for DNA Sequencing

BASE QUALITIES



Bases and qualities line up:

```
AGCTCTGGTGACCCATGGGCAGCTGCTAGGGA
||||||||||||||||||||||||||||||||
HHHHHHHHHHHHHHHGCGC5FEFFFGHHHHHH
```

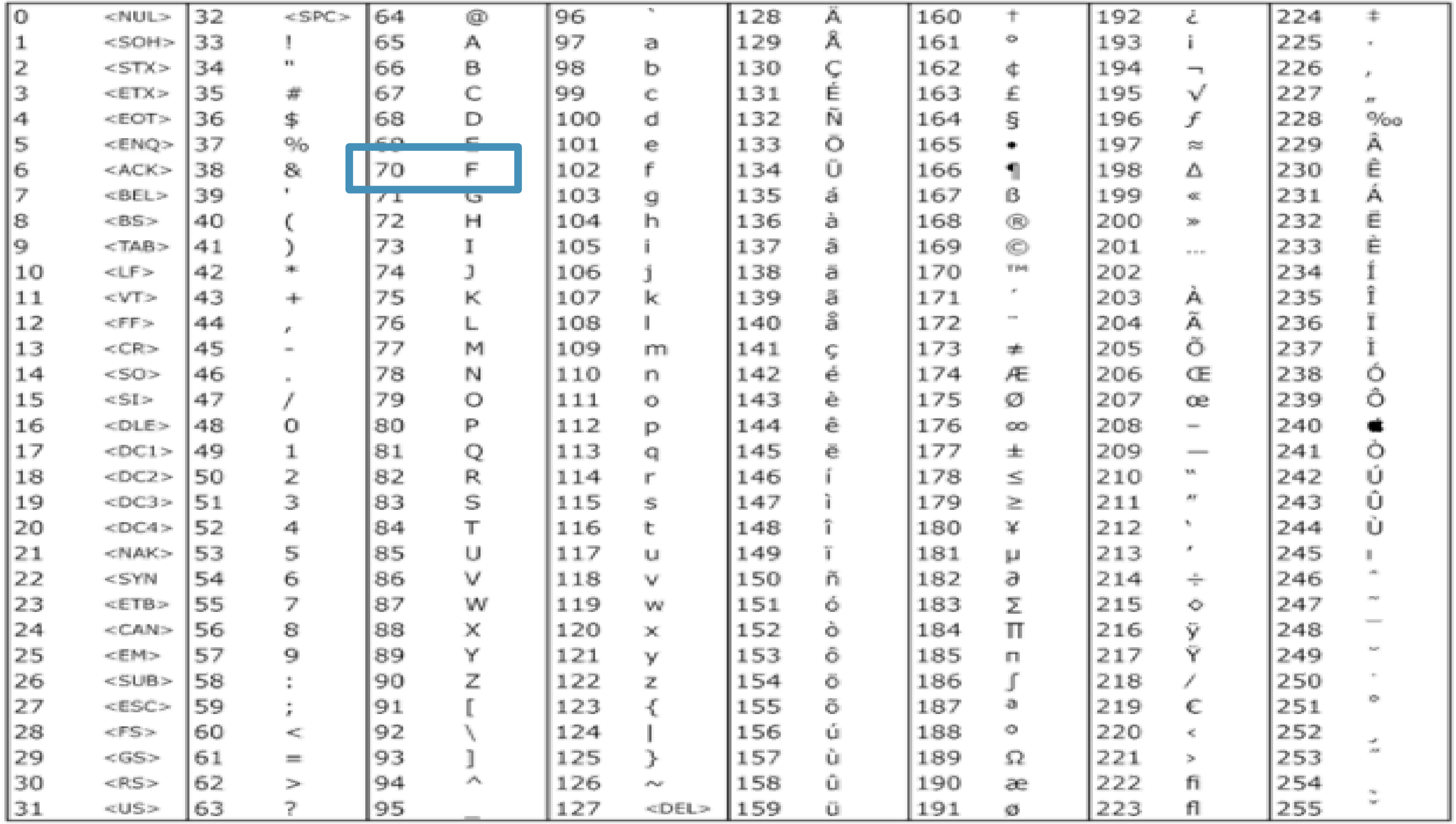
Base quality is ASCII-encoded version of $Q = -10 \log_{10} p$

Usual ASCII encoding is "Phred+33":

take Q, rounded to integer, add 33, convert to character

Slide Credit: Ben Langmead course of Algorithms for DNA Sequencing

ASCII



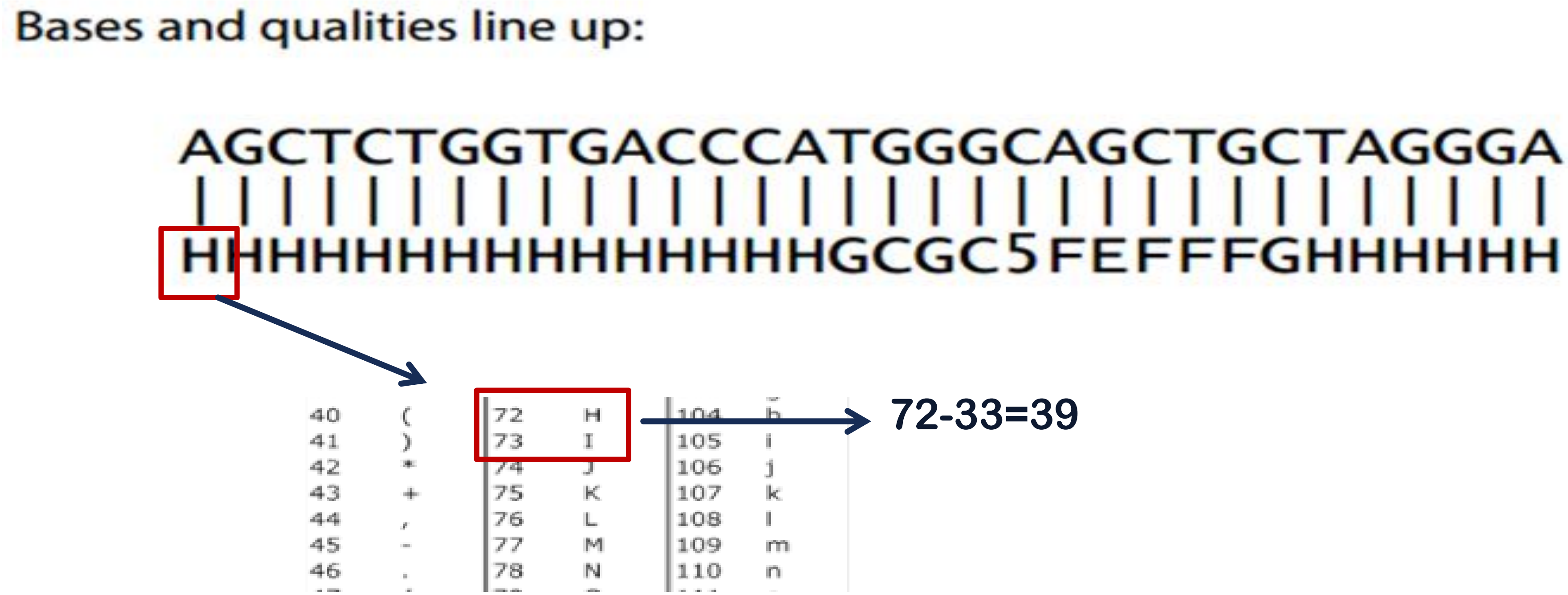
0	<NUL>	32	<SPC>	64	@	96	`	128	Ä	160	†	192	¿	224	‡
1	<SOH>	33	!	65	A	97	a	129	Å	161	°	193	¡	225	·
2	<STX>	34	"	66	B	98	b	130	Ç	162	¢	194	¬	226	‚
3	<ETX>	35	#	67	C	99	c	131	É	163	£	195	√	227	„
4	<EOT>	36	$	68	D	100	d	132	Ñ	164	§	196	ƒ	228	‰
5	<ENQ>	37	%	69	E	101	e	133	Ö	165	•	197	≈	229	Â
6	<ACK>	38	&	70	F	102	f	134	Ü	166	¶	198	∆	230	Ê
7	<BEL>	39	'	71	G	103	g	135	á	167	ß	199	«	231	Á
8	<BS>	40	(	72	H	104	h	136	à	168	®	200	»	232	Ë
9	<TAB>	41	)	73	I	105	i	137	â	169	©	201	…	233	È
10	<LF>	42	*	74	J	106	j	138	ä	170	™	202		234	Í
11	<VT>	43	+	75	K	107	k	139	ã	171	´	203	À	235	Î
12	<FF>	44	,	76	L	108	l	140	å	172	¨	204	Ã	236	Ï
13	<CR>	45	-	77	M	109	m	141	ç	173	≠	205	Õ	237	Ì
14	<SO>	46	.	78	N	110	n	142	é	174	Æ	206	Œ	238	Ó
15	<SI>	47	/	79	O	111	o	143	è	175	Ø	207	œ	239	Ô
16	<DLE>	48	0	80	P	112	p	144	ê	176	∞	208	–	240	
17	<DC1>	49	1	81	Q	113	q	145	ë	177	±	209	—	241	Ò
18	<DC2>	50	2	82	R	114	r	146	í	178	≤	210	“	242	Ú
19	<DC3>	51	3	83	S	115	s	147	ì	179	≥	211	”	243	Û
20	<DC4>	52	4	84	T	116	t	148	î	180	¥	212	‘	244	Ù
21	<NAK>	53	5	85	U	117	u	149	ï	181	µ	213	’	245	ı
22	<SYN	54	6	86	V	118	v	150	ñ	182	∂	214	÷	246	ˆ
23	<ETB>	55	7	87	W	119	w	151	ó	183	∑	215	◊	247	˜
24	<CAN>	56	8	88	X	120	x	152	ò	184	∏	216	ÿ	248	¯
25	<EM>	57	9	89	Y	121	y	153	ô	185	π	217	Ÿ	249	˘
26	<SUB>	58	:	90	Z	122	z	154	ö	186	∫	218	⁄	250	˙
27	<ESC>	59	;	91	[	123	{	155	õ	187	ª	219	€	251	˚
28	<FS>	60	<	92	\	124	|	156	ú	188	º	220	‹	252	¸
29	<GS>	61	=	93	]	125	}	157	ù	189	Ω	221	›	253	˝
30	<RS>	62	>	94	^	126	~	158	û	190	æ	222	ﬁ	254	˛
31	<US>	63	?	95	_	127	<DEL>	159	ü	191	ø	223	ﬂ	255	ˇ

Example: Q=36.7

Phred+33= 37+33=70

= F

BASE QUALITIES



Bases and qualities line up:

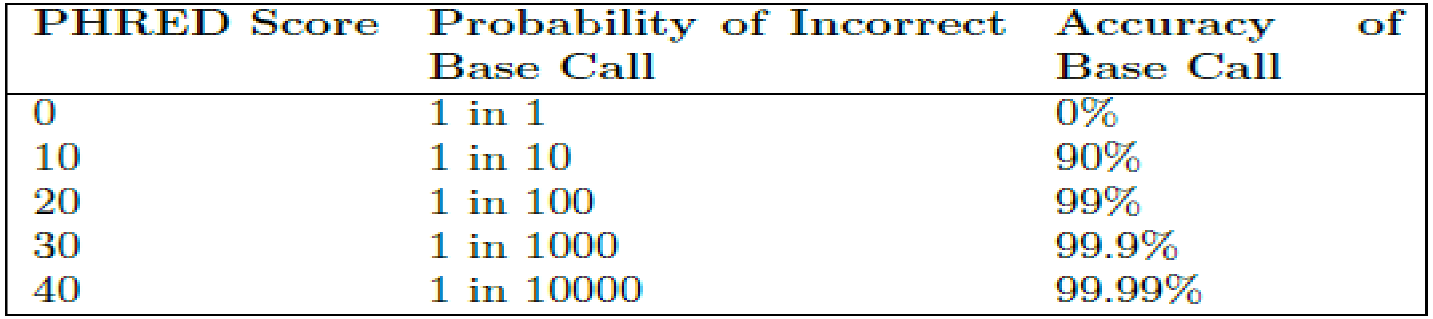
```
AGCTCTGGTGACCCATGGGCAGCTGCTAGGGA
||||||||||||||||||||||||||||||||
HHHHHHHHHHHHHHHGCGC5FEFFFGHHHHHH
```

40	(	72	H	104	h
41	)	73	I	105	i
42	*	74	J	106	j
43	+	75	K	107	k
44	,	76	L	108	l
45	-	77	M	109	m
46	.	78	N	110	n

72-33=39

Slide Credit: Ben Langmead course of Algorithms for DNA Sequencing

GENOMIC DATA
(A READ IN FASTQ)

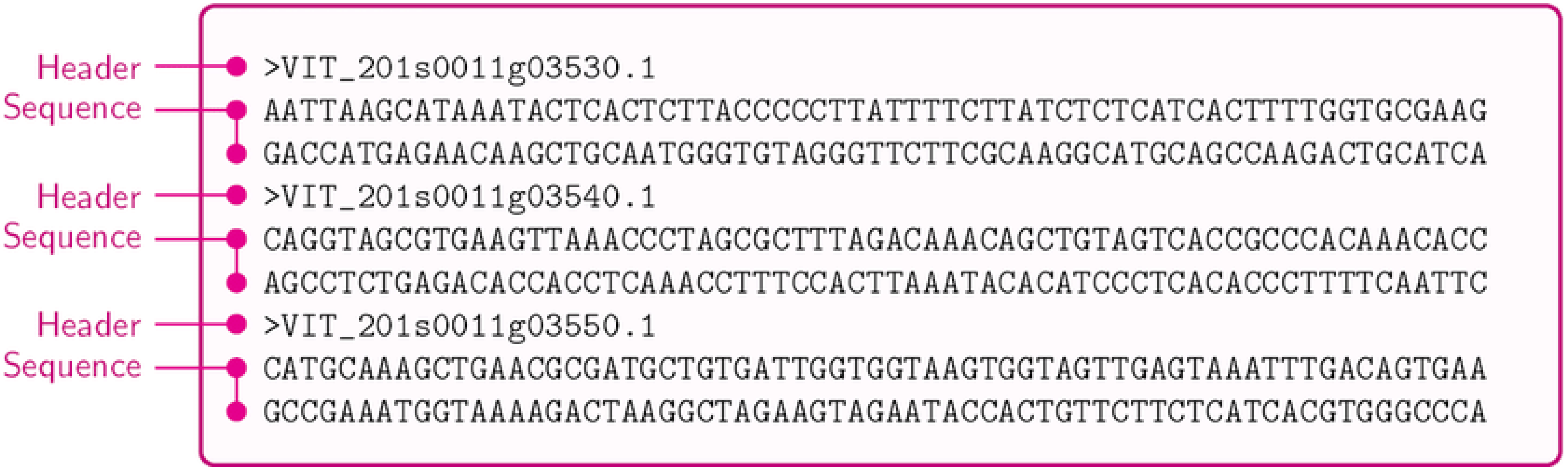


PHRED Score	Probability of Incorrect Base Call	Accuracy of Base Call
0	1 in 1	0%
10	1 in 10	90%
20	1 in 100	99%
30	1 in 1000	99.9%
40	1 in 10000	99.99%

- 10 corresponds to 10% error (1/10),
- 20 corresponds to 1% error (1/100),
- 30 corresponds to 0.1% error (1/1,000) and
- 40 corresponds to one error every 10,000 measurements (1/10,000) that is an error rate of 0.01%.

https://www.youtube.com/playlist?list=PL2mpR0RYFQsBiCWVJSvVAO3OJ2t7DzoHA

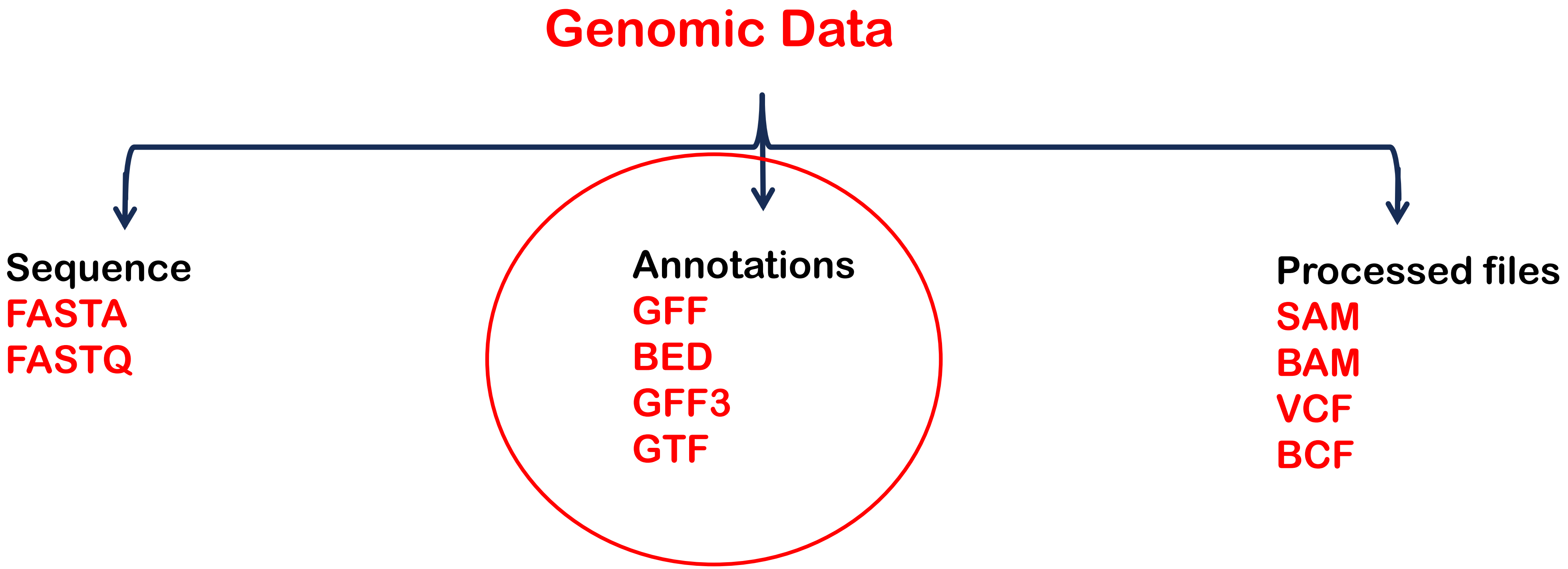
GENOMIC DATA
(A DATA IN FASTA)



```
>VIT_201s0011g03530.1
AATTAAGCATAAATACTCACTCTTACCCCCTTATTTTCTTATCTCTCATCACTTTTGGTGCGAAG
GACCATGAGAACAAGCTGCAATGGGTGTAGGGTTCTTCGCAAGGCATGCAGCCAAGACTGCATCA
>VIT_201s0011g03540.1
CAGGTAGCGTGAAGTTAAACCCTAGCGCTTTAGACAAACAGCTGTAGTCACCGCCCACAAACACC
AGCCTCTGAGACACCACCTCAAACCTTTCCACTTAAATACACATCCCTCACACCCTTTTCAATTC
>VIT_201s0011g03550.1
CATGCAAAGCTGAACGCGATGCTGTGATTGGTGGTAAGTGGTAGTTGAGTAAATTTGACAGTGAA
GCCGAAATGGTAAAAGACTAAGGCTAGAAGTAGAATACCACTGTTCTTCTCATCACGTGGCCCA
```

Slide Credit: Ben Langmead course of Algorithms for DNA Sequencing

GENOMIC DATA



Genomic Data

Sequence
FASTA
FASTQ

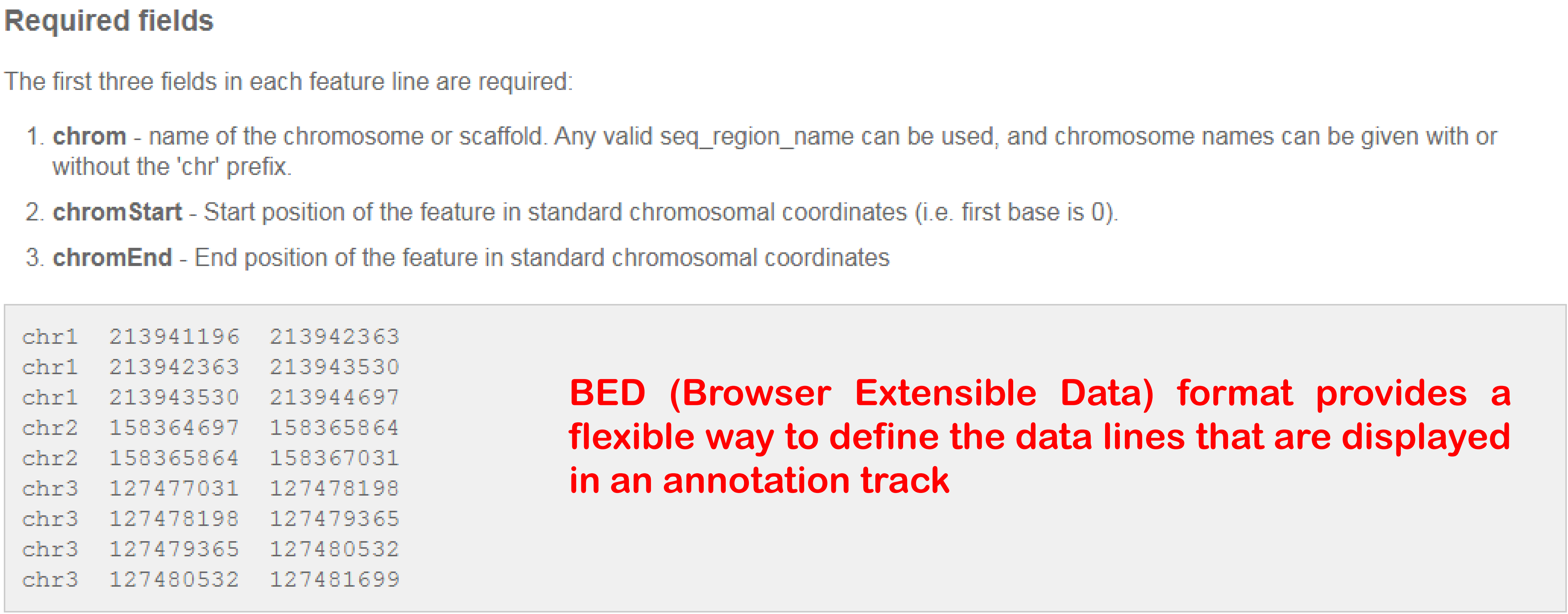
Annotations
GFF
BED
GFF3
GTF

Processed files
SAM
BAM
VCF
BCF

GENOME ANNOTATIONS

- Genome annotation is the process of identifying the locations of genes and all of the coding regions in a genome and determining what those genes do.
- Genome annotation consists of three main steps:
  - identifying portions of the genome that do not code for proteins
  - identifying elements on the genome, a process called gene prediction
  - attaching biological information to these elements
- Descriptions of features – e.g. genes, transcripts, SNPs, start codons – that appear in genomes or transcripts. Annotations typically include coordinates (chromosome name, chromosome positions, and a chromosome strand), as well as properties (gene name, function, GO terms, et c) of a given feature.

GENOMIC DATA (BED FORMAT)



Required fields

The first three fields in each feature line are required:

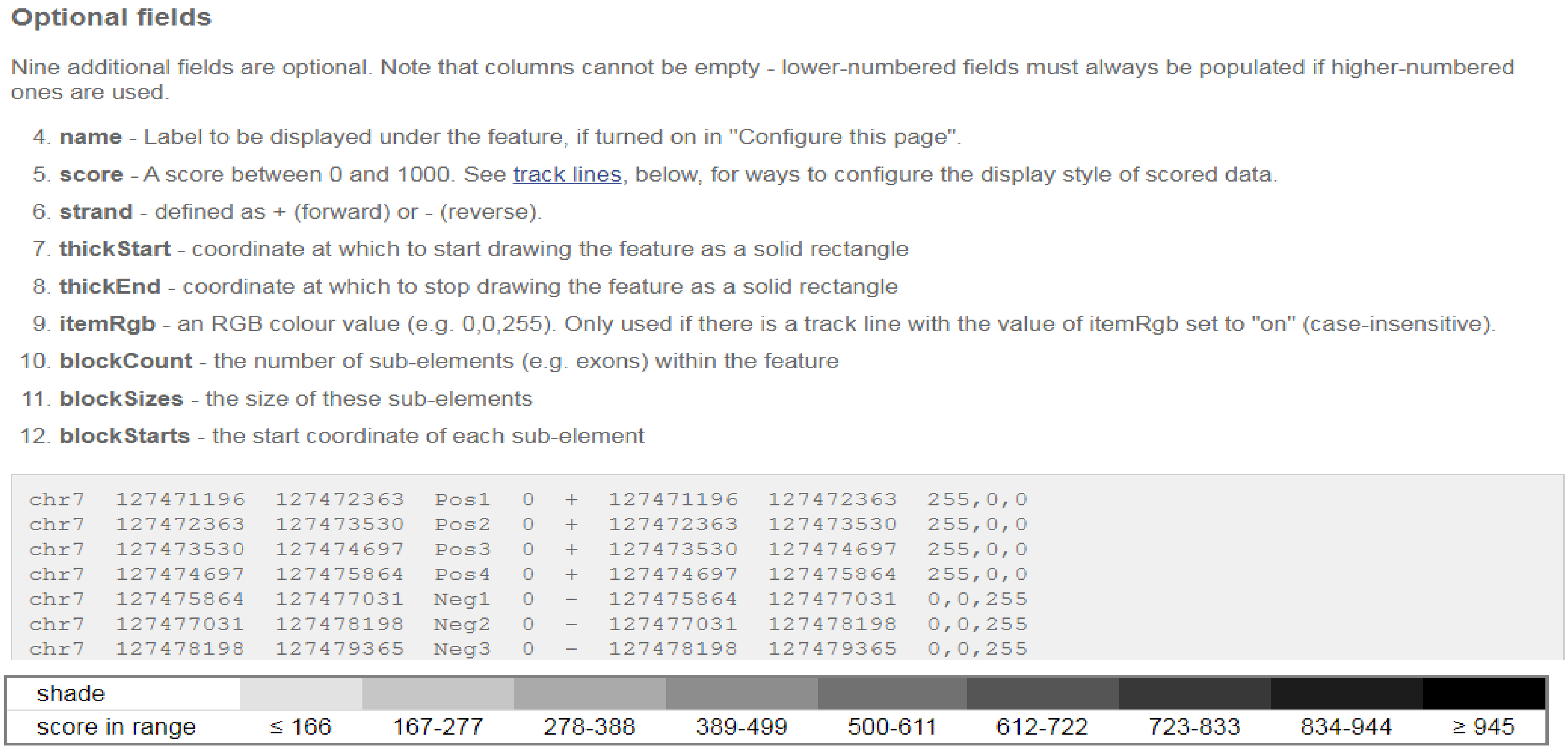
1. **chrom** - name of the chromosome or scaffold. Any valid seq_region_name can be used, and chromosome names can be given with or without the 'chr' prefix.
2. **chromStart** - Start position of the feature in standard chromosomal coordinates (i.e. first base is 0).
3. **chromEnd** - End position of the feature in standard chromosomal coordinates

```
chr1  213941196  213942363
chr1  213942363  213943530
chr1  213943530  213944697
chr2  158364697  158365864
chr2  158365864  158367031
chr3  127477031  127478198
chr3  127478198  127479365
chr3  127479365  127480532
chr3  127480532  127481699
```

BED (Browser Extensible Data) format provides a flexible way to define the data lines that are displayed in an annotation track

Image credit: http://www.ensembl.org/info/website/upload/bed.html#tracklines

GENOMIC DATA
(BED FORMAT)



Optional fields

Nine additional fields are optional. Note that columns cannot be empty - lower-numbered fields must always be populated if higher-numbered ones are used.

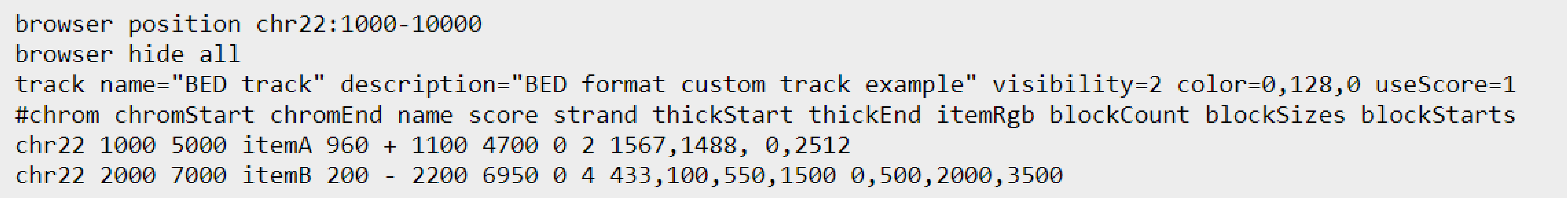
4. **name** - Label to be displayed under the feature, if turned on in "Configure this page".
5. **score** - A score between 0 and 1000. See track lines, below, for ways to configure the display style of scored data.
6. **strand** - defined as + (forward) or - (reverse).
7. **thickStart** - coordinate at which to start drawing the feature as a solid rectangle
8. **thickEnd** - coordinate at which to stop drawing the feature as a solid rectangle
9. **itemRgb** - an RGB colour value (e.g. 0,0,255). Only used if there is a track line with the value of itemRgb set to "on" (case-insensitive).
10. **blockCount** - the number of sub-elements (e.g. exons) within the feature
11. **blockSizes** - the size of these sub-elements
12. **blockStarts** - the start coordinate of each sub-element

```
chr7  127471196  127472363  Pos1  0  +  127471196  127472363  255,0,0
chr7  127472363  127473530  Pos2  0  +  127472363  127473530  255,0,0
chr7  127473530  127474697  Pos3  0  +  127473530  127474697  255,0,0
chr7  127474697  127475864  Pos4  0  +  127474697  127475864  255,0,0
chr7  127475864  127477031  Neg1  0  -  127475864  127477031  0,0,255
chr7  127477031  127478198  Neg2  0  -  127477031  127478198  0,0,255
chr7  127478198  127479365  Neg3  0  -  127478198  127479365  0,0,255
```

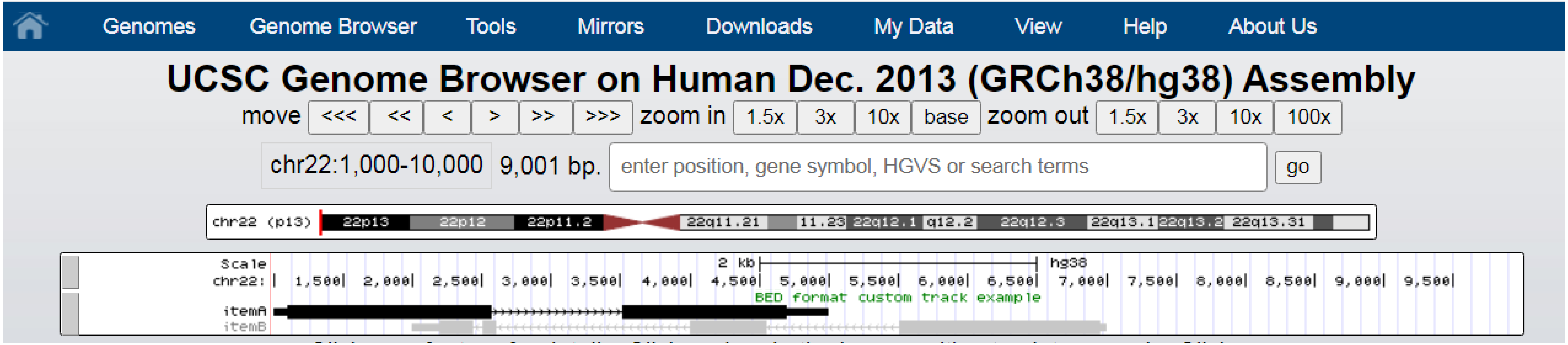
shade	░	░	▒	▒	▒	▓	▓	█	█
score in range	≤ 166	167-277	278-388	389-499	500-611	612-722	723-833	834-944	≥ 945

Image credit: http://www.ensembl.org/info/website/upload/bed.html#tracklines

GENOMIC DATA (BED FORMAT)



```
browser position chr22:1000-10000
browser hide all
track name="BED track" description="BED format custom track example" visibility=2 color=0,128,0 useScore=1
#chrom chromStart chromEnd name score strand thickStart thickEnd itemRgb blockCount blockSizes blockStarts
chr22 1000 5000 itemA 960 + 1100 4700 0 2 1567,1488, 0,2512
chr22 2000 7000 itemB 200 - 2200 6950 0 4 433,100,550,1500 0,500,2000,3500
```



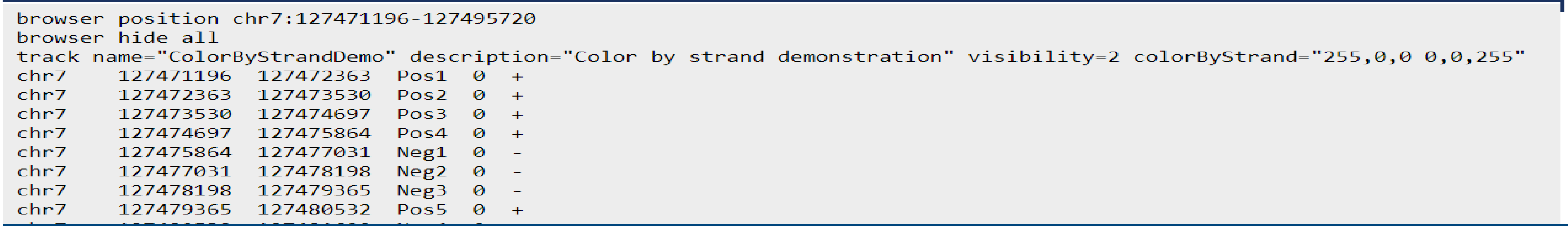
UCSC Genome Browser on Human Dec. 2013 (GRCh38/hg38) Assembly

chr22:1,000-10,000 9,001 bp.

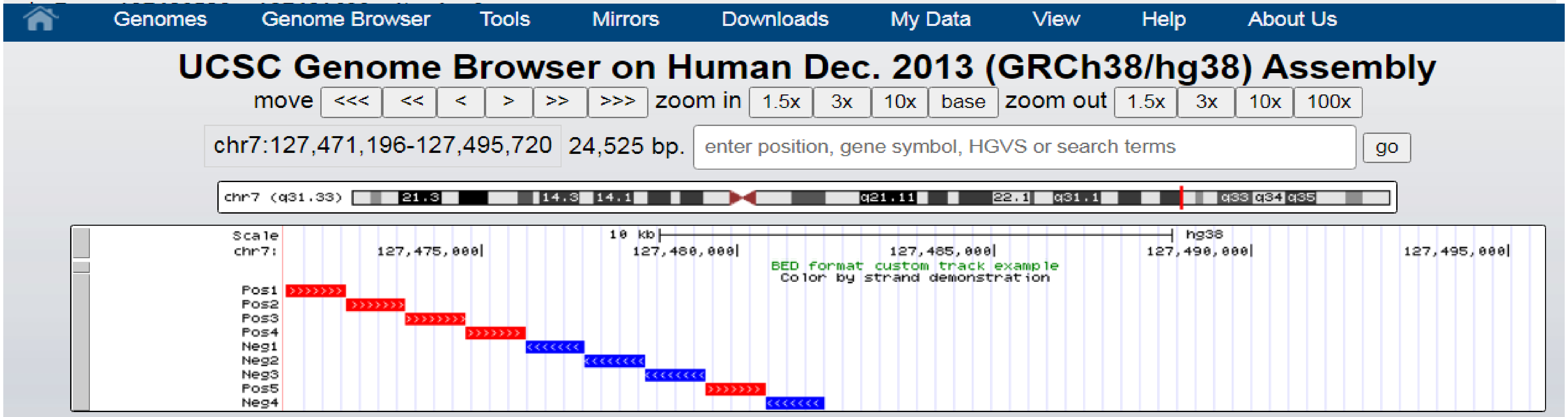
Genomes
Genome Browser
Tools
Mirrors
Downloads
My Data
View
Help
About Us
move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x 100x
enter position, gene symbol, HGVS or search terms
go
chr22 (p13) 22p13 22p12 22p11.2 22q11.21 11.23 22q12.1 q12.2 22q12.3 22q13.1 22q13.2 22q13.31
Scale
2 kb
hg38
chr22: 1,500 2,000 2,500 3,000 3,500 4,000 4,500 5,000 5,500 6,000 6,500 7,000 7,500 8,000 8,500 9,000 9,500
BED format custom track example
itemA
itemB

Image credit: https://genome.ucsc.edu/goldenPath/help/customTrack.html

GENOMIC DATA (BED FORMAT)



```
browser position chr7:127471196-127495720
browser hide all
track name="ColorByStrandDemo" description="Color by strand demonstration" visibility=2 colorByStrand="255,0,0 0,0,255"
chr7    127471196  127472363  Pos1  0  +
chr7    127472363  127473530  Pos2  0  +
chr7    127473530  127474697  Pos3  0  +
chr7    127474697  127475864  Pos4  0  +
chr7    127475864  127477031  Neg1  0  -
chr7    127477031  127478198  Neg2  0  -
chr7    127478198  127479365  Neg3  0  -
chr7    127479365  127480532  Pos5  0  +
```

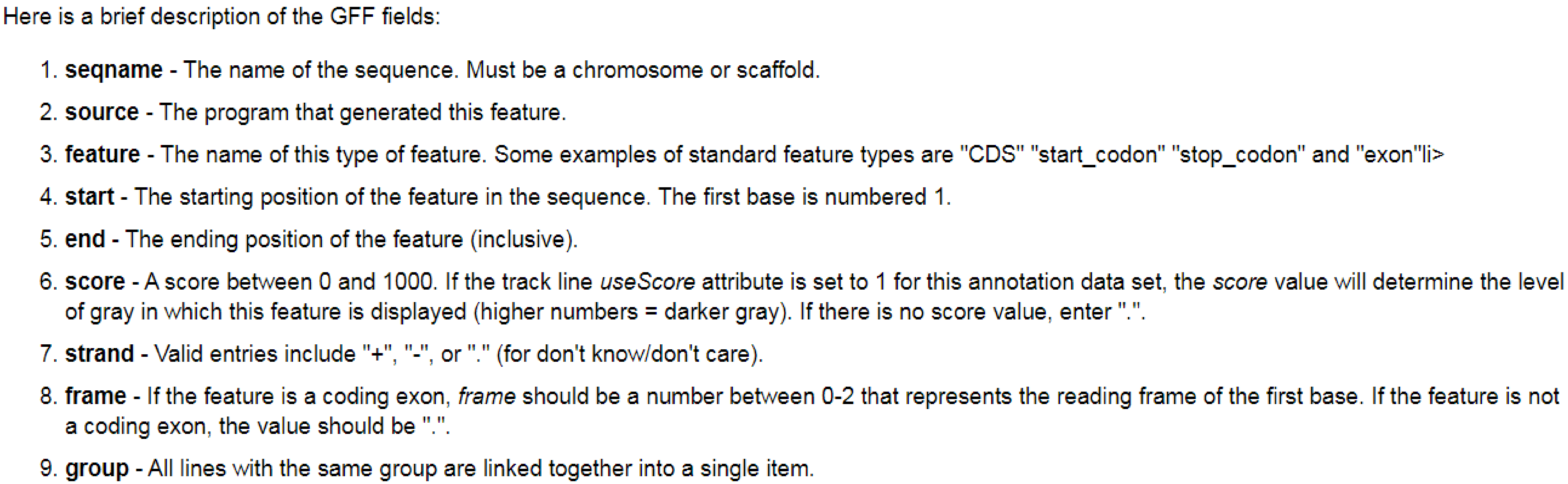


UCSC Genome Browser on Human Dec. 2013 (GRCh38/hg38) Assembly

chr7:127,471,196-127,495,720 24,525 bp.

Image credit: https://genome.ucsc.edu/FAQ/FAQformat.html

GENOMIC DATA
(GFF FORMAT)



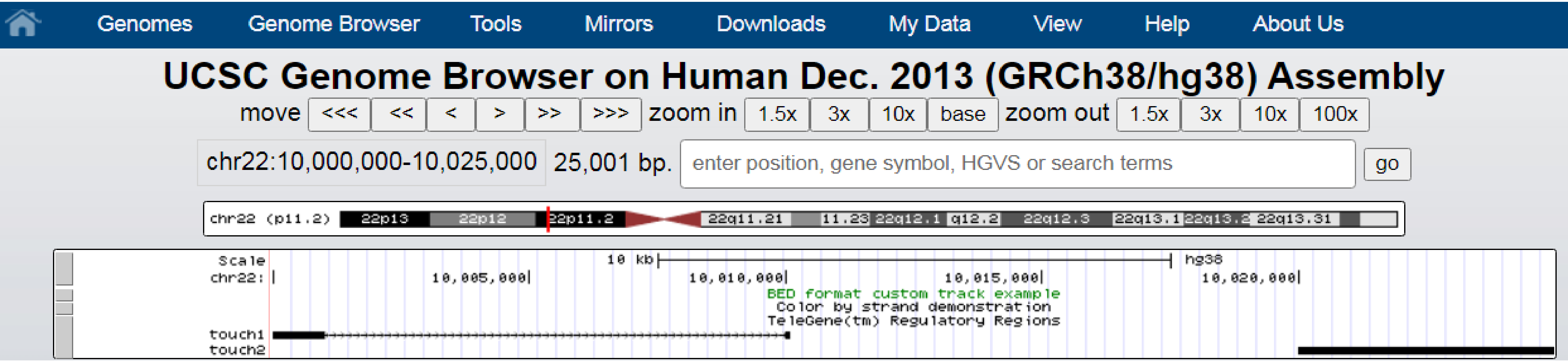
Here is a brief description of the GFF fields:

1. **seqname** - The name of the sequence. Must be a chromosome or scaffold.
2. **source** - The program that generated this feature.
3. **feature** - The name of this type of feature. Some examples of standard feature types are "CDS" "start_codon" "stop_codon" and "exon"li>
4. **start** - The starting position of the feature in the sequence. The first base is numbered 1.
5. **end** - The ending position of the feature (inclusive).
6. **score** - A score between 0 and 1000. If the track line *useScore* attribute is set to 1 for this annotation data set, the *score* value will determine the level of gray in which this feature is displayed (higher numbers = darker gray). If there is no score value, enter ".".
7. **strand** - Valid entries include "+", "-", or "." (for don't know/don't care).
8. **frame** - If the feature is a coding exon, *frame* should be a number between 0-2 that represents the reading frame of the first base. If the feature is not a coding exon, the value should be ".".
9. **group** - All lines with the same group are linked together into a single item.

Slides credit: https://genome.ucsc.edu/FAQ/FAQformat.html#format3

GENOMIC DATA
(GFF FORMAT)

```
browser position chr22:10000000-10025000
browser hide all
track name=regulatory description="TeleGene(tm) Regulatory Regions" visibility=2
chr22   TeleGene        enhancer        10000000        10001000        500     +       .       touch1
chr22   TeleGene        promoter        10010000        10010100        900     +       .       touch1
chr22   TeleGene        promoter        10020000        10025000        800     -       .       touch2
```

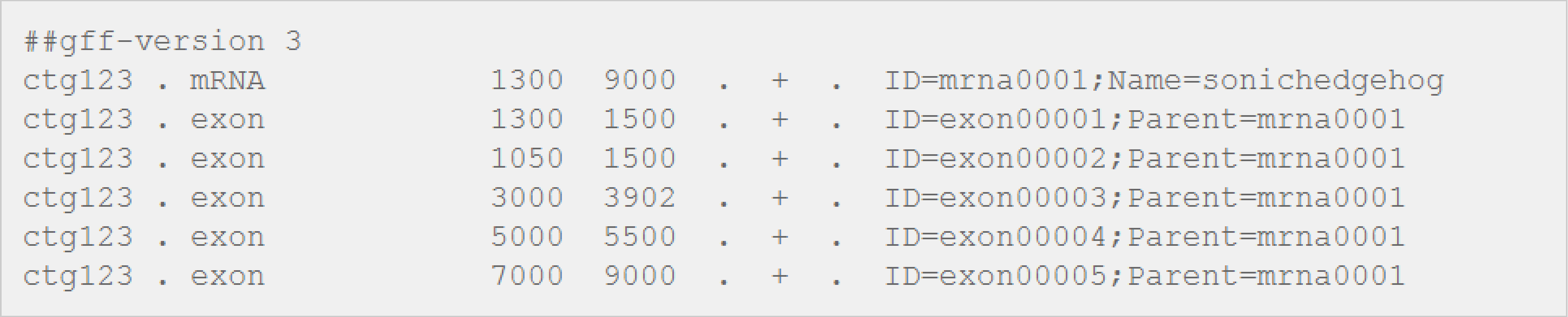


UCSC Genome Browser on Human Dec. 2013 (GRCh38/hg38) Assembly

chr22:10,000,000-10,025,000 25,001 bp.

BED format custom track example
Color by strand demonstration
TeleGene(tm) Regulatory Regions

GENOMIC DATA (GFF3 FORMAT)



```
##gff-version 3
ctg123 . mRNA            1300  9000  .  +  .  ID=mrna0001;Name=sonichedgehog
ctg123 . exon            1300  1500  .  +  .  ID=exon00001;Parent=mrna0001
ctg123 . exon            1050  1500  .  +  .  ID=exon00002;Parent=mrna0001
ctg123 . exon            3000  3902  .  +  .  ID=exon00003;Parent=mrna0001
ctg123 . exon            5000  5500  .  +  .  ID=exon00004;Parent=mrna0001
ctg123 . exon            7000  9000  .  +  .  ID=exon00005;Parent=mrna0001
```

Slides credit: http://www.ensembl.org/info/website/upload/gff3.html

Thank you!