

Plataformas de secuenciación

Primera generación

Sanger or chain-termination

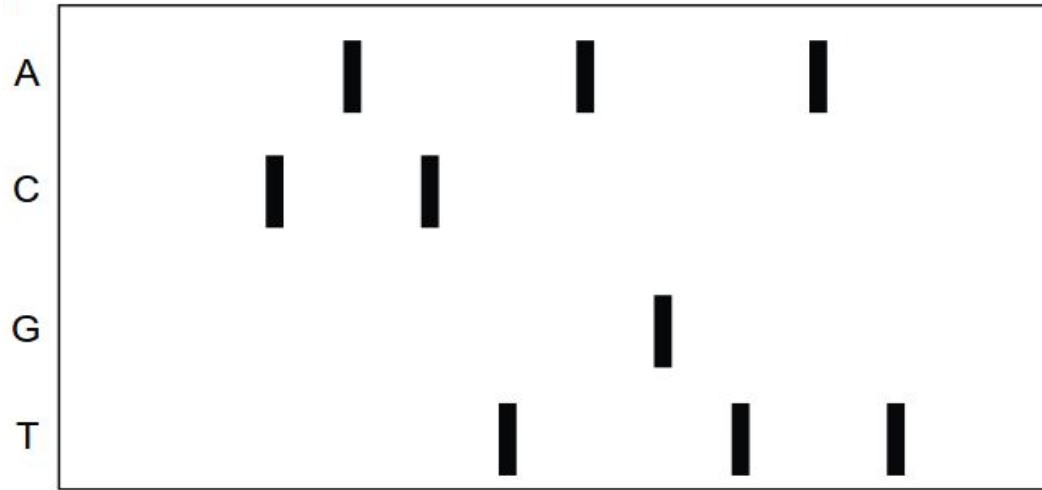
Produce cadenas de 100 - 1000 pb

Problemas cuando hay más de una variante alélica en la misma posición

Usos actuales:

- Validación de variantes
- Brindar cobertura a regiones poco representadas por técnicas NGS

Sanger or chain-termination



Segunda generación

Next Generation Sequencing High Throughput Sequencing

Sequencing by synthesis (Illumina)

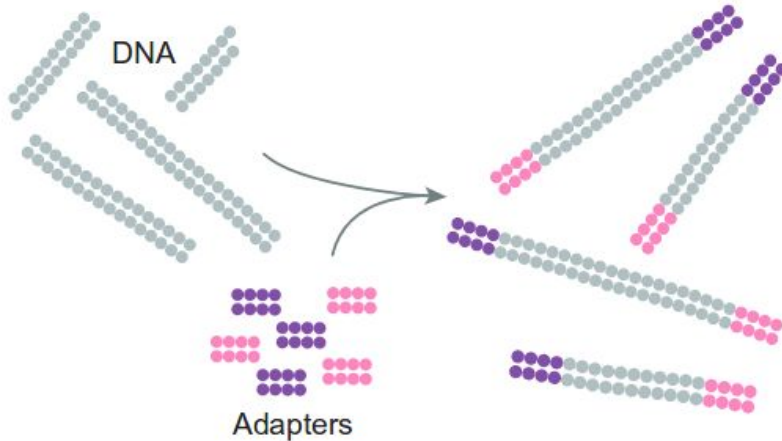
Nucleótidos de terminación de cadena marcados fluorescentemente reversibles

Muestras: sangre, médula ósea, tejido fresco o tejido embebido en parafina fijado con formalina (FFPE)

Es importante una extracción de ADN con buena calidad para luego armar la librería

Sequencing by synthesis (Illumina)

(A)



Prepare genomic DNA sample

Randomly fragment genomic DNA and ligate adapters to both ends of the fragments.

Fragmentado:

- sonicación
- digestión enzimática con DNAsa I o fragmentasa

Etiquetado (adaptadores, etiquetas, códigos de barras y cebadores):

- Ligación
- PCR

Fragmentado + Etiquetado (Nextera Tagmentation):

- enzima transposasa para fragmentar e insertar secuencias simultáneamente

Library

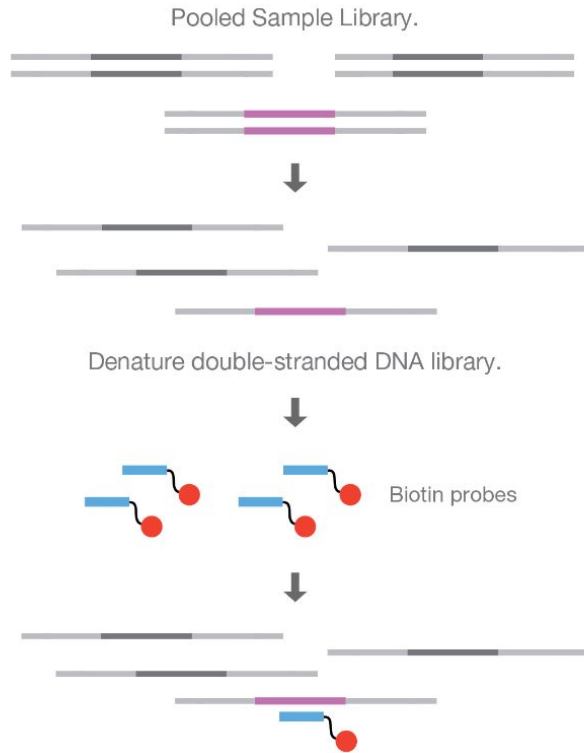
Target sequencing:

- Exomas
- Paneles de genes de interés
- Genotipado

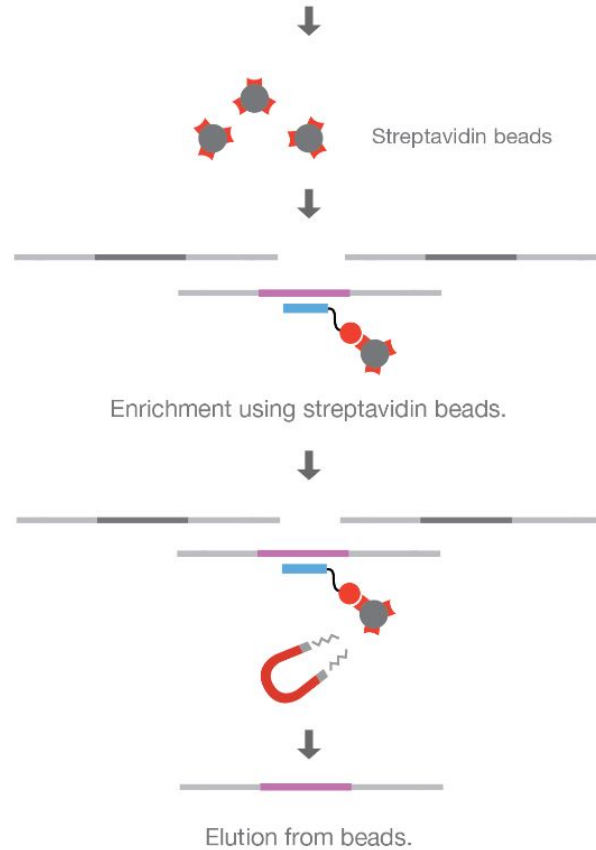
Método de captura

- Amplificación por PCR
- Hibridación con oligonucleótidos
- Selección de poliA (ARNm)

Target enrichment



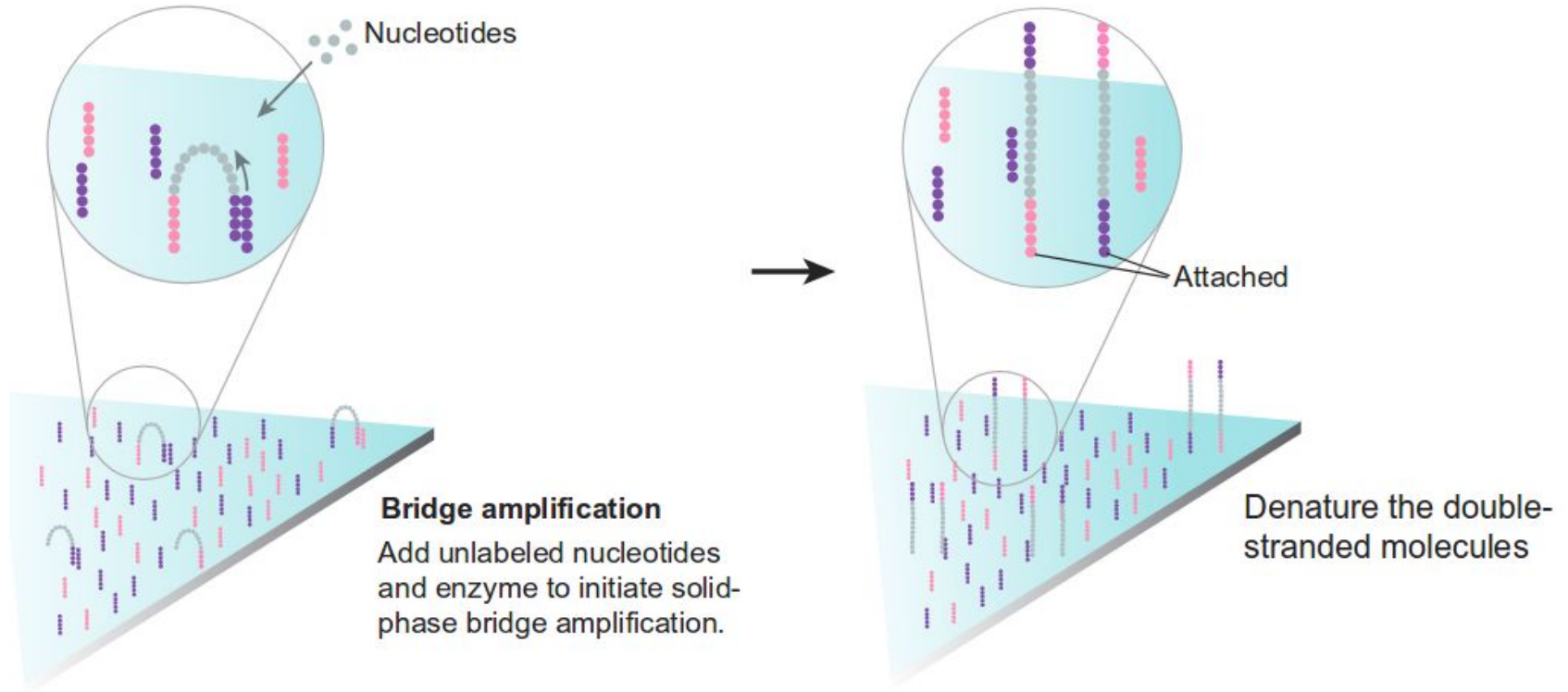
Hybridize biotinylated probes to targeted regions.



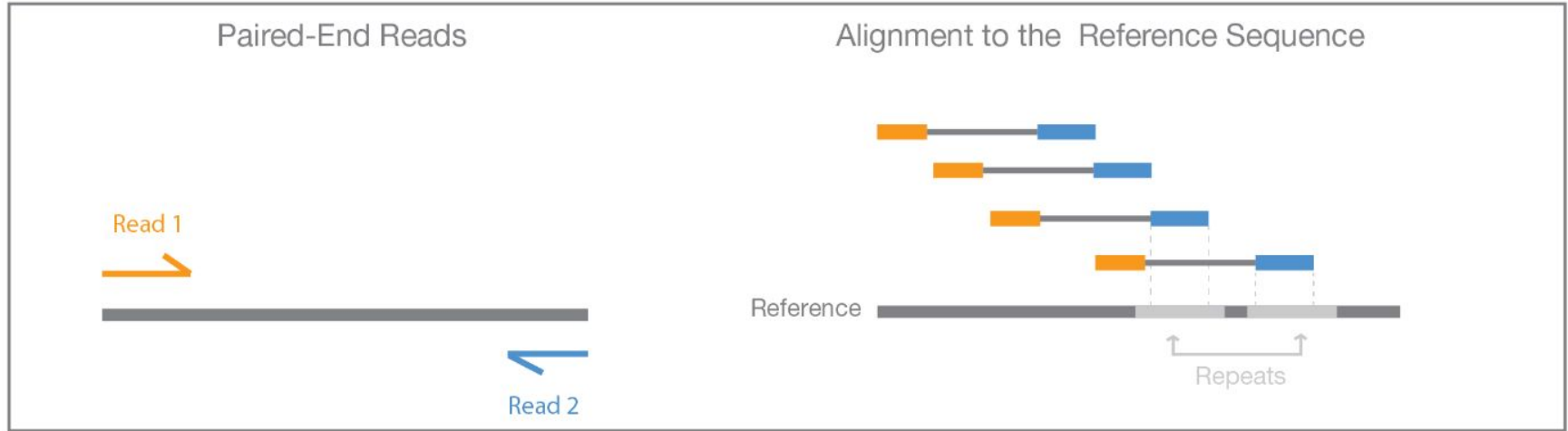
Library - PCR limitations

- Solo amplifica regiones objetivo sin mutaciones
- Sesgo de PCR causado por diferente longitud de fragmentos, variaciones aleatorias en la cantidad de fragmentos o en la eficiencia por ciclo
- Productos espurios de PCR

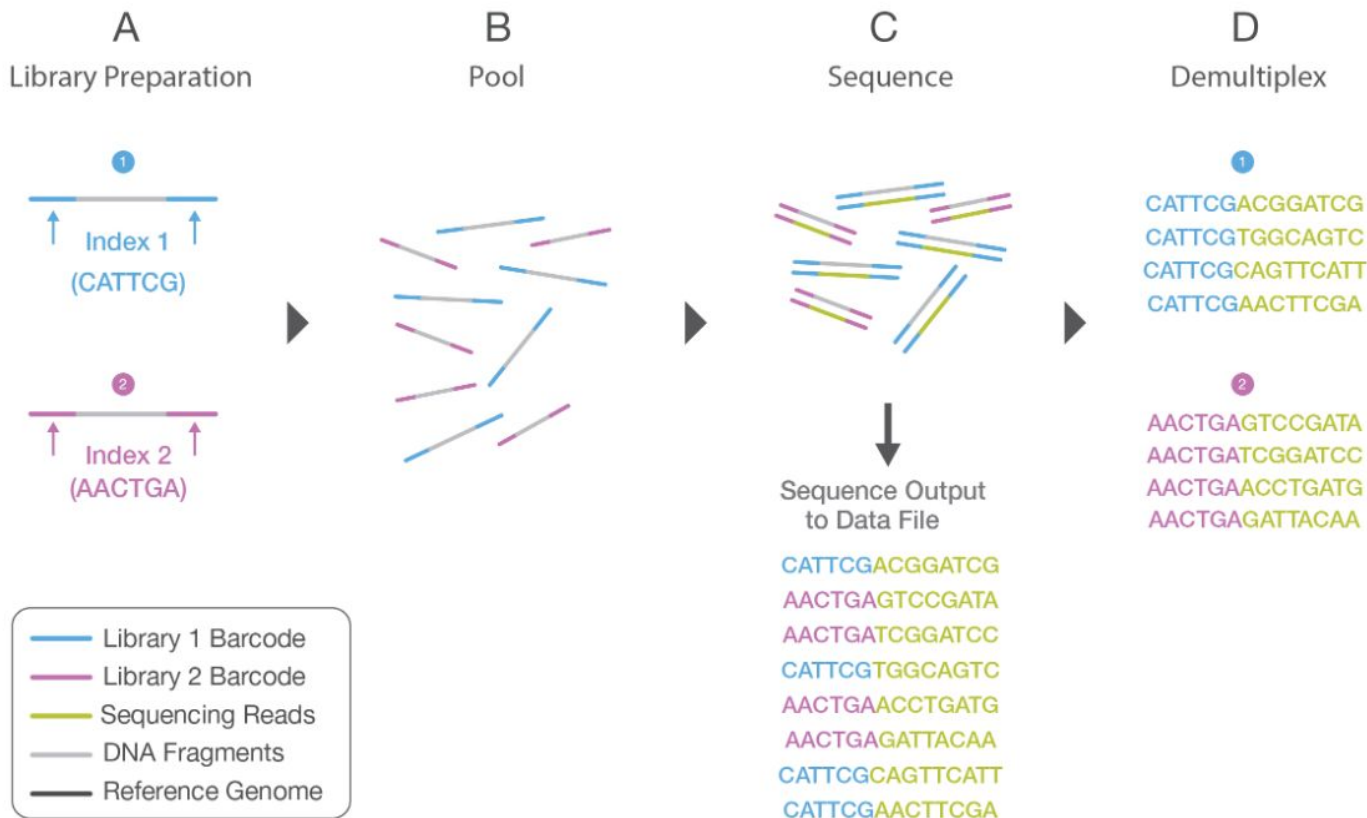
Sequencing by synthesis (Illumina)



Paired-End Sequencing

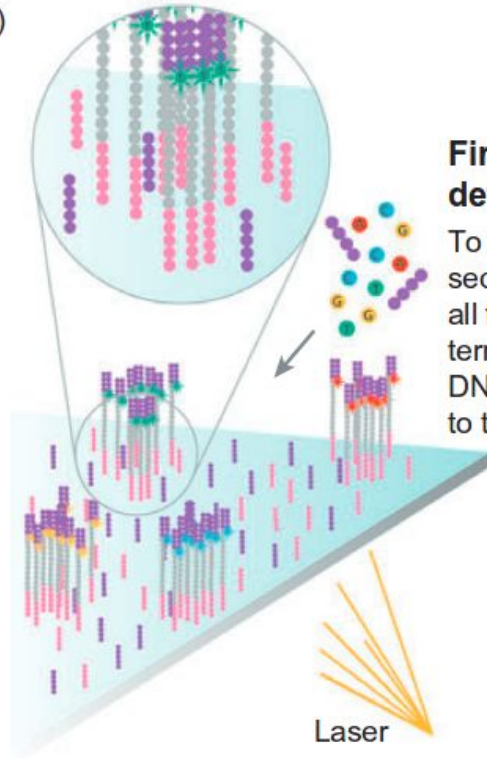


Multiplex



Sequencing by synthesis (Illumina)

(B)



First chemistry cycle: determine first base

To initiate the first sequencing cycle, add all four labeled reversible terminators, primers, and DNA polymerase enzyme to the flow cell.

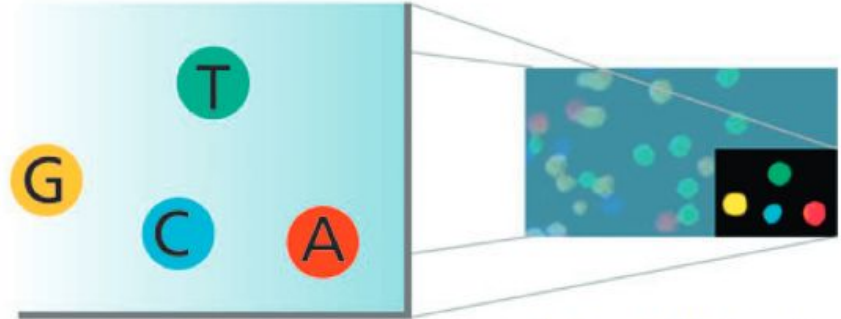


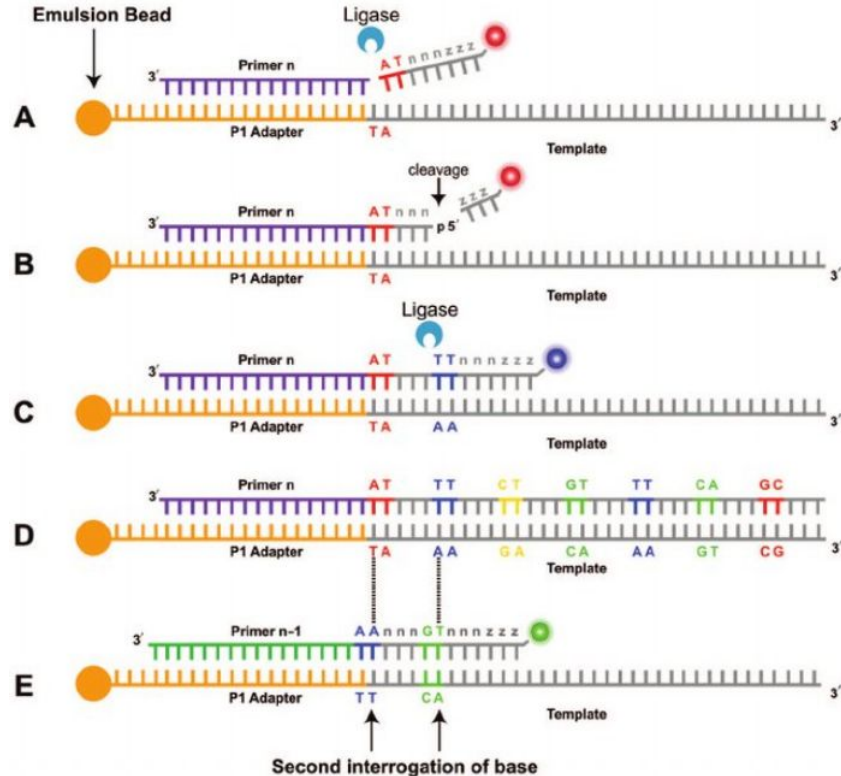
Image of first chemistry cycle

After laser excitation, capture the image of emitted fluorescence from each cluster on the flow cell. Record the identity of the first base for each cluster.

Before initiating the next chemistry cycle

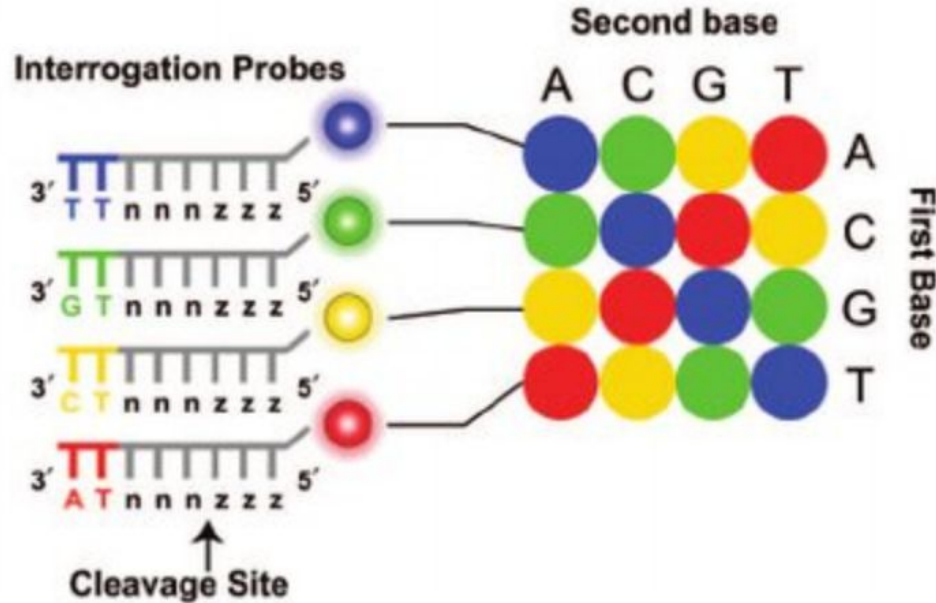
The blocked 3' terminus and the fluorophore from each incorporated base are removed.

Sequencing by Oligonucleotide Ligation and Detection (SOLiD)



Voelkerding, K. V., Dames, S. A., & Durtschi, J. D. (2009). Next-generation sequencing: from basic research to diagnostics. *Clinical chemistry*, 55(4), 641-658.

Sequencing by Oligonucleotide Ligation and Detection (SOLiD)



Pyrosequencing (ROCHE 454)

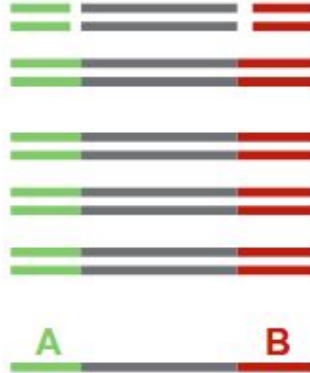
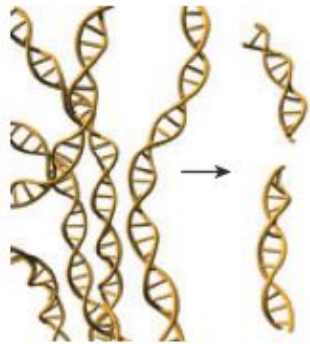
Cada nucleótido incorporado causa la liberación de un pirofosfato que se combina con fosfosulfato de adenosina para formar ATP en una reacción catalizada por ATP sulfurilasa.

En presencia de ATP, la luciferasa convierte luciferina en oxiluciferina y se emite luz, lo que da como resultado una señal.

Pyrosequencing

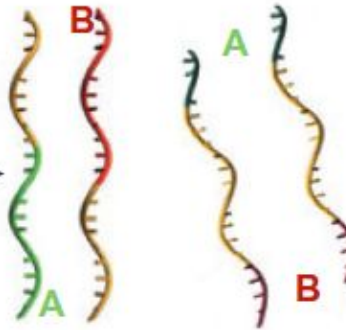
DNA library preparation

4.5 h



Ligation

Selection
(isolate AB
fragments
only)



- Genome fragmented by nebulization
- No cloning; no colony picking
- sstDNA library created with adaptors
- A/B fragments selected using avidin-biotin purification

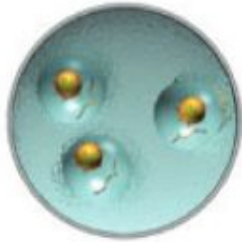
gDNA

sstDNA library

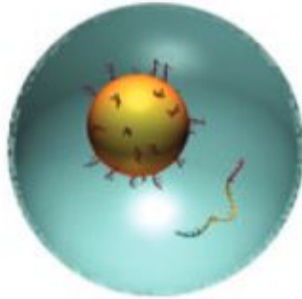
Pyrosequencing

Emulsion PCR

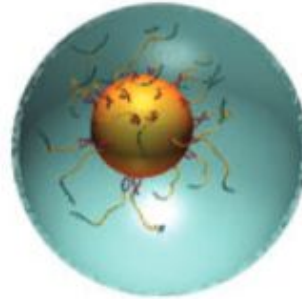
8 h



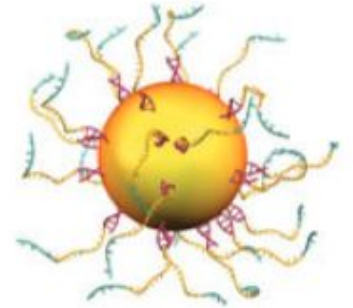
Anneal sstDNA to an excess of DNA capture beads



Emulsify beads and PCR reagents in water-in-oil microreactors



Clonal amplification occurs inside microreactors



Break microreactors and enrich for DNA-positive beads

sstDNA library

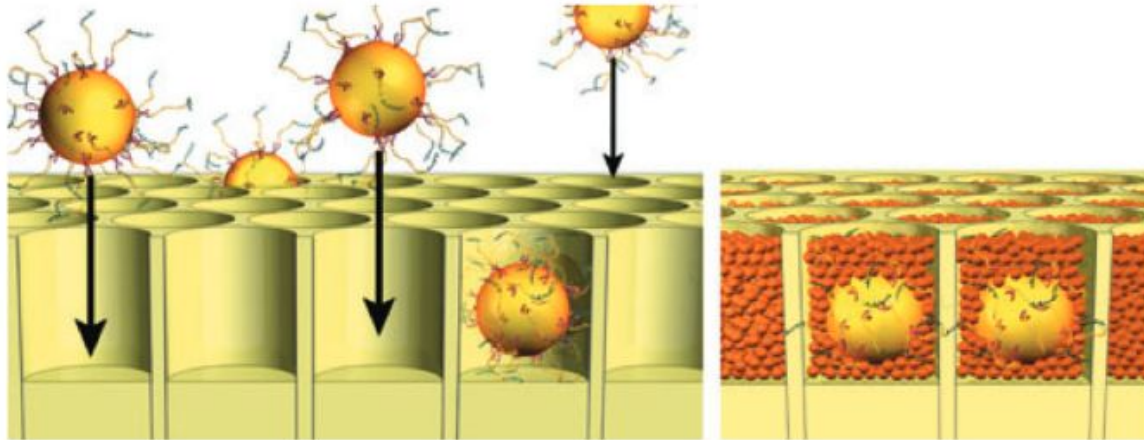


Bead-amplified sstDNA library

Pyrosequencing

Sequencing

7.5 h



- Well diameter: average of 44 μm
- 400,000 reads obtained in parallel
- A single cloned amplified sstDNA bead is deposited per well

Amplified sstDNA library beads



Quality filtered bases

Ion Torrent

Detecta cambios pequeños en el pH que ocurren como resultado de la liberación de H cuando los nucleótidos se incorporan en una secuencia de alargamiento

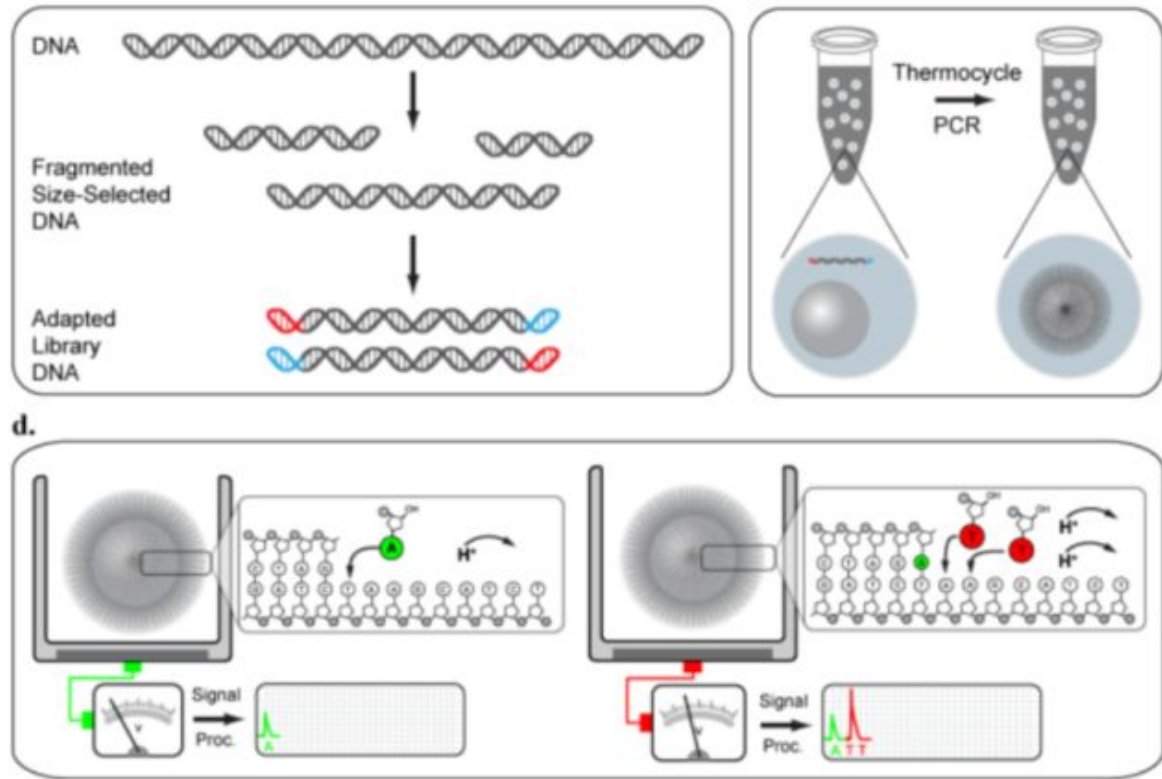
Se incorpora un tipo de desoxinucleótido (dA, dC, dG, dT) a la vez

La reacción de secuenciación es análoga a la pirosecuenciación, con la diferencia de que se detecta pH en lugar de luz.

Las lecturas son relativamente largas, y los tiempos de reacción son muy cortos.

Alta tasa de error

Ion Torrent



http://www.genomics.cn/en/navigation/show_navigation?nid=2640

NGS platforms/company/max output per run	Read length per run (bp)	No. reads per run	Time (h or days)	Cost per 10 ⁶ bases	Raw error rate (%)	Platform cost (USD approx.)	Chemistry
<i>First generation</i>							
Sanger/Life Technologies/84 kb	800	1	2 h	2400	0.3	95,000	Dideoxy terminator
<i>Second generation</i>							
454 GS FLX+/Roche/0.7 Gb	700	1×10 ⁶	24/48 h	10	1	500,000	Pyrosequencing
GS Junior/Roche/70 Mb	500	1×10 ⁵	18 h	9		100,000	Pyrosequencing
HiSeq/Illumina/1500 Gb	2×150	5×10 ⁹	27/240 h	0.1	0.8	750,000	Reversible terminators
MiSeq/Illumina/15 Gb	2×300	3×10 ⁸	27 h	0.13	0.8	125,000	Reversible terminators
SOLiD/Life Technologies/120 Gb	50	1×10 ⁹	14 days	0.13	0.01	350,000	Ligation
Retrovolocity/BGI/3000 Gb	50	1×10 ⁹	14 days	0.01	0.01	12×10 ⁶	Nanoball/ligation

Tercera generación

Los métodos de tercera generación secuencian moléculas individuales, sin la necesidad de un paso de enriquecimiento o amplificación.

La principal desventaja de los métodos de una sola molécula es que requieren la capacidad de detectar señales pequeñas, sin comprometer la precisión.

Aplicaciones

Genómica

- Whole Genome Sequencing
- Whole Exome Sequencing
- De novo Sequencing
- Targeted Sequencing

RNAseq

Pasos para secuenciar ARN

- Aislar el ARN
- Convertir el ARN en cDNA con retrotranscripción
- Fragmentar el cDNA y construir la librería
- Secuenciar

Transcriptómica

Es posible secuenciar:

- ARN total
- ARNm
- Target
- Small and non-coding

Esta técnica permite cuantificar el nivel de expresión de cada ARN en el momento de la extracción.

Es importante tener una buena cobertura para evitar que sólo se secuencien ARN naturalmente abundantes como el ARNr.

T-cell receptor repertoire analysis

Decisiones a tomar de acuerdo al experimento

Tipo de material genético

- DNA

- Mayor estabilidad
- Una sola copia por célula -> mejor cuantificación

- RNA

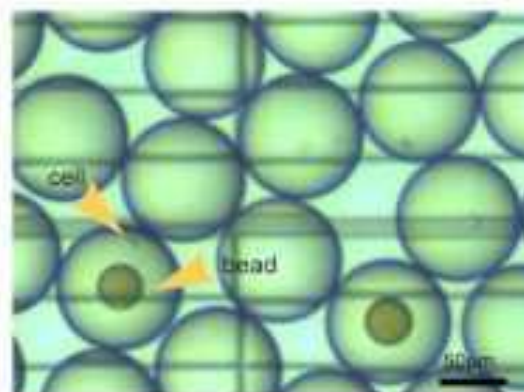
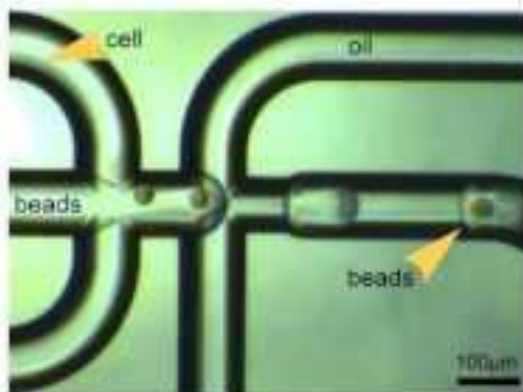
- Información sobre el nivel de expresión
- No hay errores relacionados con intrones o rearrreglos

Sequencing methods

- Bulk sequencing of pooled immune cell populations
 - Más barato
- Single cell sequencing (identify pairs of alpha and beta chains)
 - 10X Chromium 3' -> microdroplet-based
 - Smart-seq2
 - selección manual de células con pipetas micro capilares
 - transcripción inversa de alta fidelidad



Single cell RNA-Seq



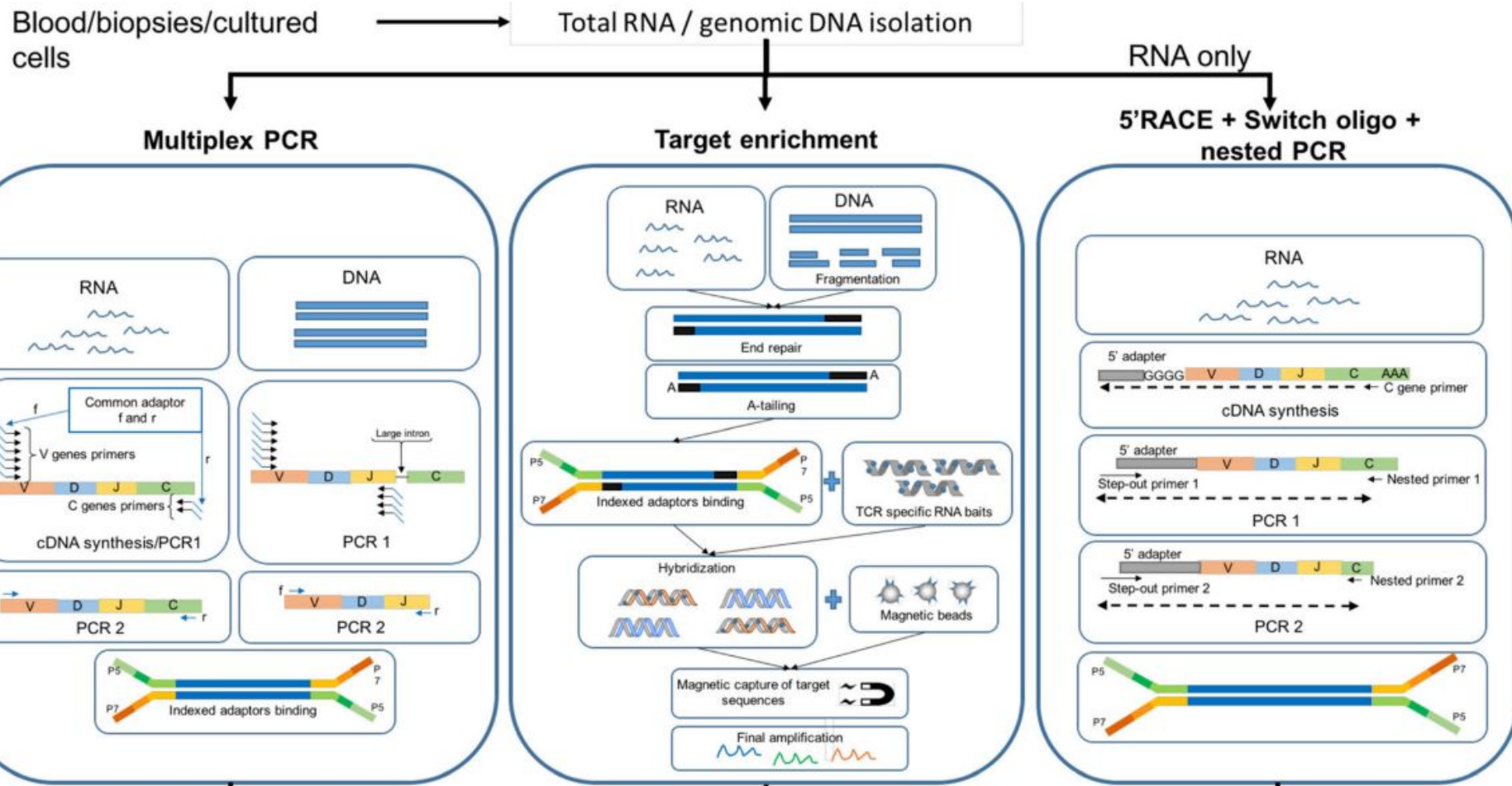
Sequencing depth

- Low coverage (MiSeq)
 - Los análisis orientados a la enfermedad buscan TCR altamente expresados y clonalmente expandidos
- High coverage (HiSeq)

Target sequence

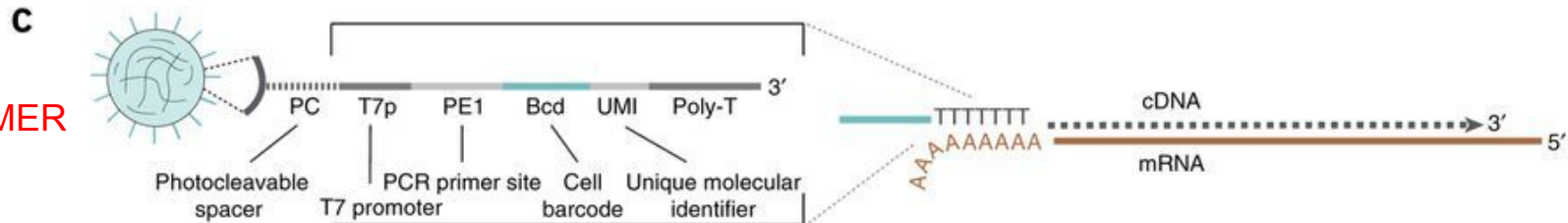
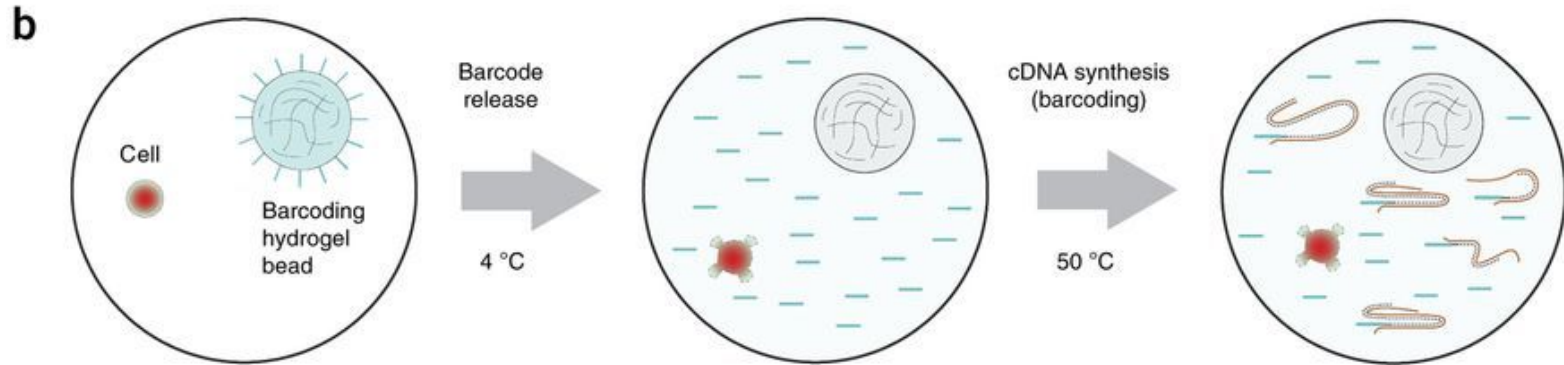
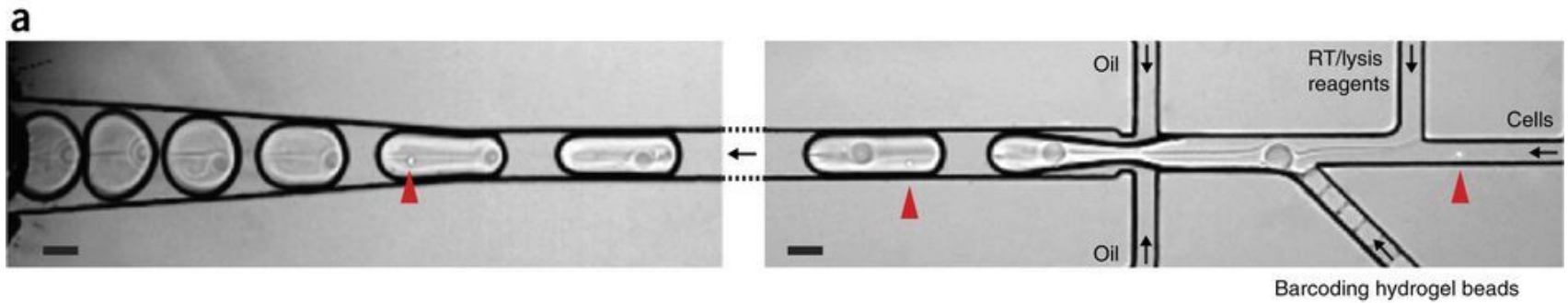
- $\alpha\beta$ and $\gamma\delta$
- $\alpha\beta$
- β -> unique in each single cell
 - CDR3 region

Library preparation



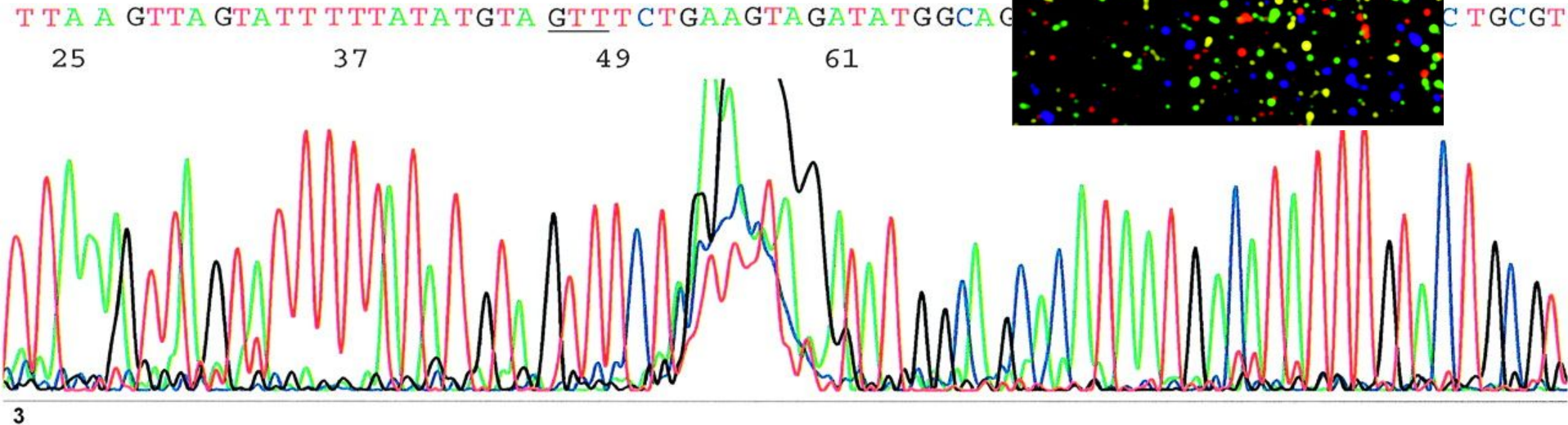
Rosati, E., Dowds, C. M., Liaskou, E., Henriksen, E. K. K., Karlsen, T. H., & Franke, A. (2017). Overview of methodologies for T-cell receptor repertoire analysis. *BMC Biotechnology*, 17(1), 61. <https://doi.org/10.1186/s12896-017-0379-9>

Barcoding Unique Molecular Identifiers



Bioinformática

Base calling



Phred quality score

$$Q = -10 \log_{10} P$$

Phred Quality Score	Probability of incorrect base call	Base call accuracy
10	1 in 10	90%
20	1 in 100	99%
30	1 in 1000	99.9%
40	1 in 10,000	99.99%
50	1 in 100,000	99.999%
60	1 in 1,000,000	99.9999%

FASTQ files

```
@K00114:390:HCGGLBBXX:1:1101:1722:2264 1:N:0
```

```
TGTTTGAGGAATTAGCCAGAGTAATAGTGTCAAGTTTCCATCACAATTTGAAGTCAGACATTATTTCTAACCAAAT  
TATACTCTTTTGTCTGCCAAGCTACCTTGACACTTTTAATATCTGAAGGCAGGCCTATCTGATAGTAGATCGGA
```

```
+
```

```
AAAFFJ-JJJFFFJFFFJJ<FAAJJJJFF<FJJJFJJFJJ<JAFFJJFJFFFJFFFJJFJFJJFJJFJJFFJAJA  
7JFJJJJJ<FJFFFJ7F<AFJJJ<J<FJJJFJJJJFFJJJFJJJJJJJAFJJ<-FJJJFJ<FFJJJAJJJJFFFF
```

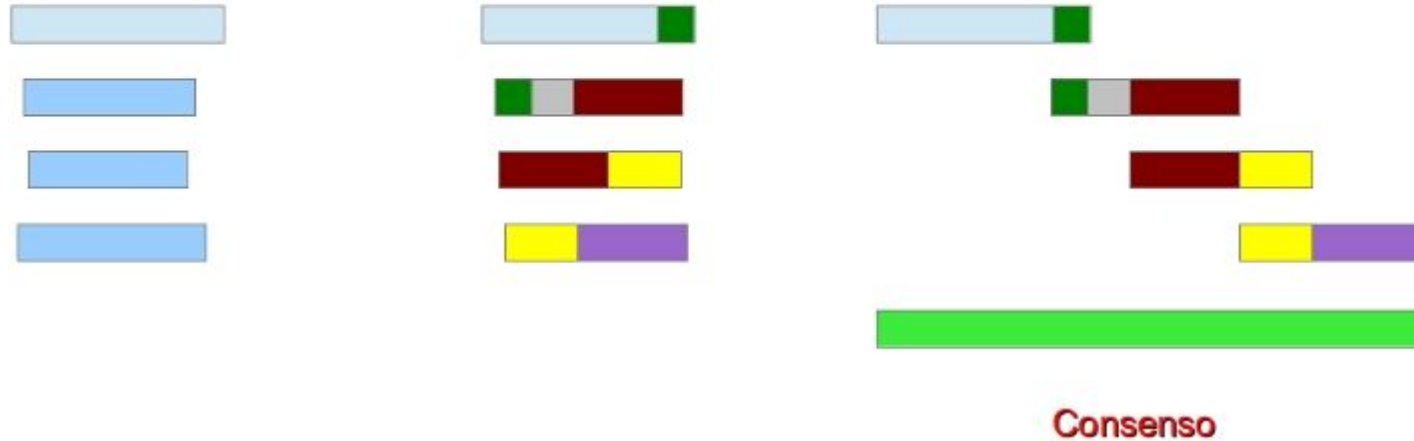
Codificación ASCII de valores Q

ASCII Table

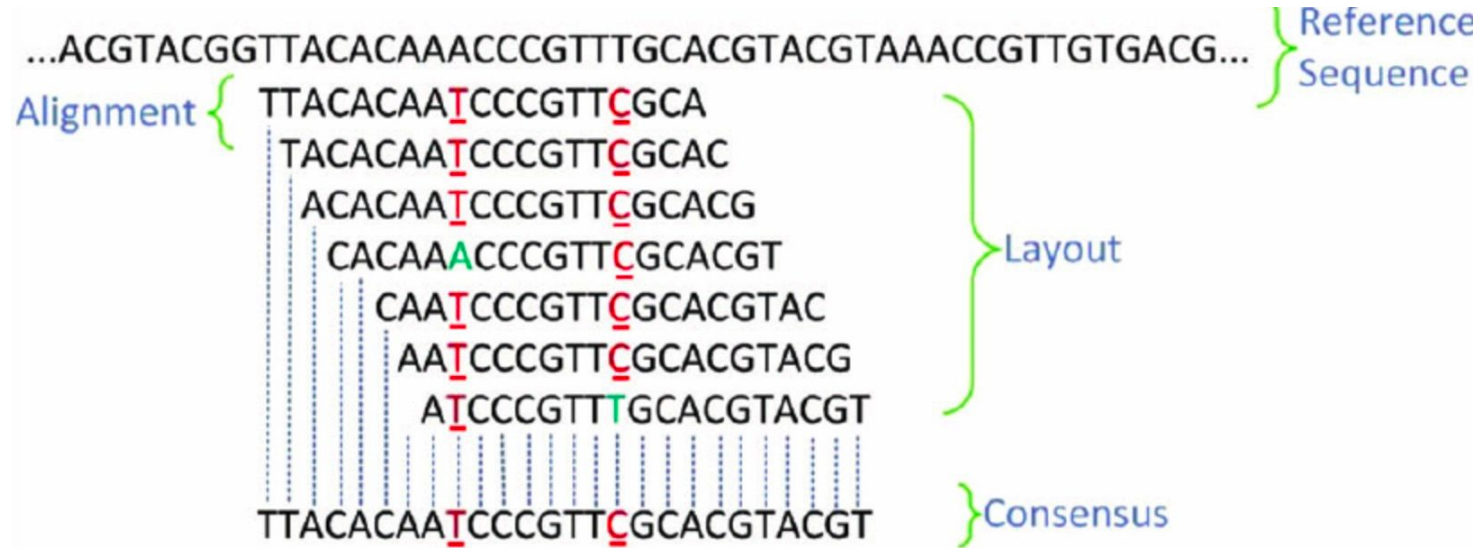
Dec	Hex	Oct	Char	Dec	Hex	Oct	Char	Dec	Hex	Oct	Char	Dec	Hex	Oct	Char
0	0	0		32	20	40	[space]	64	40	100	@	96	60	140	`
1	1	1		33	21	41	!	65	41	101	A	97	61	141	a
2	2	2		34	22	42	"	66	42	102	B	98	62	142	b
3	3	3		35	23	43	#	67	43	103	C	99	63	143	c
4	4	4		36	24	44	\$	68	44	104	D	100	64	144	d
5	5	5		37	25	45	%	69	45	105	E	101	65	145	e
6	6	6		38	26	46	&	70	46	106	F	102	66	146	f
7	7	7		39	27	47	'	71	47	107	G	103	67	147	g
8	8	10		40	28	50	(72	48	110	H	104	68	150	h
9	9	11		41	29	51)	73	49	111	I	105	69	151	i
10	A	12		42	2A	52	*	74	4A	112	J	106	6A	152	j
11	B	13		43	2B	53	+	75	4B	113	K	107	6B	153	k
12	C	14		44	2C	54	,	76	4C	114	L	108	6C	154	l
13	D	15		45	2D	55	-	77	4D	115	M	109	6D	155	m
14	E	16		46	2E	56	.	78	4E	116	N	110	6E	156	n
15	F	17		47	2F	57	/	79	4F	117	O	111	6F	157	o
16	10	20		48	30	60	0	80	50	120	P	112	70	160	p
17	11	21		49	31	61	1	81	51	121	Q	113	71	161	q
18	12	22		50	32	62	2	82	52	122	R	114	72	162	r
19	13	23		51	33	63	3	83	53	123	S	115	73	163	s
20	14	24		52	34	64	4	84	54	124	T	116	74	164	t
21	15	25		53	35	65	5	85	55	125	U	117	75	165	u
22	16	26		54	36	66	6	86	56	126	V	118	76	166	v
23	17	27		55	37	67	7	87	57	127	W	119	77	167	w
24	18	30		56	38	70	8	88	58	130	X	120	78	170	x
25	19	31		57	39	71	9	89	59	131	Y	121	79	171	y
26	1A	32		58	3A	72	:	90	5A	132	Z	122	7A	172	z
27	1B	33		59	3B	73	;	91	5B	133	[123	7B	173	{
28	1C	34		60	3C	74	<	92	5C	134	\	124	7C	174	
29	1D	35		61	3D	75	=	93	5D	135]	125	7D	175	}
30	1E	36		62	3E	76	>	94	5E	136	^	126	7E	176	~
31	1F	37		63	3F	77	?	95	5F	137	_	127	7F	177	

Ensamblado de novo

Si no existe un genoma de referencia es necesario ensamblar



Mapeo de lecturas al genoma de referencia



SAM / BAM file

```
K00114:390:HCGGLBBXX:1:2215:23663:9789      163      1      3000857      60      120M      =      3000959      223
TTTCAATATTTTTGTATCTGTTGAGGACTTTTTGTGAGTGACTATATGGTCAATTTTGGAGGATTTGGTACTGAGAAGAAGGTATATATCCTTTTGTCTTATGATAAAATGT
TCTGTAGA
CABFCE4CDCB>6GC>DG=EDCGDGF4?>=>B>BDE7FEHAFG?E<DHDDE?AA=DDHFFHG>ECDHGE@DGFAFCGHE:CECCC@FEFIGGCEEIFFHEDEHC?CBGEFIF
DFGHE9D?
BD:Z:QQIRTSSRRRHHSNSRTPTNRRNMSROORHHHSLNLMNLNNNSQQQRRRSTSRRHHSRRNSRTRHSRRSTOUPPOSSOSSTSTSSSSSUSTSIIUPURTSTUQVTT
LMUVQUTTXOTUO      MD:Z:120      PG:Z:MarkDuplicates      RG:Z:1
BI:Z:UUOWXXXXXXXXXOOOWUWWXRSTVTVVTVVWVTNNNUSRUTUSUUVVVVWSVUUUVVVOOWSWTWXXVNWTTWWTTWUUVVWWXWXXXXXXXXXXWPPXVWTWWXYXYXU
RQXZWYYUUUWUU      NM:i:0      AS:i:120      XS:i:87
```

```
K00114:390:HCGGLBBXX:1:2215:23663:9789      83      1      3000959      60      121M      =      3000857      -223
GATAAAATGTTCTGTAGATATCTATTAAATTCATTTGTTTCATAACTTCGGTTAGTGTCCGTGTGTCTCTGTTCAGTTTCTGCTTCCAGGATCTGTCTCTTGGTGAGAGTGT
GGTCTTGAA
@?CA<9@AAGFIHGFEFEFEFFIEFG@DCEGEFFGDHDBGFIF@C=IGFC@FBCGDCFEE>>DBGFECEHGFDEHGFCCDHGGFFECFD>>/EF0CCFEFDGFFFEF0GAC>C
@-DE4BEDD
BD:Z:USRILVWPVUQWOVURVTSUOUSSSRISRROTSJSMSKROSSQRORRSRSSRSOMMORQRMMMONNNUMSROTORJRNUTURRRRTSRTTNUMONMMQSSSMTPPP
PNNNTUQQUVURQQ      MD:Z:121      PG:Z:MarkDuplicates      RG:Z:1
BI:Z:XWVOQYYUXYV[VXXTYXXYVXXXUWPXXWXXQXTWPWWXVWVWWWWUTWTWSSWVWVSTSTXUUUYTWWTWVOVTXWVWWXTTWXXUYTXUUUWWWTSVUSU
WSTXUYVXYVWUU      NM:i:0      AS:i:121      XS:i:66
```

Cuestionario

- 1- ¿Qué información se obtiene a partir de la secuenciación de material genético?
- 2- ¿Cuáles son los beneficios de la secuenciación de nueva generación comparado con otras tecnologías?
- 3- Mencione las distintas plataformas de secuenciación vistas en clase, sus principales características y diferencias.
- 4- ¿Cuál es la utilidad de los adaptadores? ¿Qué otras secuencias se pueden agregar a los fragmentos a secuenciar?
- 5- ¿Por qué la amplificación por PCR puede generar errores? ¿Qué alternativas existen para omitir este problema?
- 6- Mencione ejemplos de uso actual en diagnóstico médico. Piense qué estudios del laboratorio clínico podrían realizarse con esta técnica y justifique.
- 7- Explicar los principios del RNAseq
- 8- ¿A qué causas puede deberse que un nucleótido no se corresponda con el genoma de referencia?
- 9- ¿Por qué es importante la longitud de los fragmentos secuenciados?
- 10- ¿En qué partes del proceso de secuenciación con la plataforma de Illumina se pueden cometer errores?
- 11- ¿Cuál es la utilidad de secuenciar células únicas? Enumere posibles aplicaciones.