

BI694

Bioinformatics & Phylogenetics

Winter Semester 2017

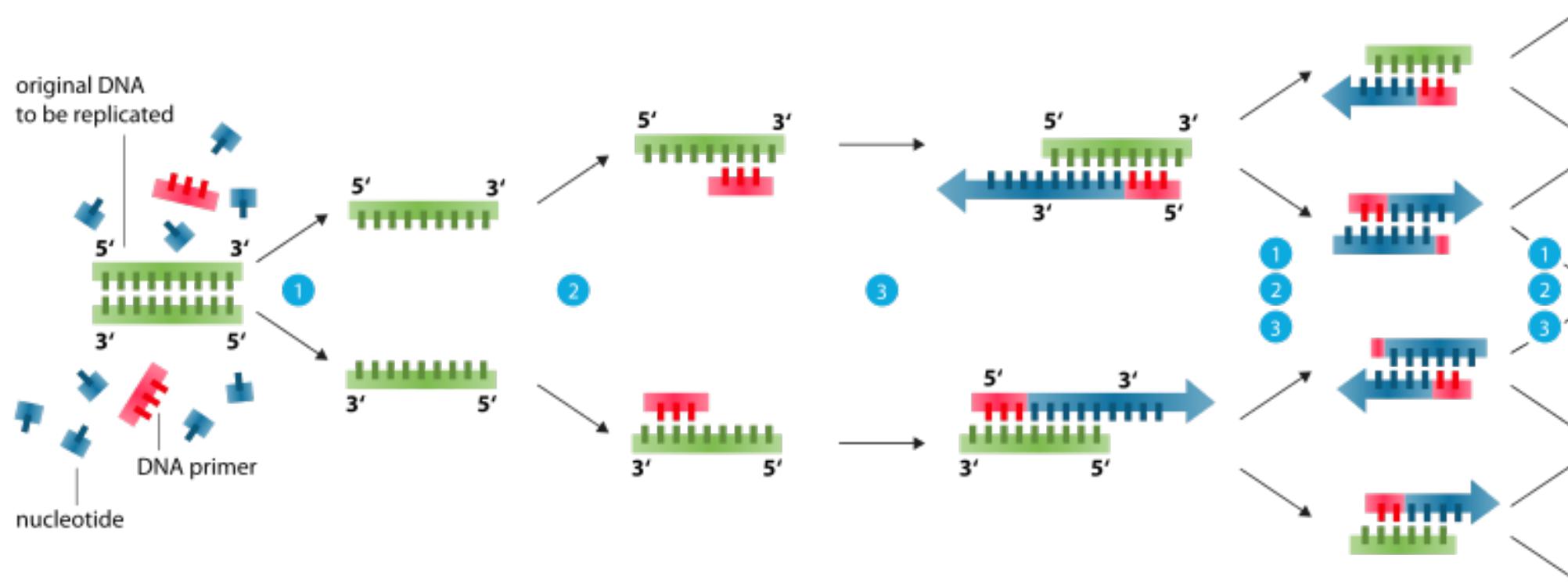
WEEK 7

PCR, Sanger Sequencing, Next Generation Sequencing

Overview

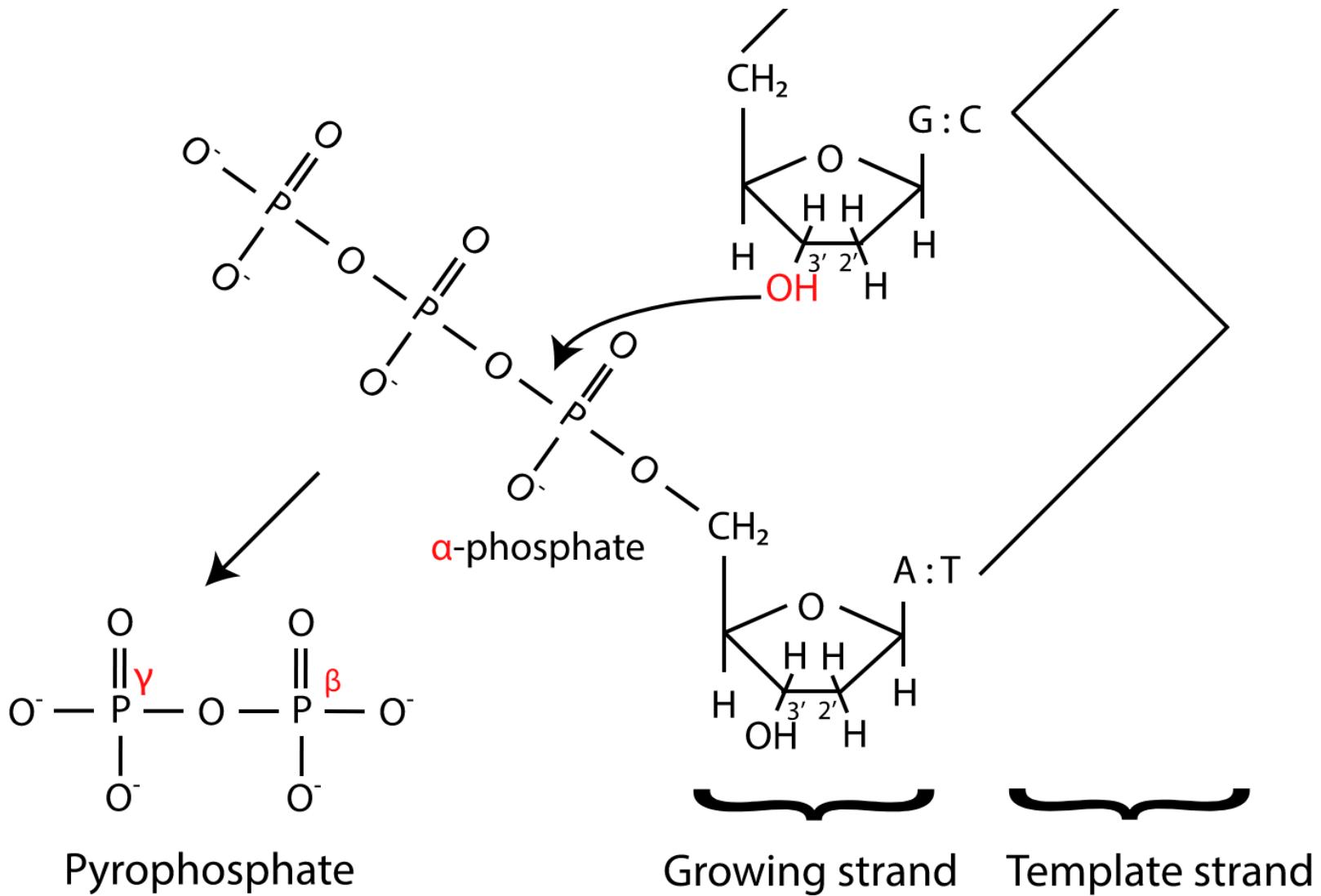
- Polymerase Chain Reaction
- Sanger Sequencing
- Pyro-Sequencing
- Semiconductor Sequencing
- Illumina Sequencing
- PacBio Sequencing
- What type of data do these new technologies generate?

Polymerase Chain Reaction (PCR)



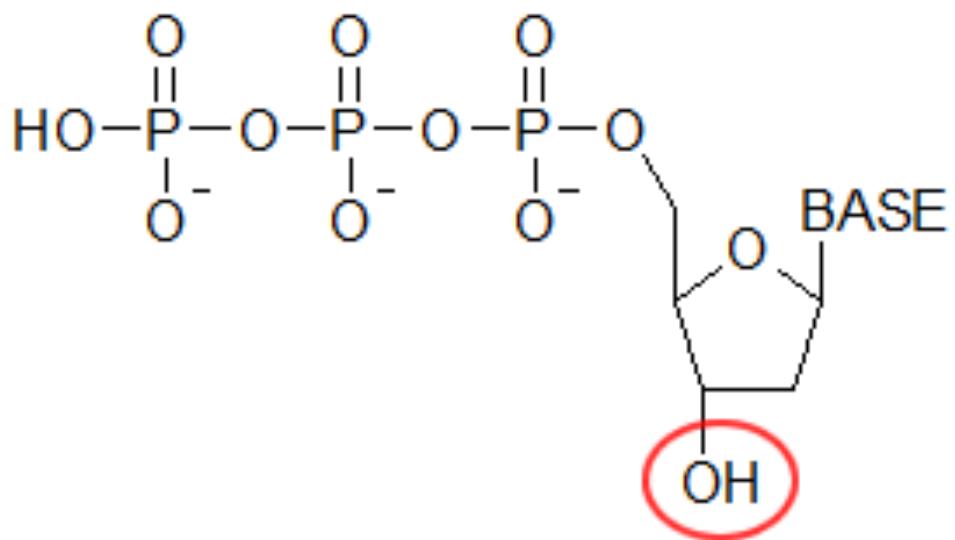
- 1 **Denaturation** at 94-96°C
- 2 **Annealing** at ~68°C
- 3 **Elongation** at ca. 72 °C

Sanger Sequencing

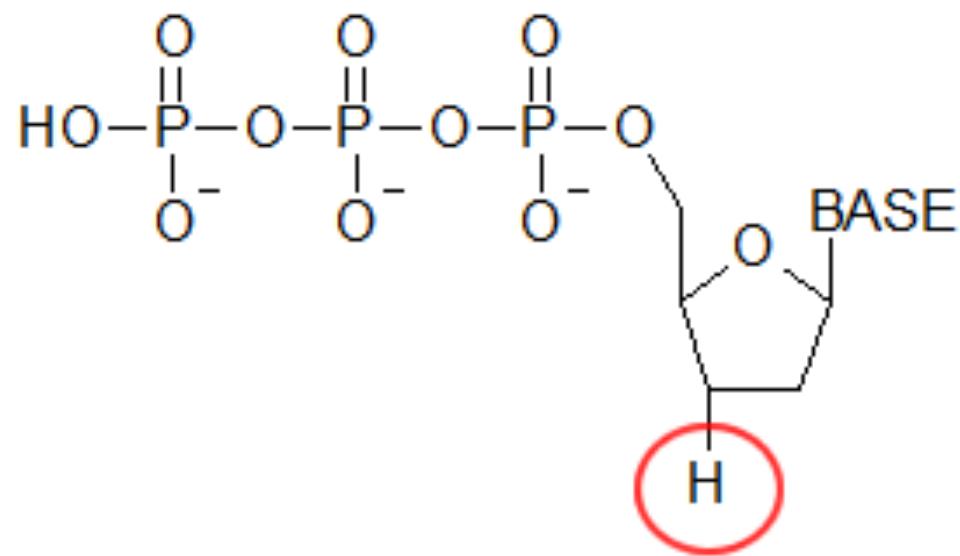


Pyrophosphate is a by-product of DNA elongation.

Sanger Sequencing

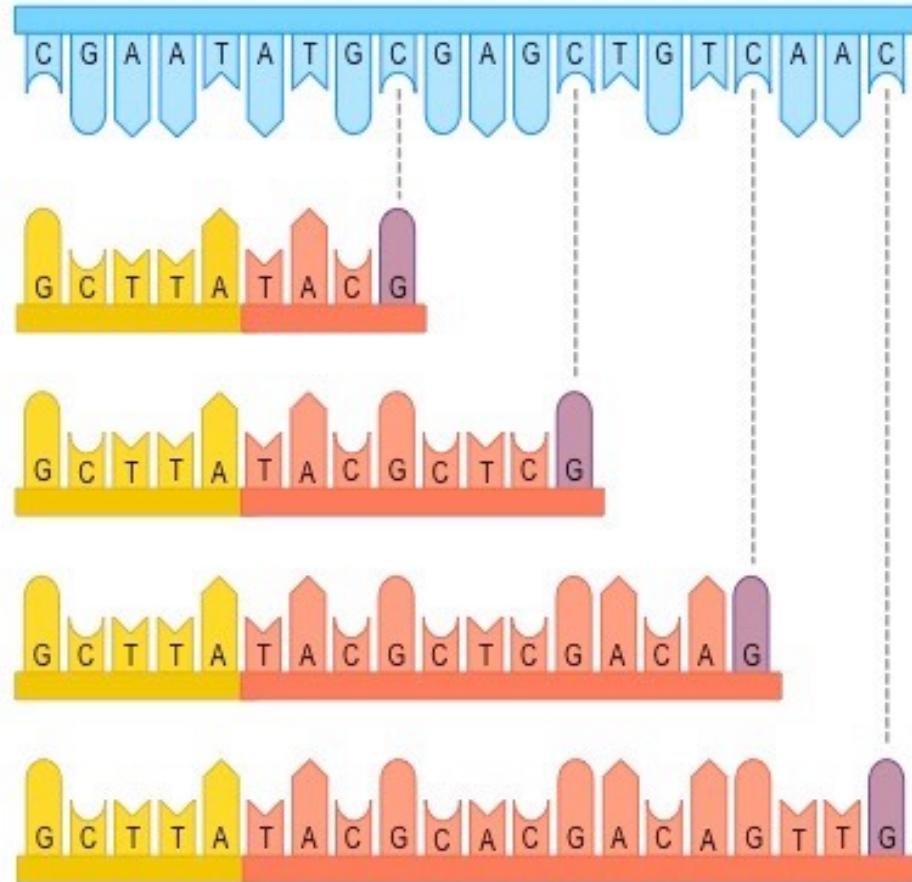


deoxynucleotide triphosphate

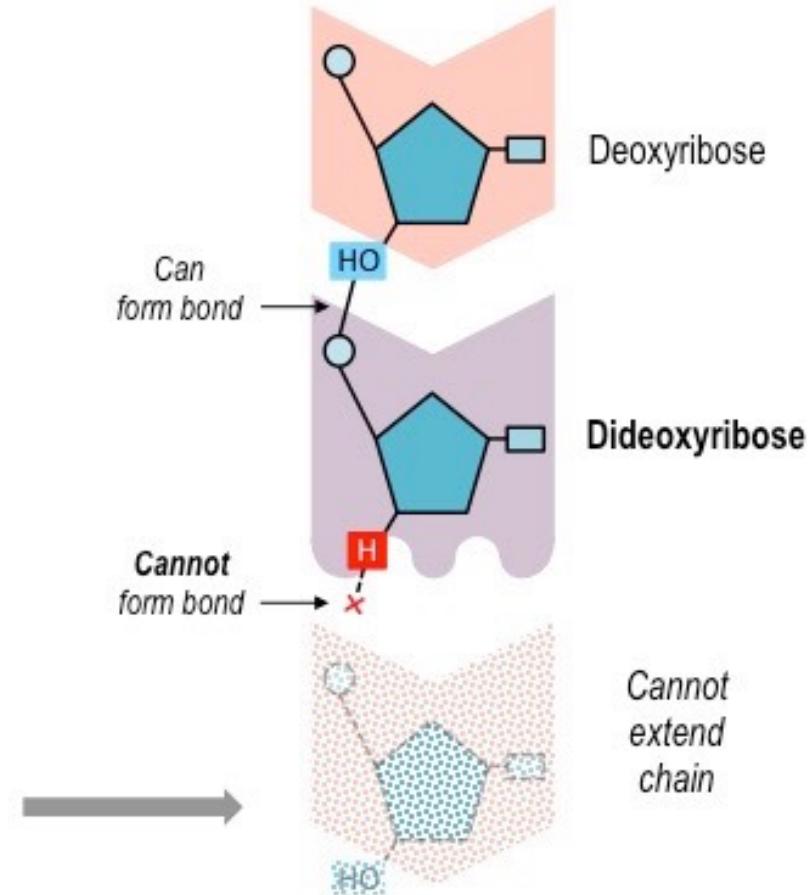


dideoxynucleotide triphosphate

Sanger Sequencing



Sequence terminates when the ddNTP is incorporated
Fragment lengths reflect base position in sequence

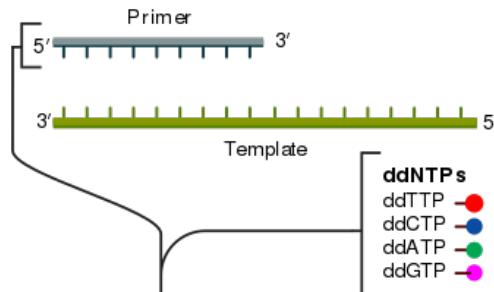


Chain termination by
dideoxynucleotides

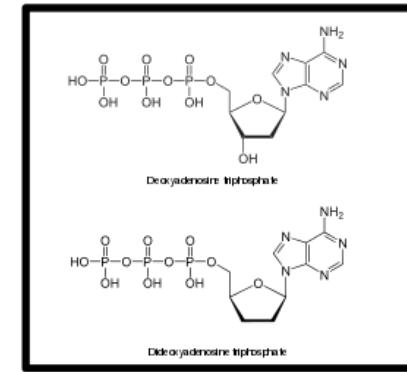
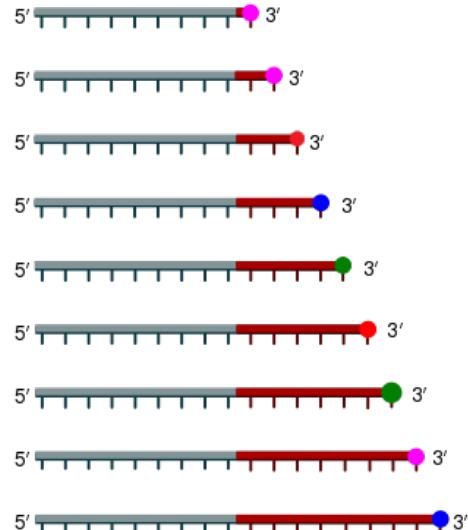
Sanger Sequencing

① Reaction mixture

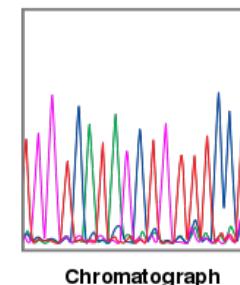
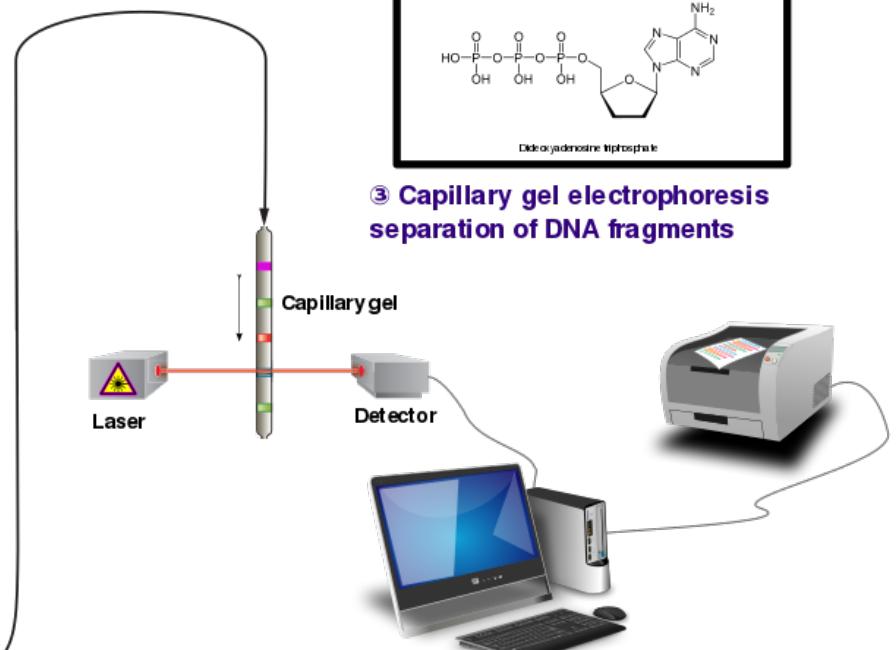
- Primer and DNA template
- DNA polymerase
- ddNTPs with flourochromes
- dNTPs (dATP, dCTP, dGTP, and dTTP)



② Primer elongation and chain termination

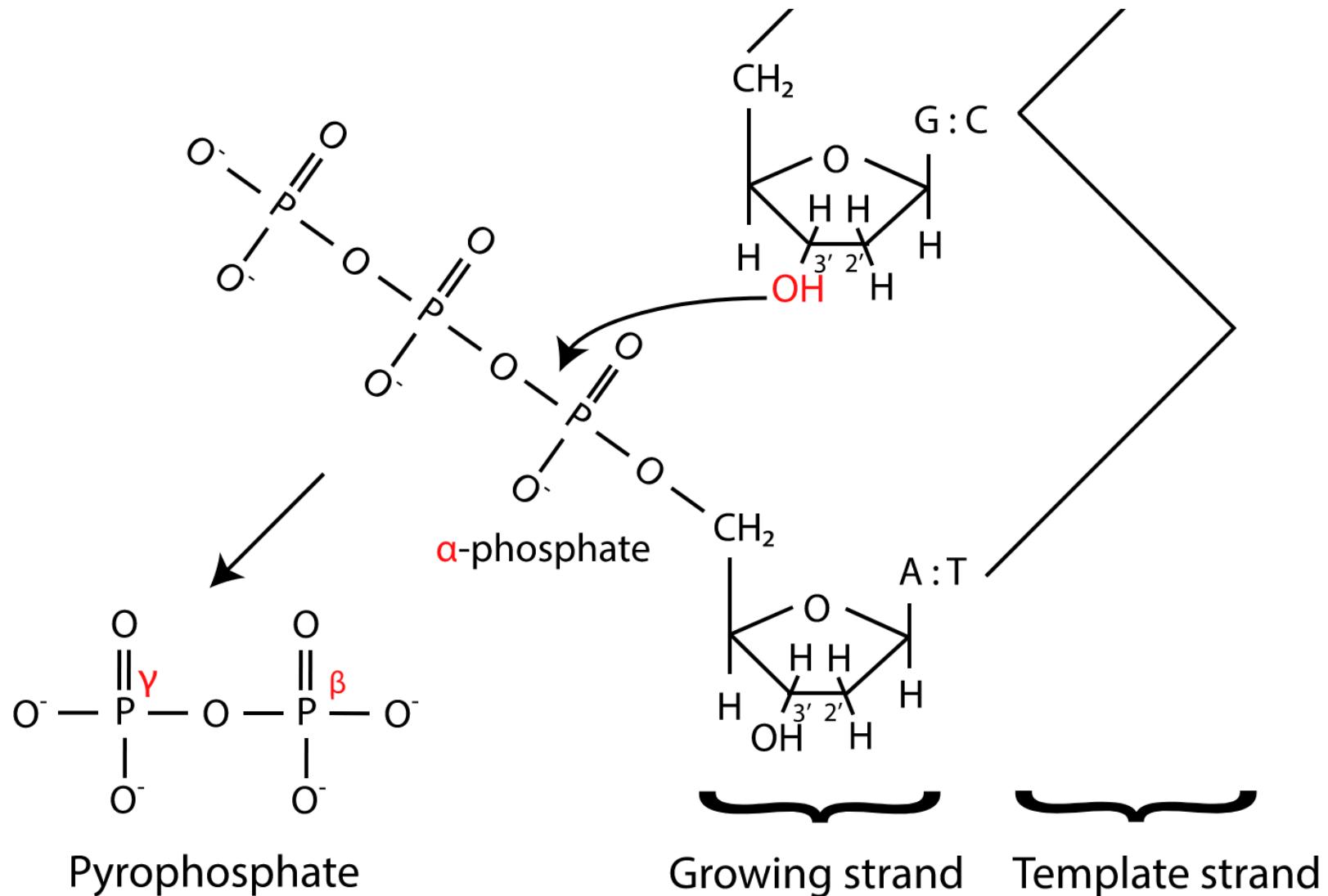


③ Capillary gel electrophoresis separation of DNA fragments



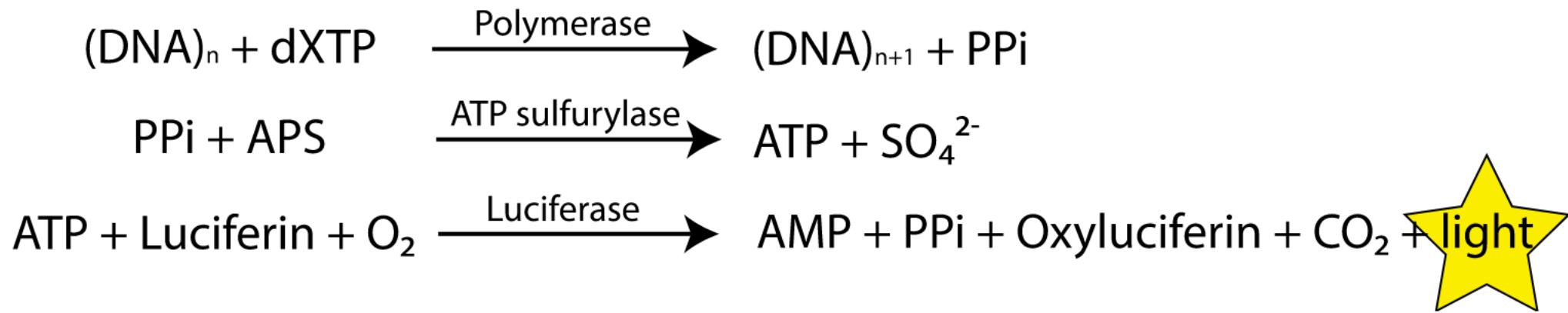
④ Laser detection of flourochromes and computational sequence analysis

Pyrosequencing – 454



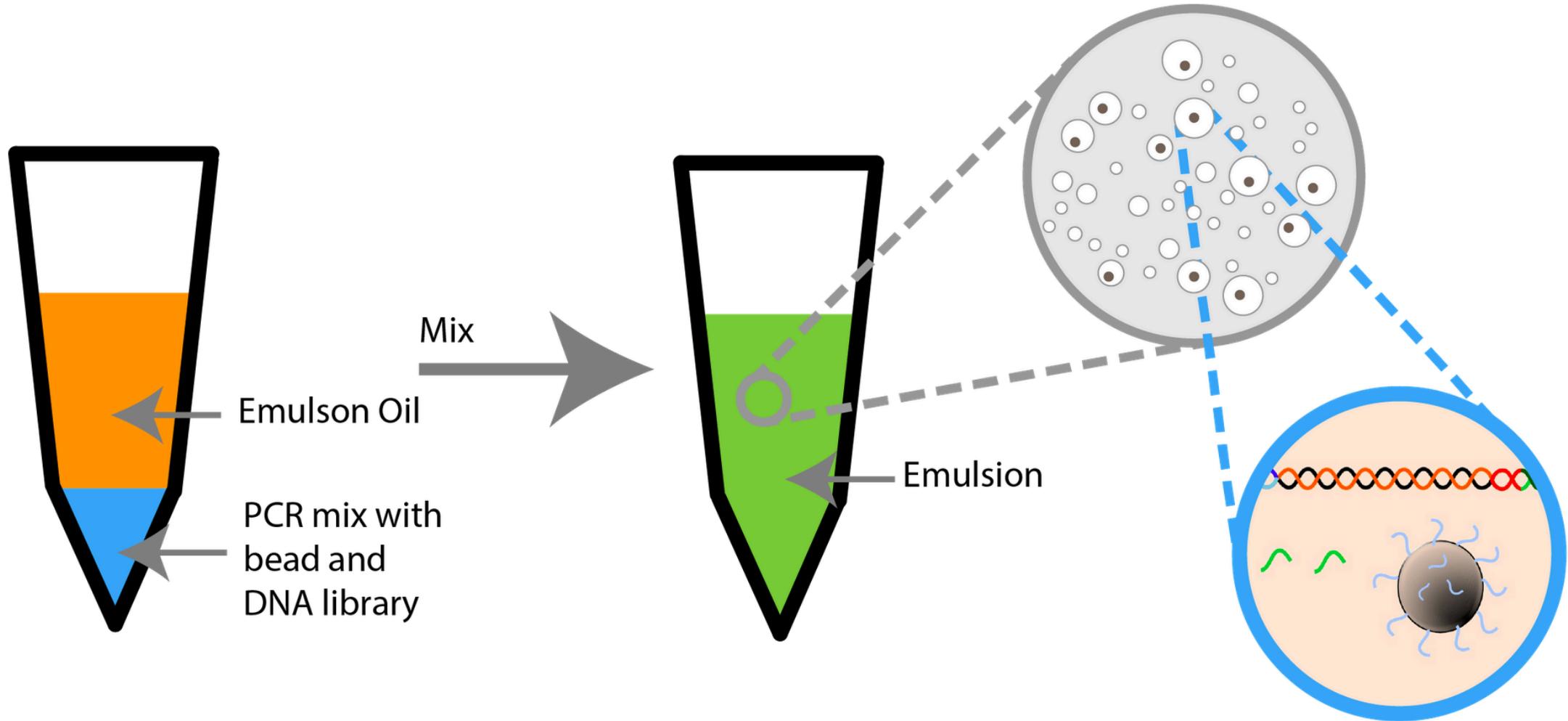
Pyrophosphate is a by-product of DNA elongation.

Pyrosequencing – 454



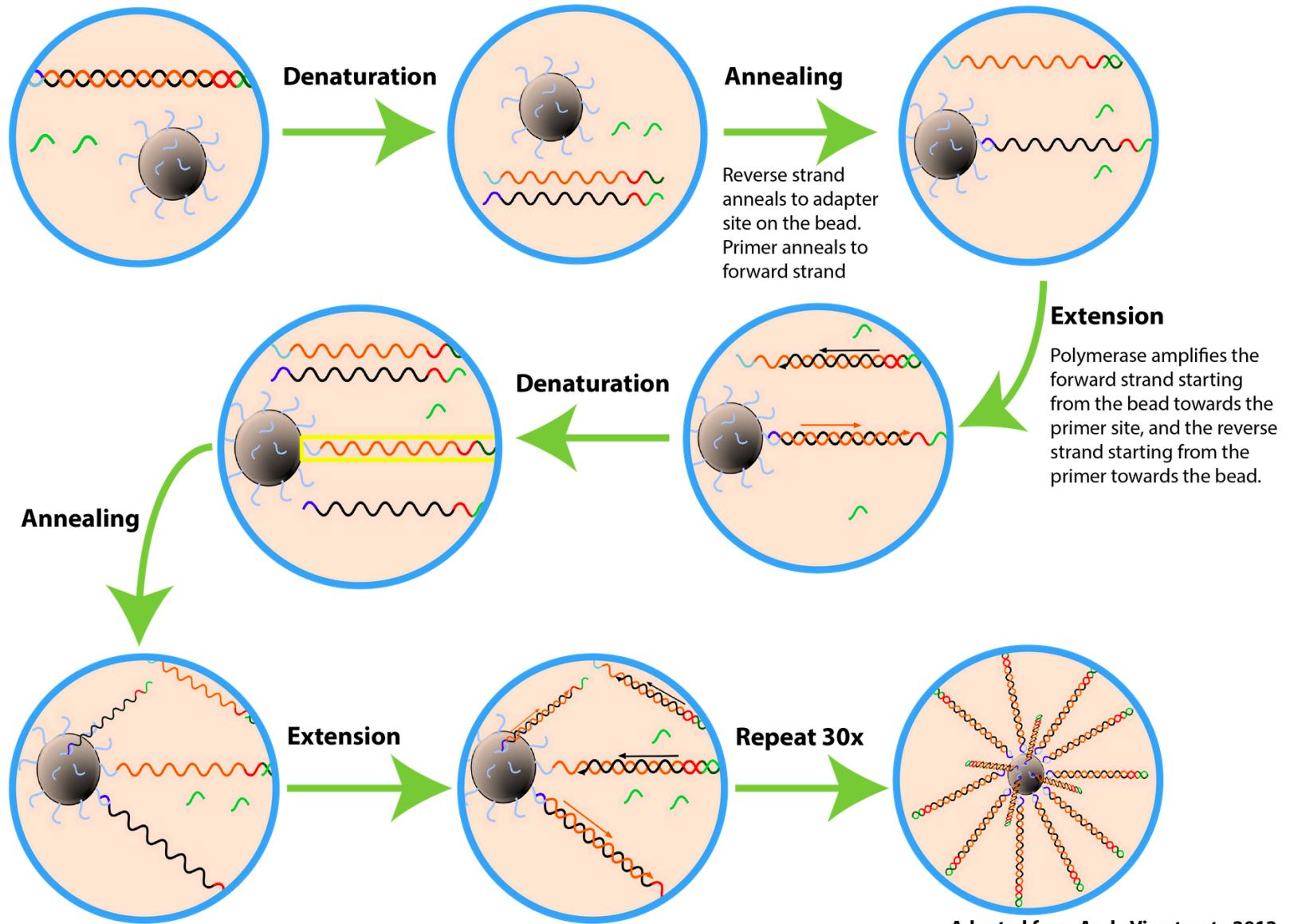
Steps to detecting a nucleotide by pyrosequencing: 1) PPi is generated if dXTP is incorporated. 2) PPi reacts and through a series of reactions, light is generated. 3) Light gets picked up by the sensor. 4) The more intense the light is, the more dXTP's are in that region of sequence.

Pyrosequencing – 454

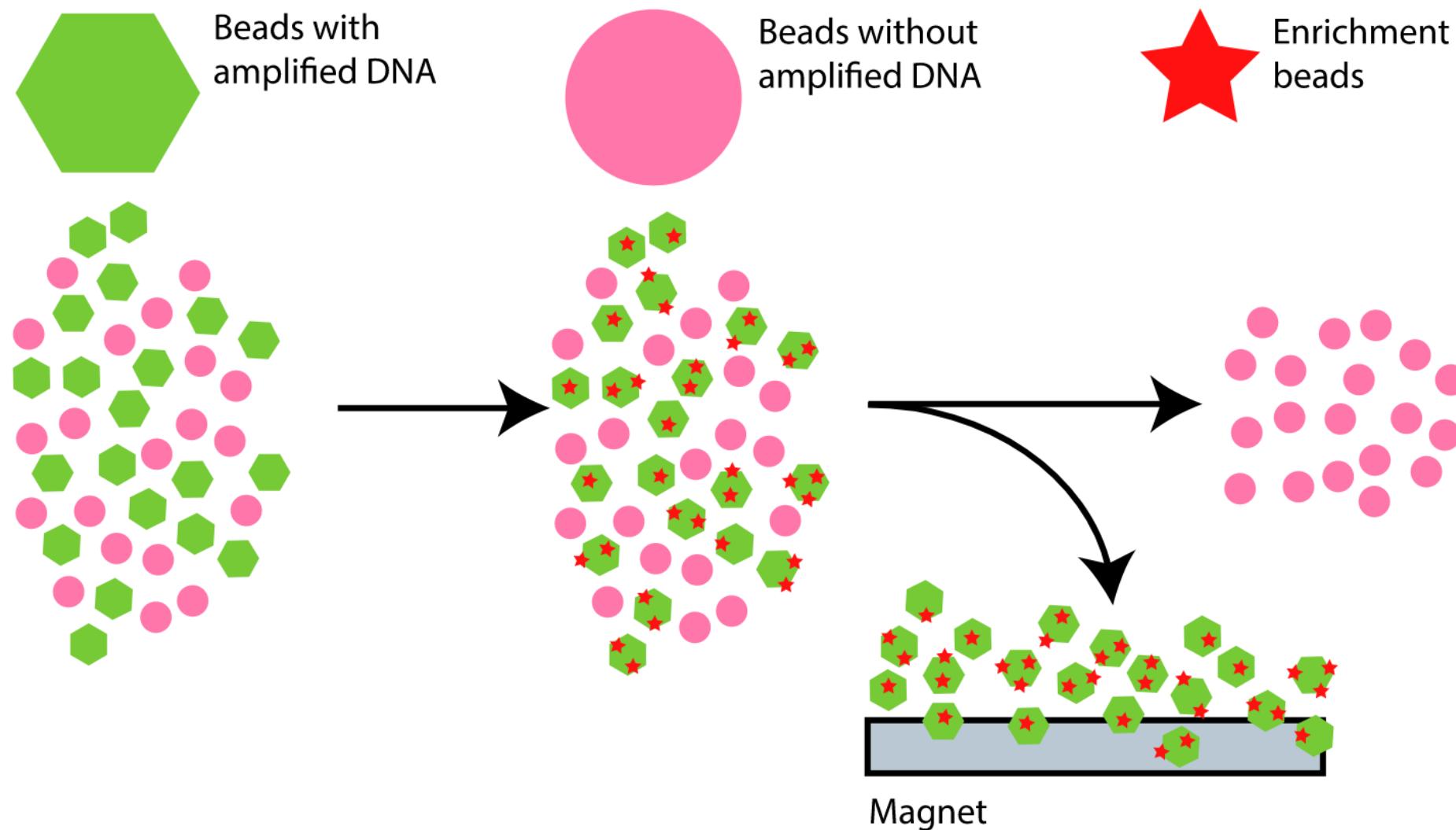


Emulsion PCR components.

Pyrosequencing – 454



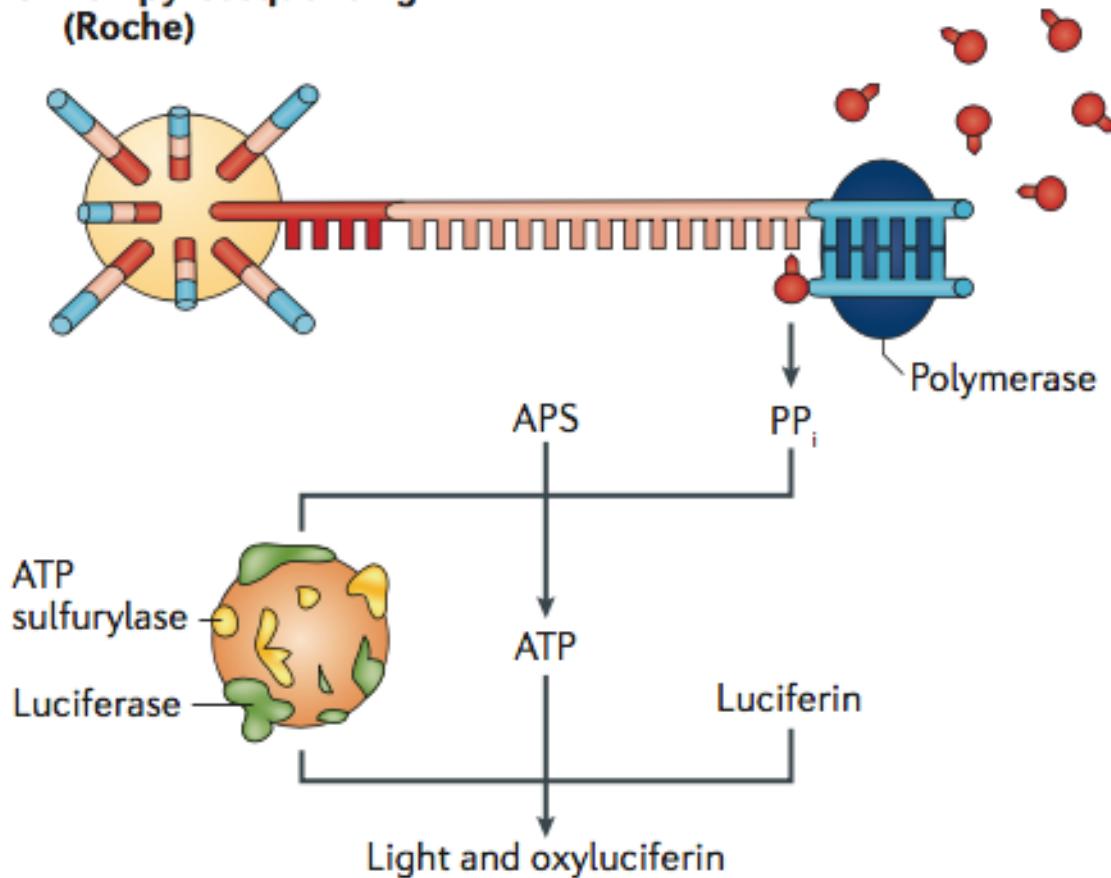
Pyrosequencing – 454



Bead enrichment using magnets.

Pyrosequencing – 454

a 454 pyrosequencing
(Roche)

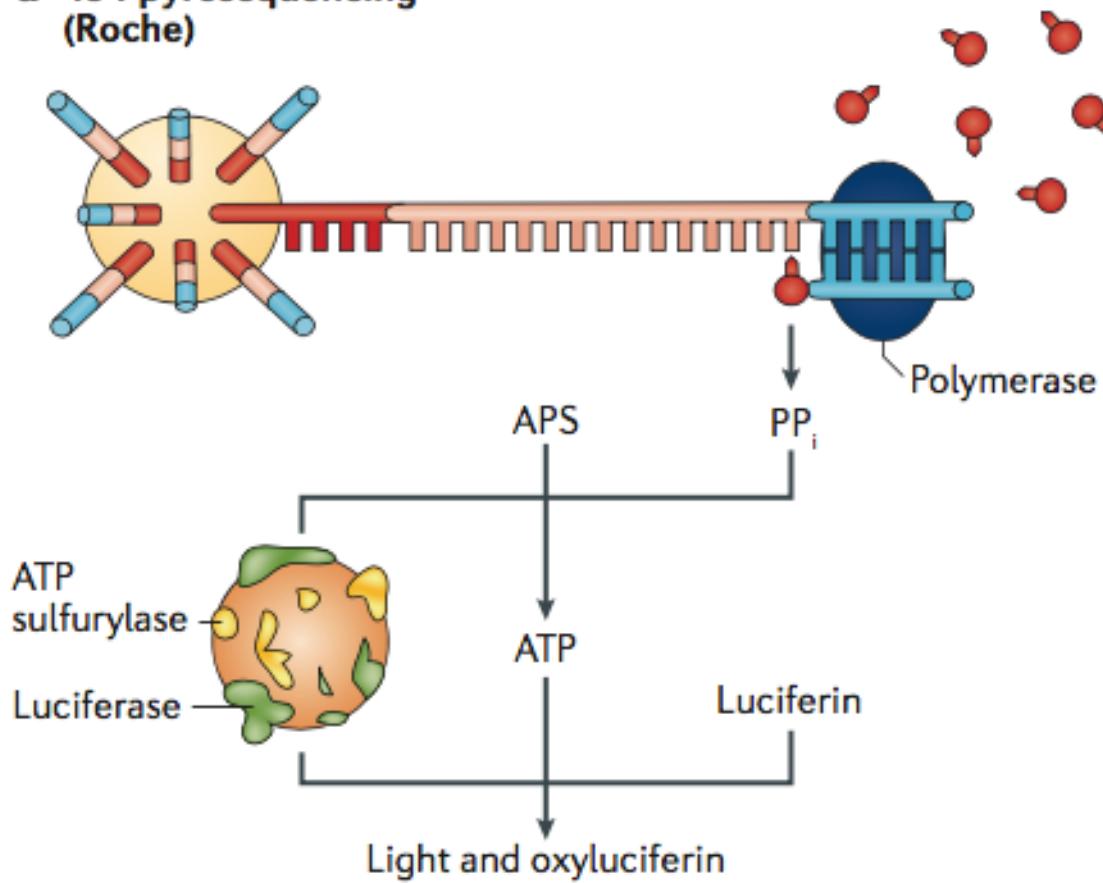


Pyrosequencing

As a base is incorporated, the release of an inorganic pyrophosphate triggers an enzyme cascade, resulting in light

Pyrosequencing – 454

a 454 pyrosequencing (Roche)

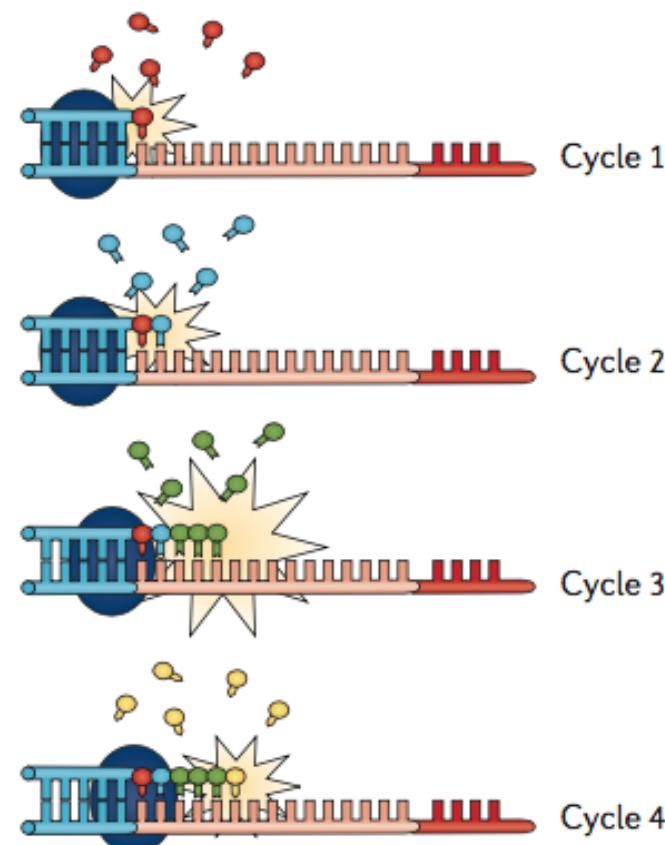


Pyrosequencing

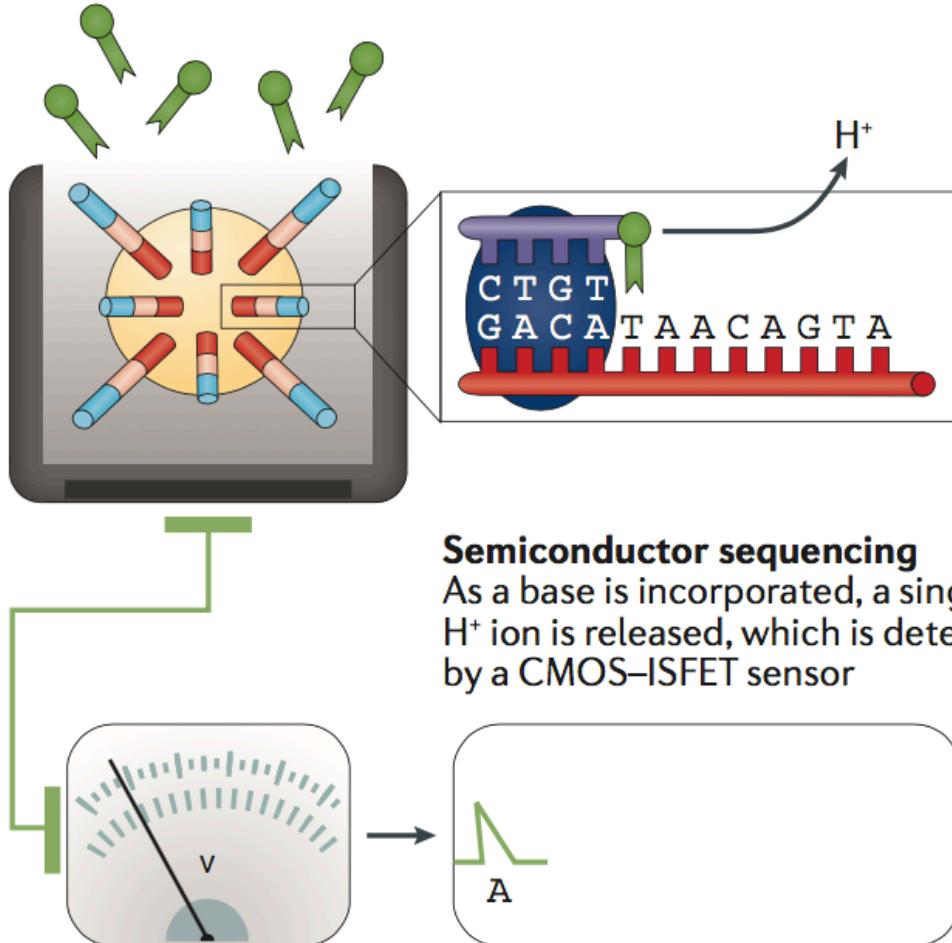
As a base is incorporated, the release of an inorganic pyrophosphate triggers an enzyme cascade, resulting in light

Single nucleotide addition

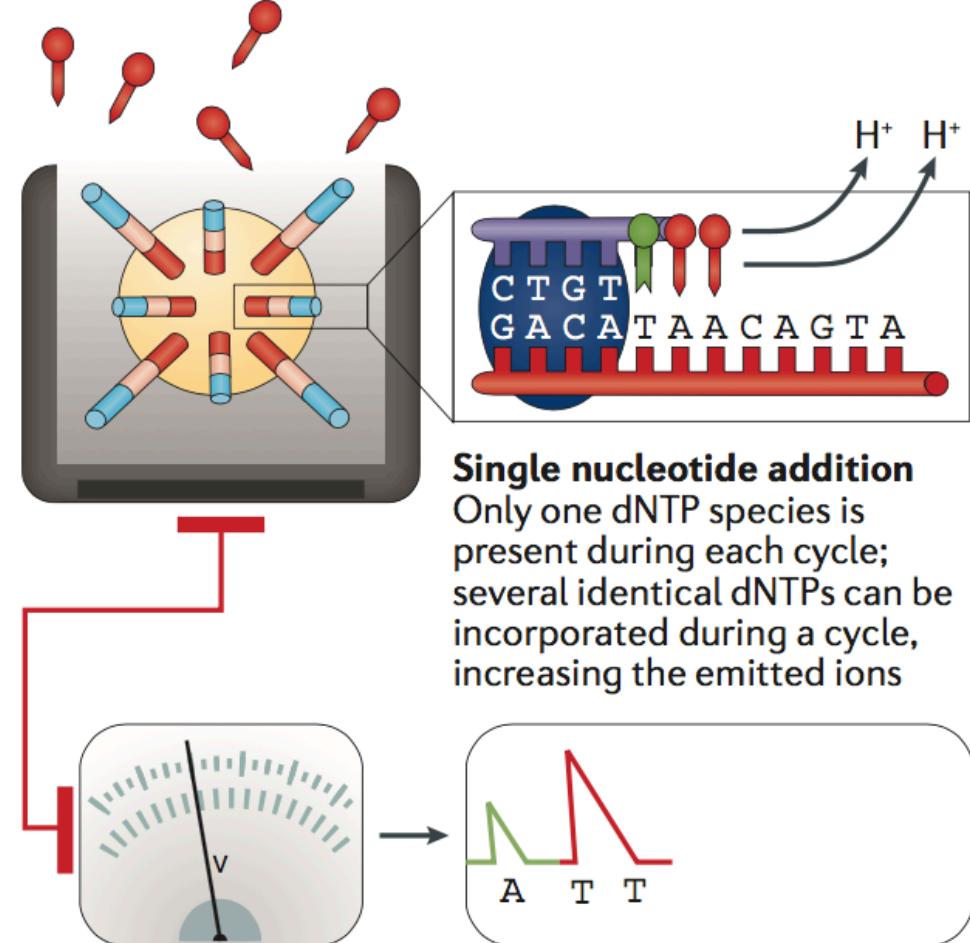
Only one dNTP species is present during each cycle; multiple identical dNTPs can be incorporated during a cycle, increasing emitted light



Semiconductor Sequencing – ION Torrent

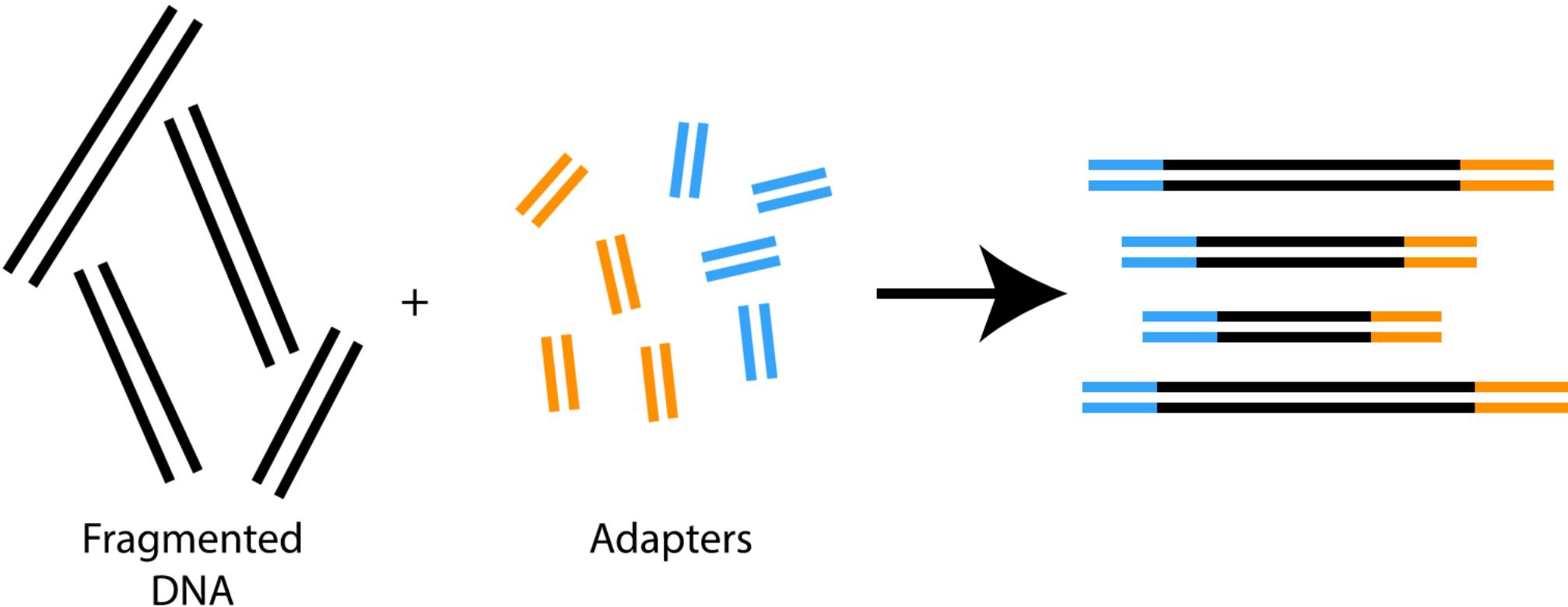


Semiconductor sequencing
As a base is incorporated, a single H^+ ion is released, which is detected by a CMOS-ISFET sensor



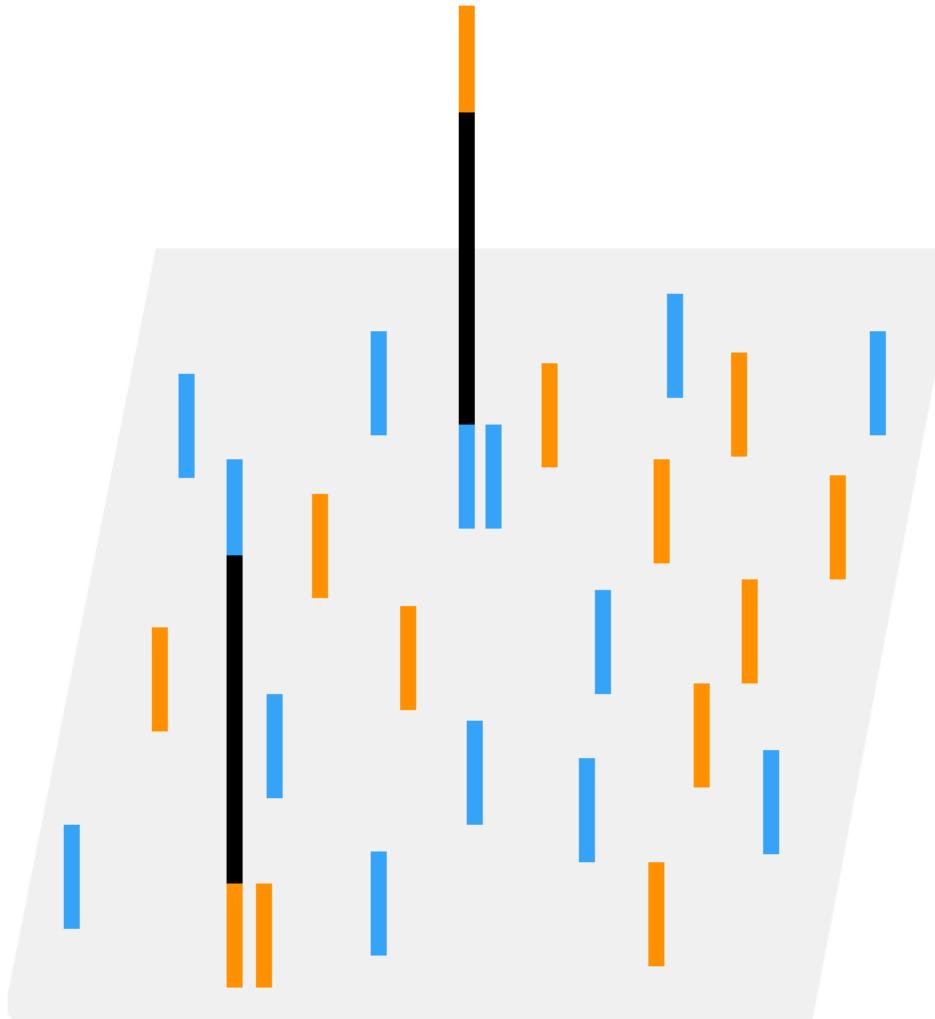
Single nucleotide addition
Only one dNTP species is present during each cycle; several identical dNTPs can be incorporated during a cycle, increasing the emitted ions

Illumina Sequencing: Library Preparation



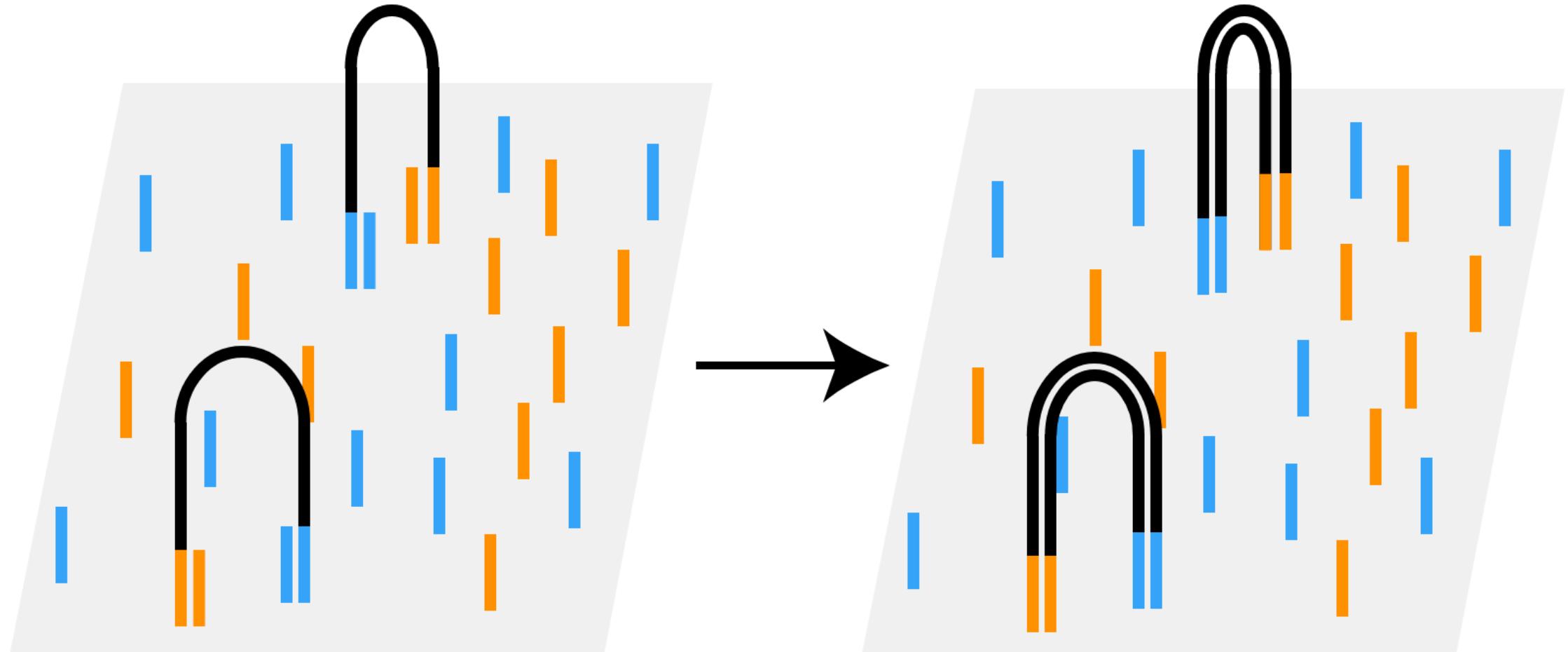
Adapters are attached to fragmented DNA.

Illumina Sequencing: Flowcell Ligation



DNA strands are ligated onto the flowcell.

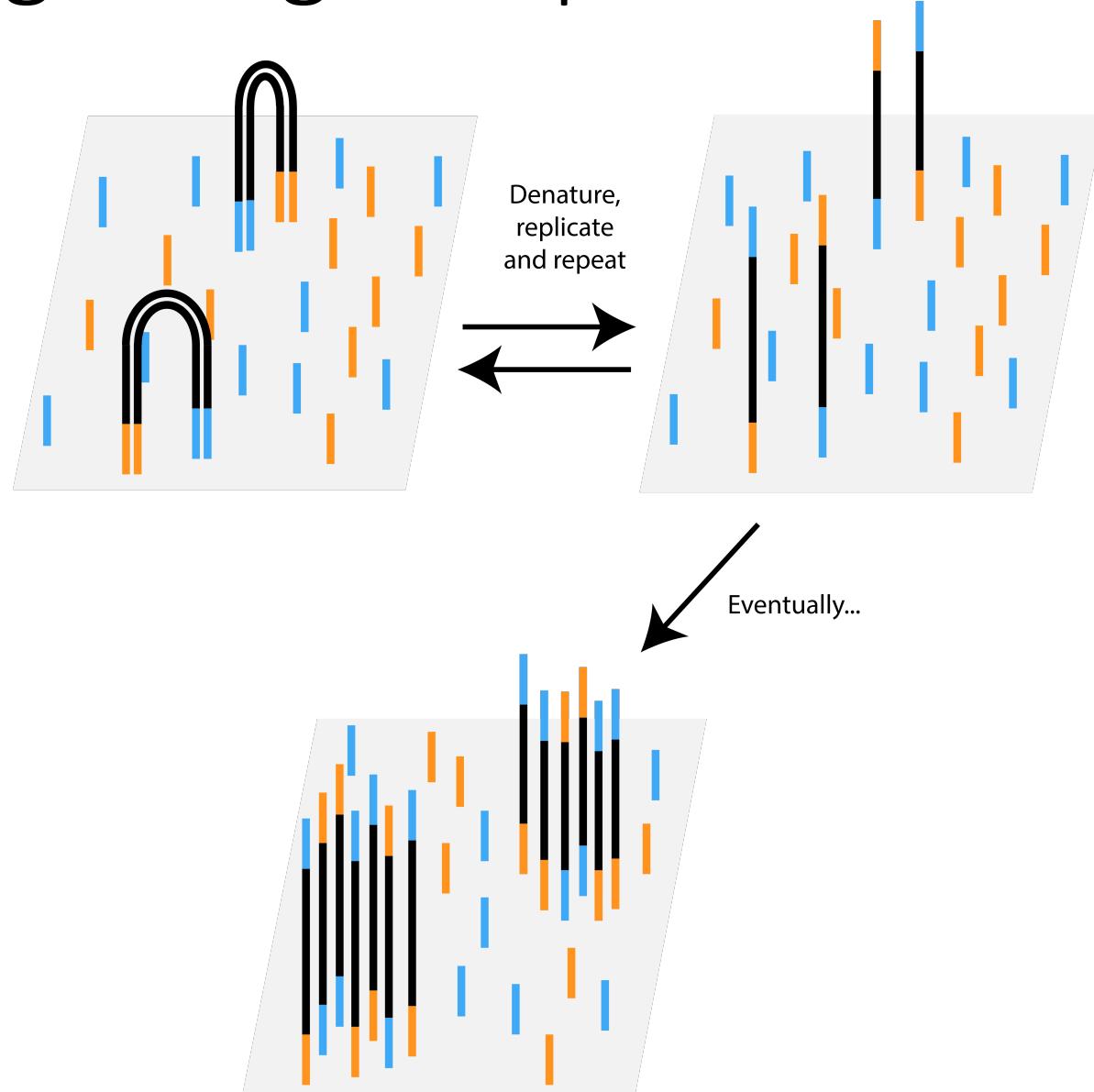
Illumina Sequencing: Bridge Amplification



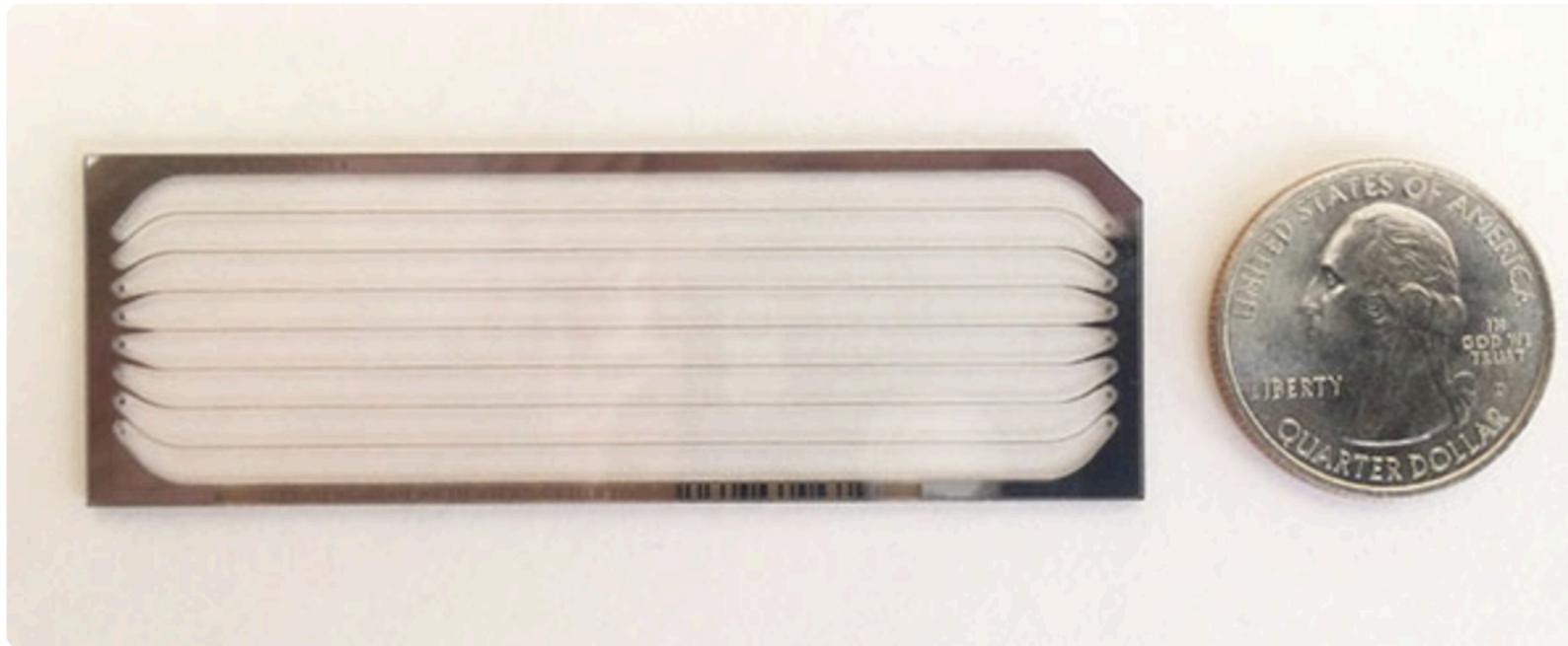
Bridge amplification occurs with regular PCR components.

Illumina Sequencing: Bridge Amplification

What is the purpose
of generating high
density clusters of
sequences?

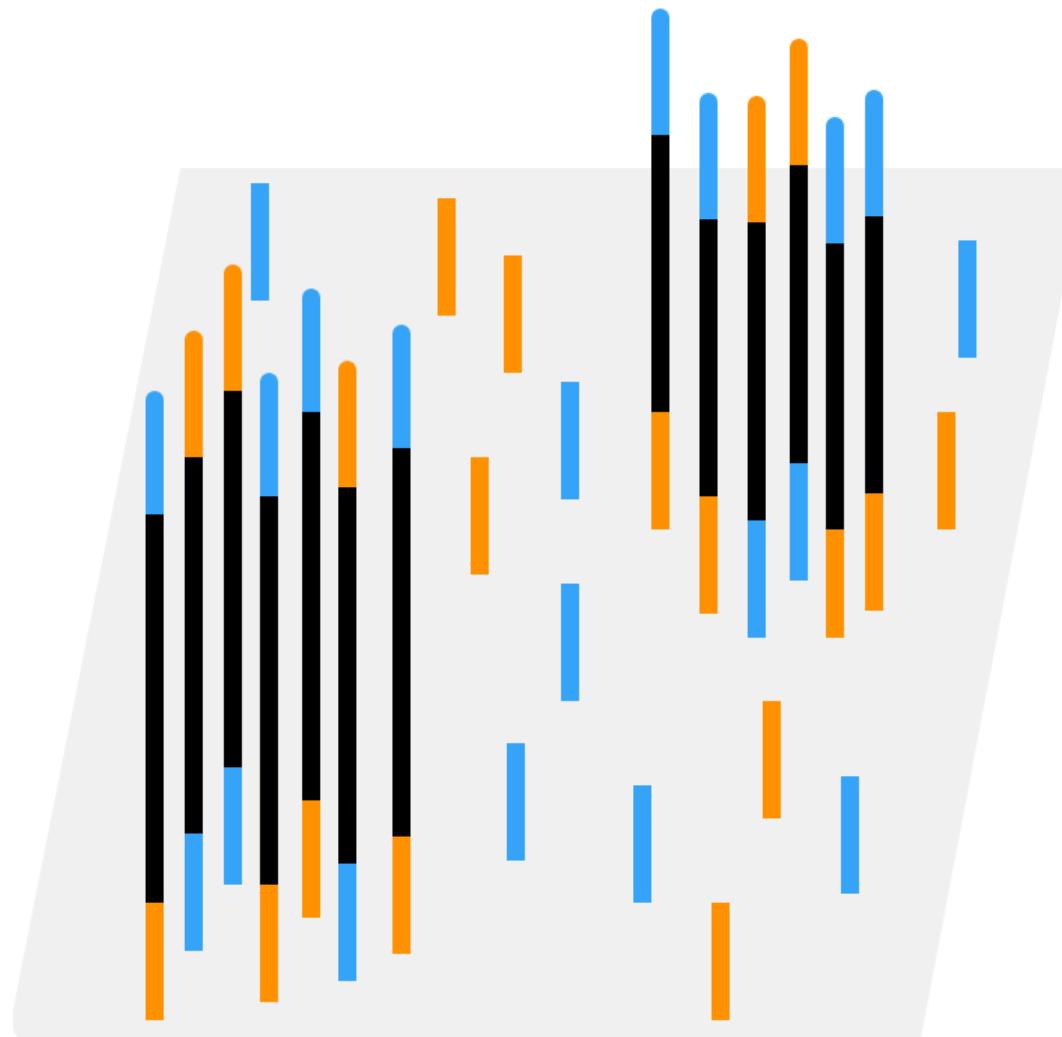


Illumina Sequencing: Flowcell



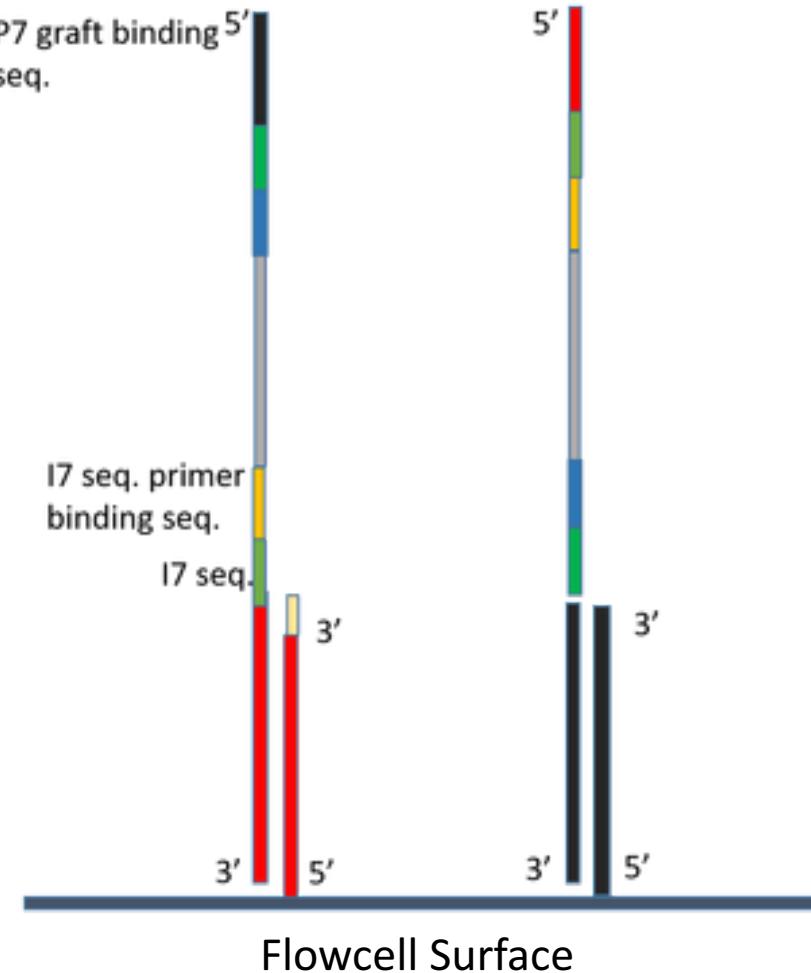
What an Illumina flowcell looks like with a US quarter for scale. Very similar to a microscope slide.

Illumina Sequencing: Sequencing by Synthesis

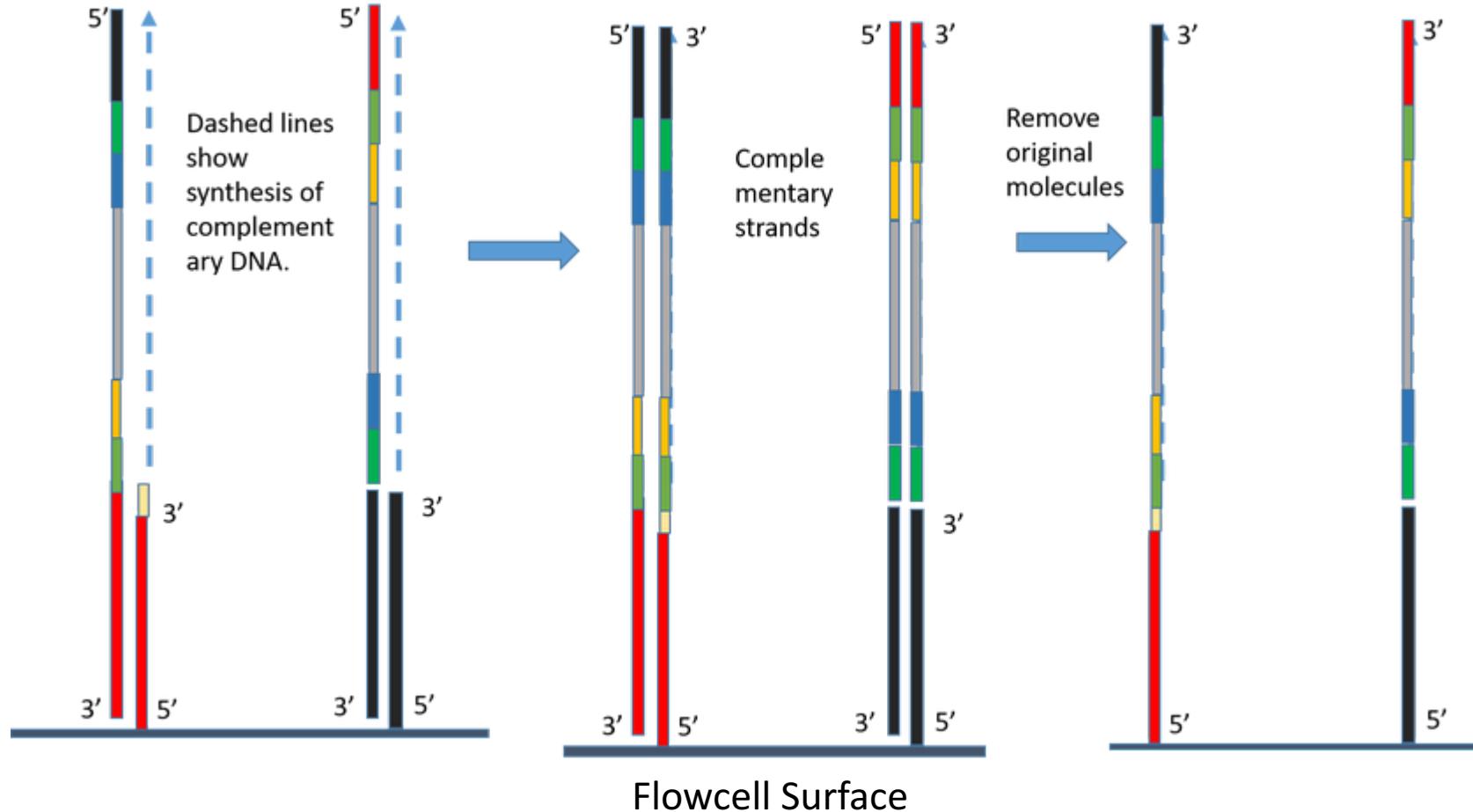


Several clusters are formed on the Illumina flow cell like the above.

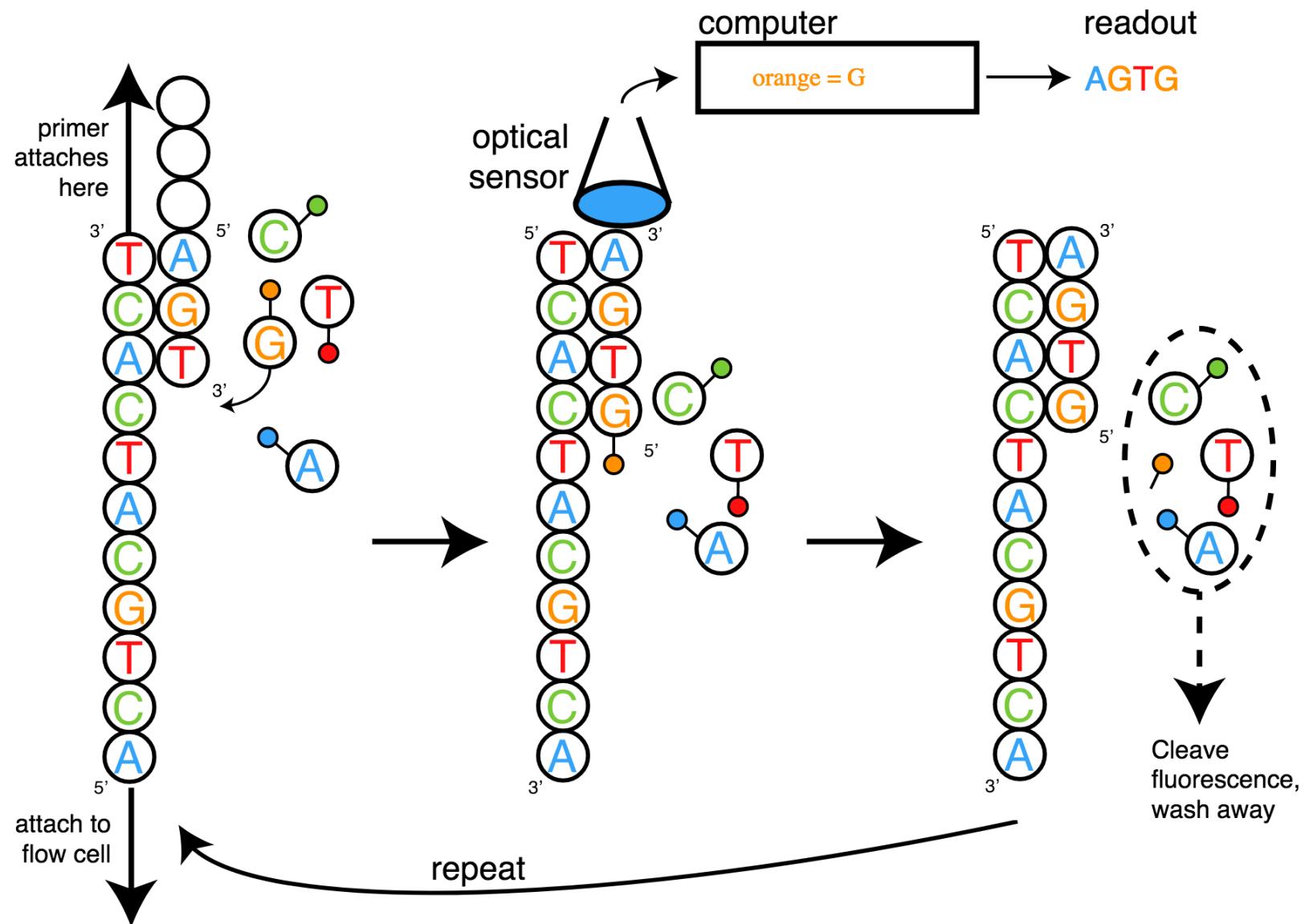
Illumina Sequencing: Sequencing by Synthesis



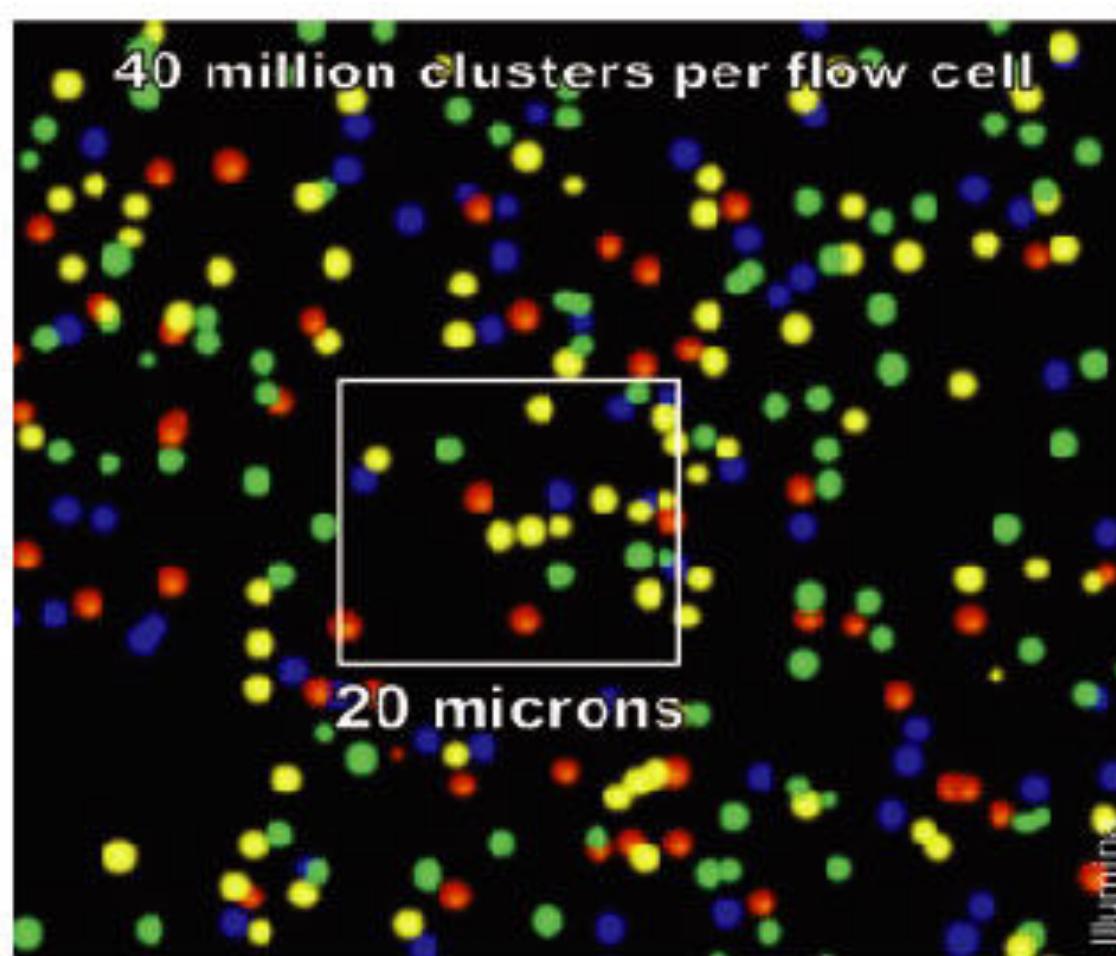
Illumina Sequencing: Sequencing by Synthesis



Illumina Sequencing: Sequencing by Synthesis



Illumina Sequencing: Sequencing by Synthesis



Long Read Sequencing: eg, PacBio

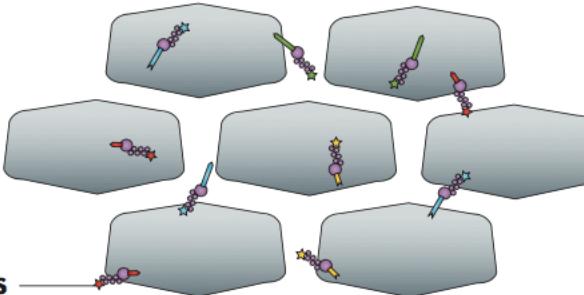
SMRTbell template

Two hairpin adapters allow continuous circular sequencing



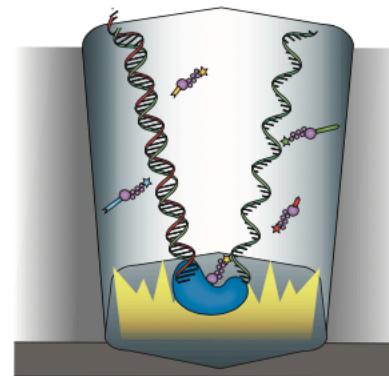
ZMW wells

Sites where sequencing takes place



Labelled nucleotides

All four dNTPs are labelled and available for incorporation



Modified polymerase

As a nucleotide is incorporated by the polymerase, a camera records the emitted light



PacBio output

A camera records the changing colours from all ZMWs; each colour change corresponds to one base

FastQ File Format...

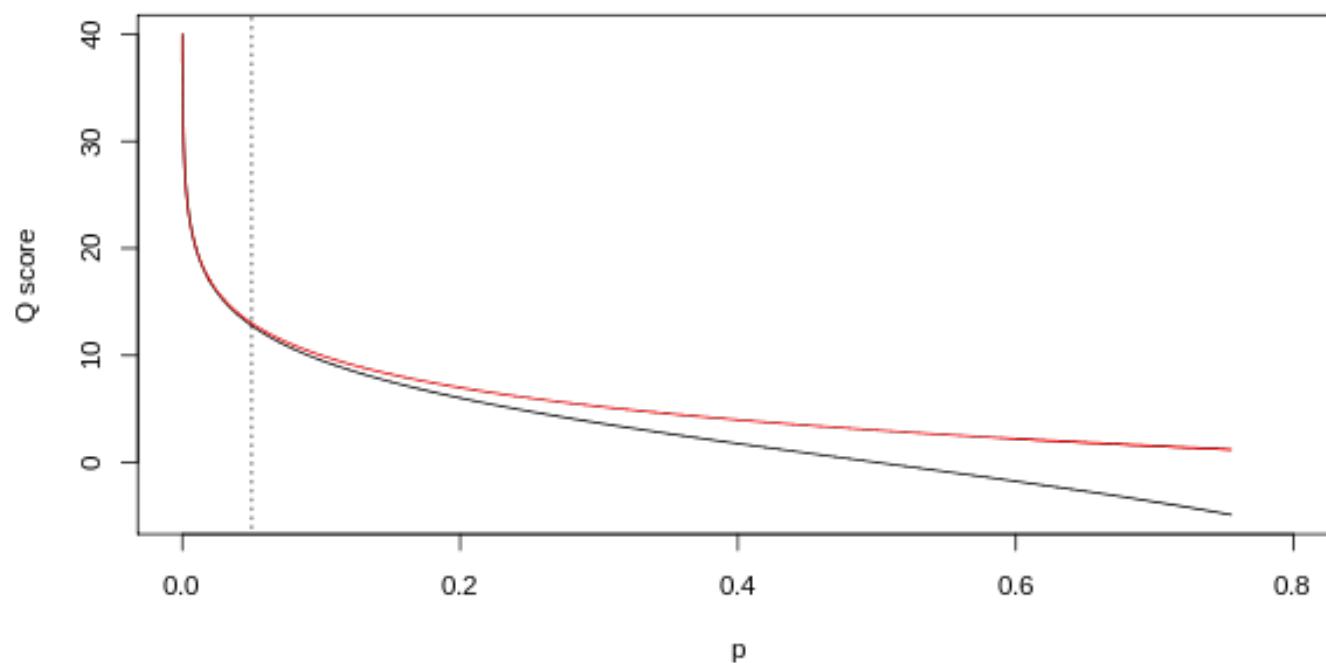
```
@SEQ_ID
GATTGGGGTTCAAAGCAGTATCGATCAAATAGTAAATCCATTGTTCAACTCACAGTT
+
! ' ' * ( ( ( ***+ ) %%++ ) %%%.1***-+* ' ) ) **55CCF>>>>CCCCCCCC65
```

Phred Quality Scores

$$Q_{\text{sanger}} = -10 \log_{10} p$$

$$Q_{\text{solexa-prior to v.1.3}} = -10 \log_{10} \frac{p}{1-p}$$

p corresponds to the probability of the basecall being incorrect; the old Solexa (Illumina) encoding reports the odds that the basecall is incorrect.



Phred Quality Scores – ASCII offset encoding

@SEQ_ID

GATTTGGGTTCAAAGCAGTATCGATAAATGAAATCCATTGTTCAI

+

! ' ' * ((((**+)) % % + +) (% % %) . 1 * * - + * ' ')) * * 55CCF>>>>

Phred score of 0 is encoded by ASCII
Character 33.

What is the phred score of *?

0	<NUL>	32	<SPC>	64	@	96	'	128	Ä	160	†	192	¿	224	‡
1	<SOH>	33	!	65	A	97	a	129	Å	161	°	193	i	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	f	228	%o
5	<ENQ>	37	%	69	E	101	e	133	Ö	165	•	197	≈	229	Â
6	<ACK>	38	&	70	F	102	f	134	Ü	166	¶	198	Δ	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	Ê
9	<TAB>	41)	73	I	105	i	137	â	169	©	201	...	233	È
10	<LF>	42	*	74	J	106	j	138	ã	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	apple
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	í
22	<SYN>	54	6	86	V	118	v	150	ñ	182	ð	214	÷	246	^
23	<ETB>	55	7	87	W	119	w	151	ó	183	Σ	215	◊	247	-
24	<CAN>	56	8	88	X	120	x	152	ò	184	Π	216	ÿ	248	„
25		57	9	89	Y	121	y	153	ô	185	π	217	Ý	249	„
26	<SUB>	58	:	90	Z	122	z	154	ö	186	ſ	218	/	250	.
27	<ESC>	59	;	91	[123	{	155	ø	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		159	ü	191	ø	223	fl	255	„