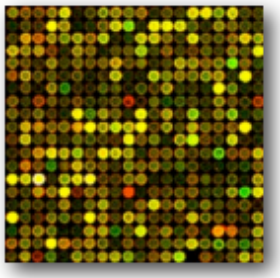
Next generation sequencing

Steven Salzberg

Sanger DNA sequencing 1977-1990s

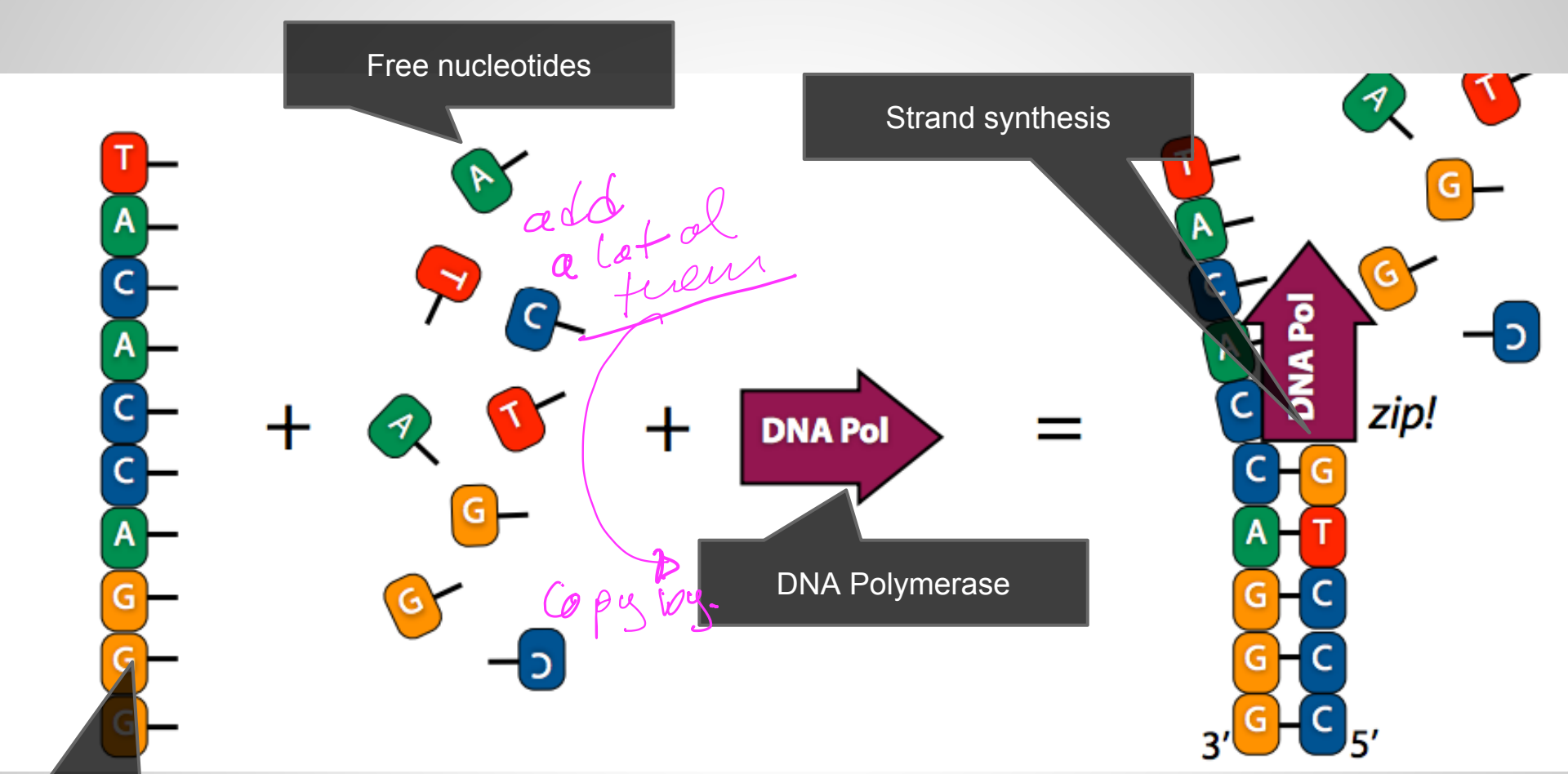


DNA Microarrays Since mid-1990s

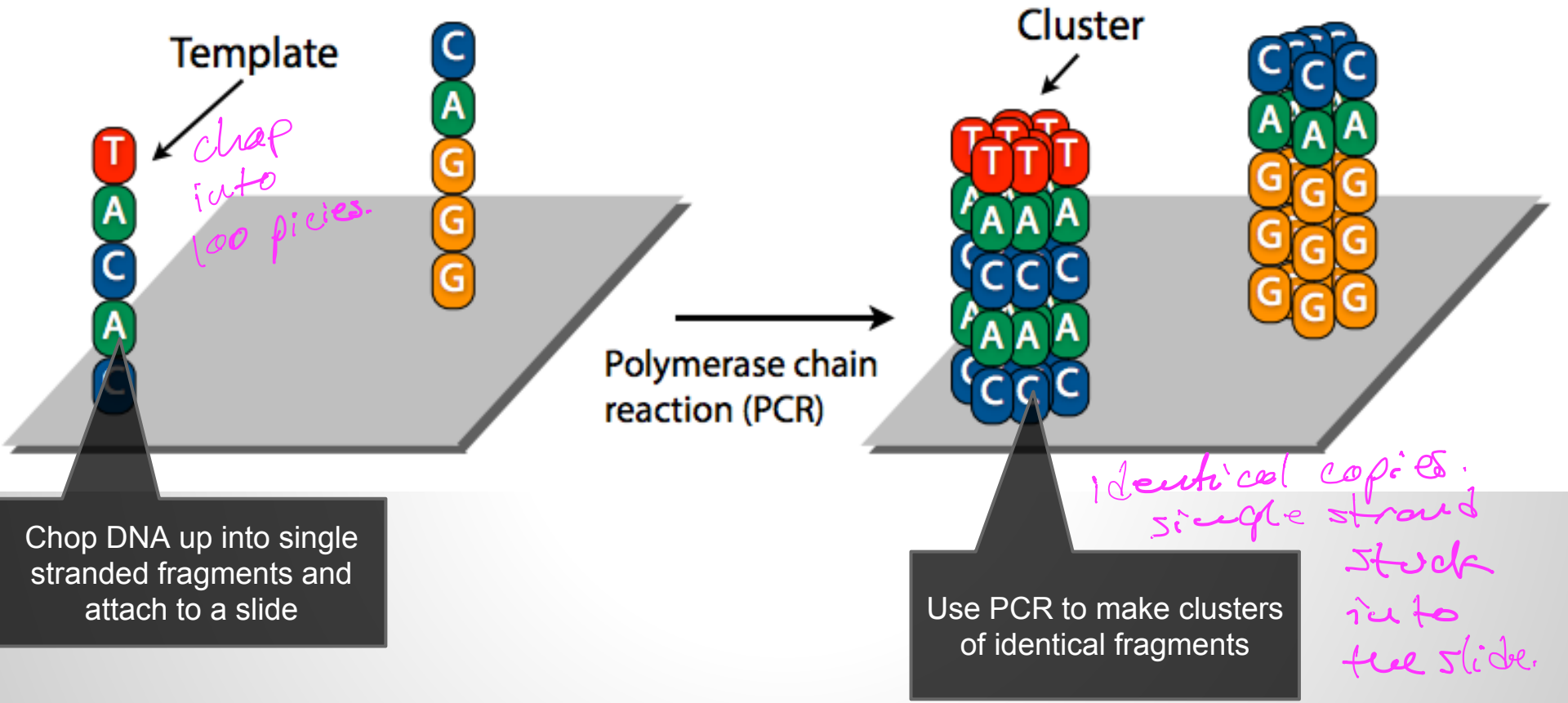
2nd Generation DNA Sequencing Since ~2007

3rd Generation & single molecule Sequencing Since ~ 2010

Slide adapted from Ben Langmead

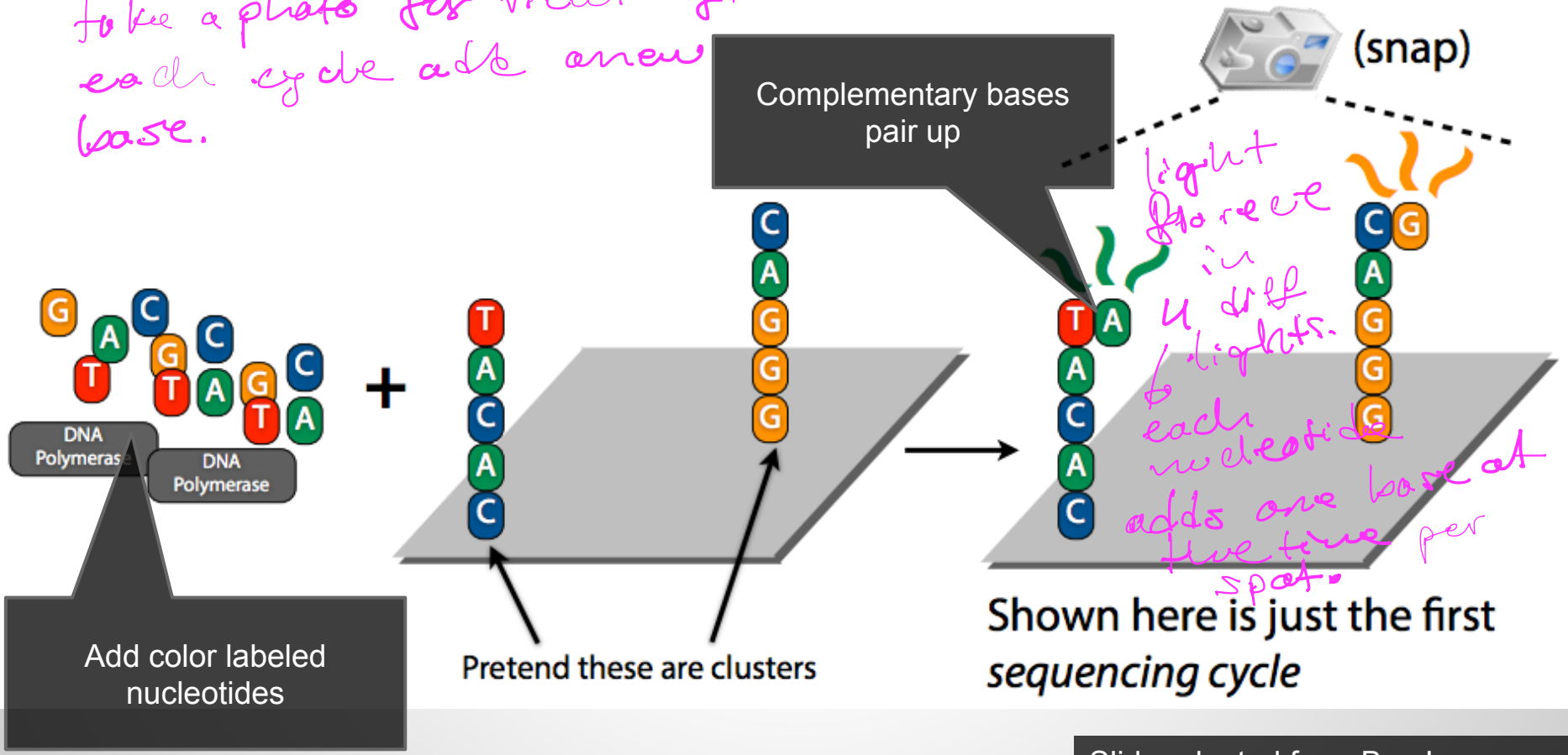


Slide adapted from Ben Langmead

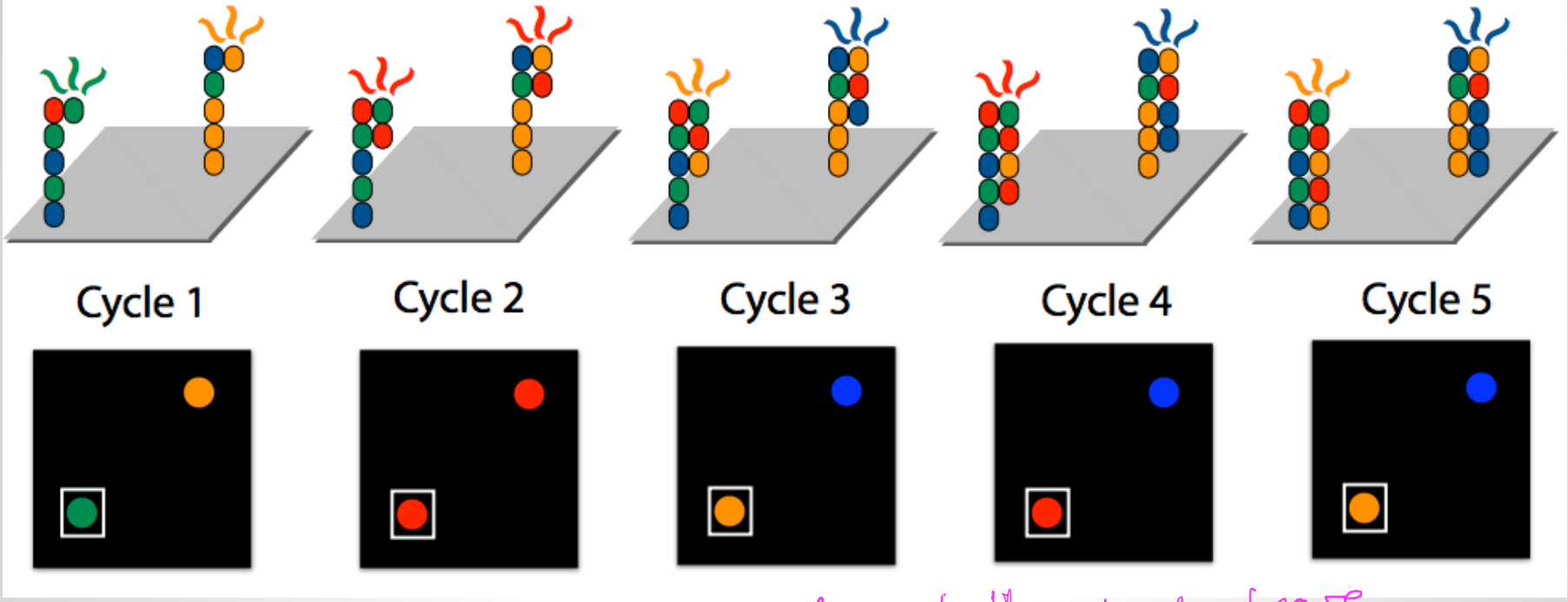


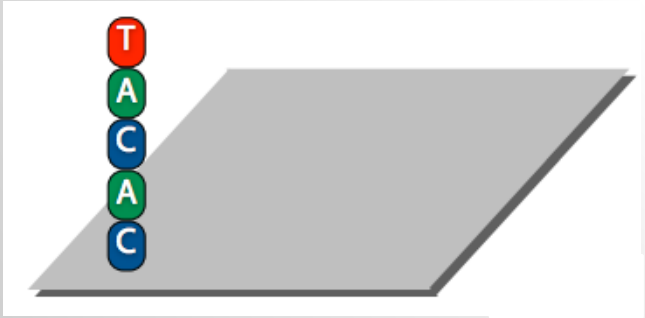
Slide adapted from Ben Langmead

take a photo for viewing. each cycle adds a new base.



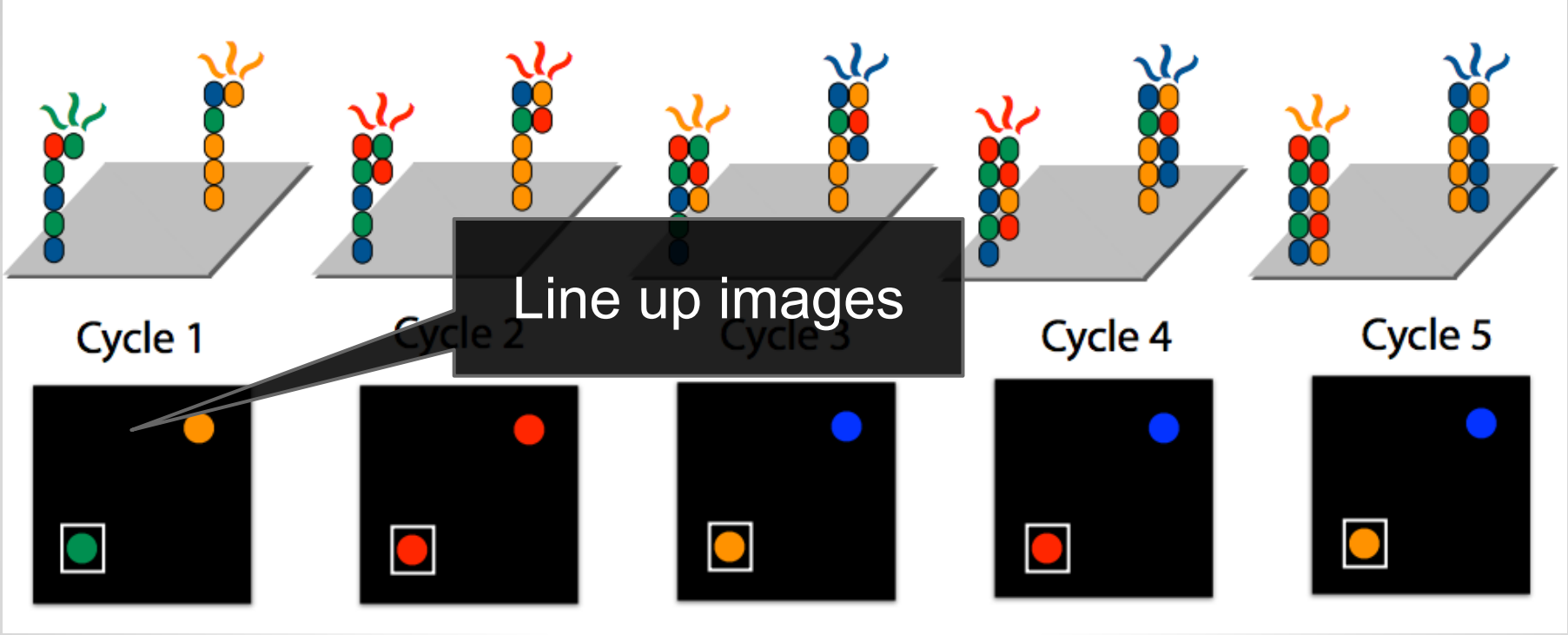
Slide adapted from Ben Langmead

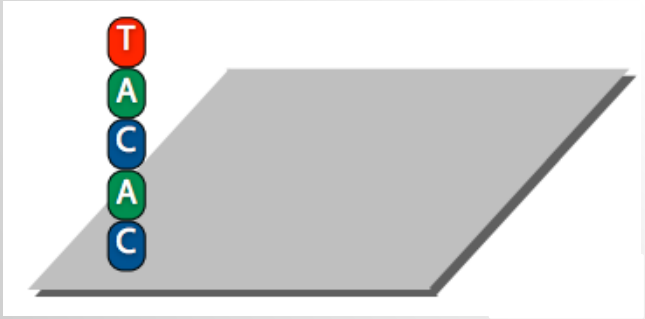




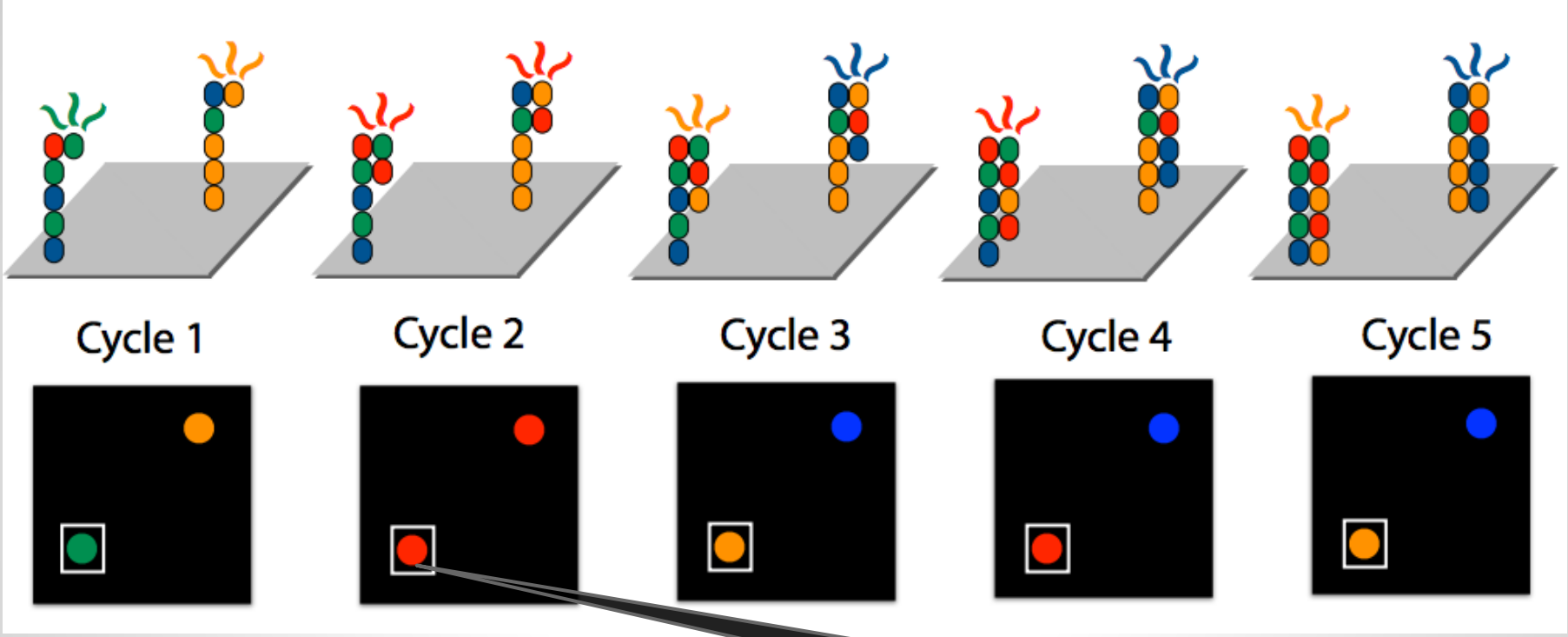
true color tells true base.

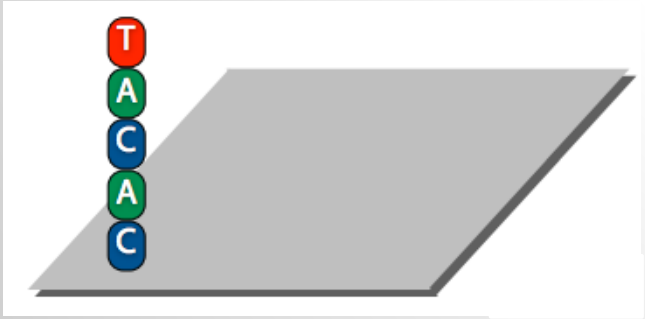
Slide adapted from Ben Langmead





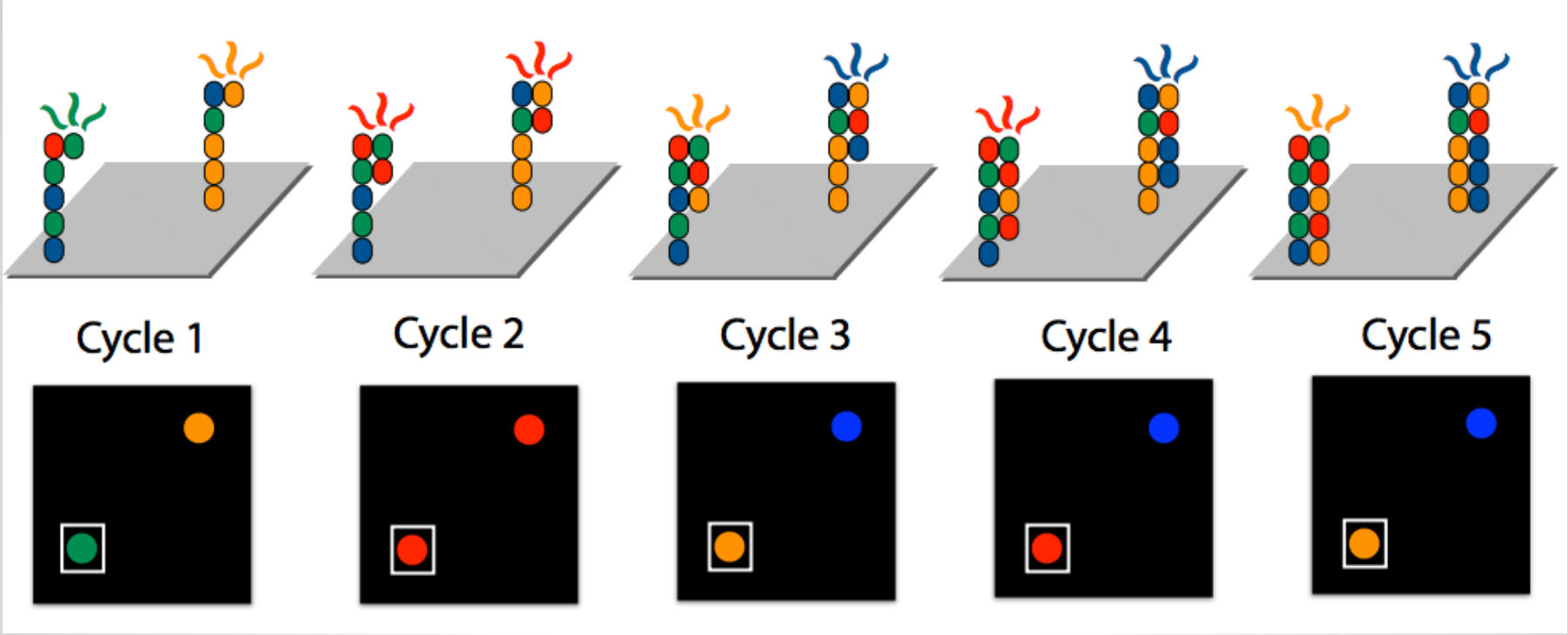
Slide adapted from Ben Langmead

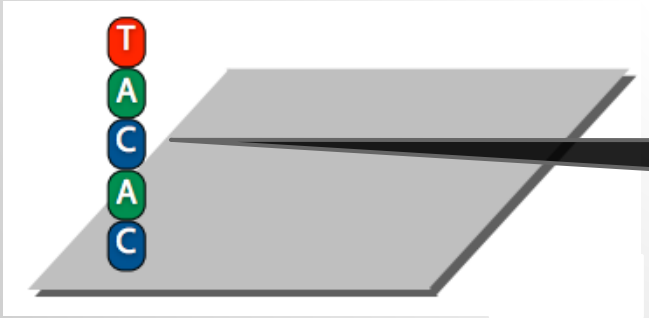




Look at sequence of colors at same spot

Slide adapted from Ben Langmead



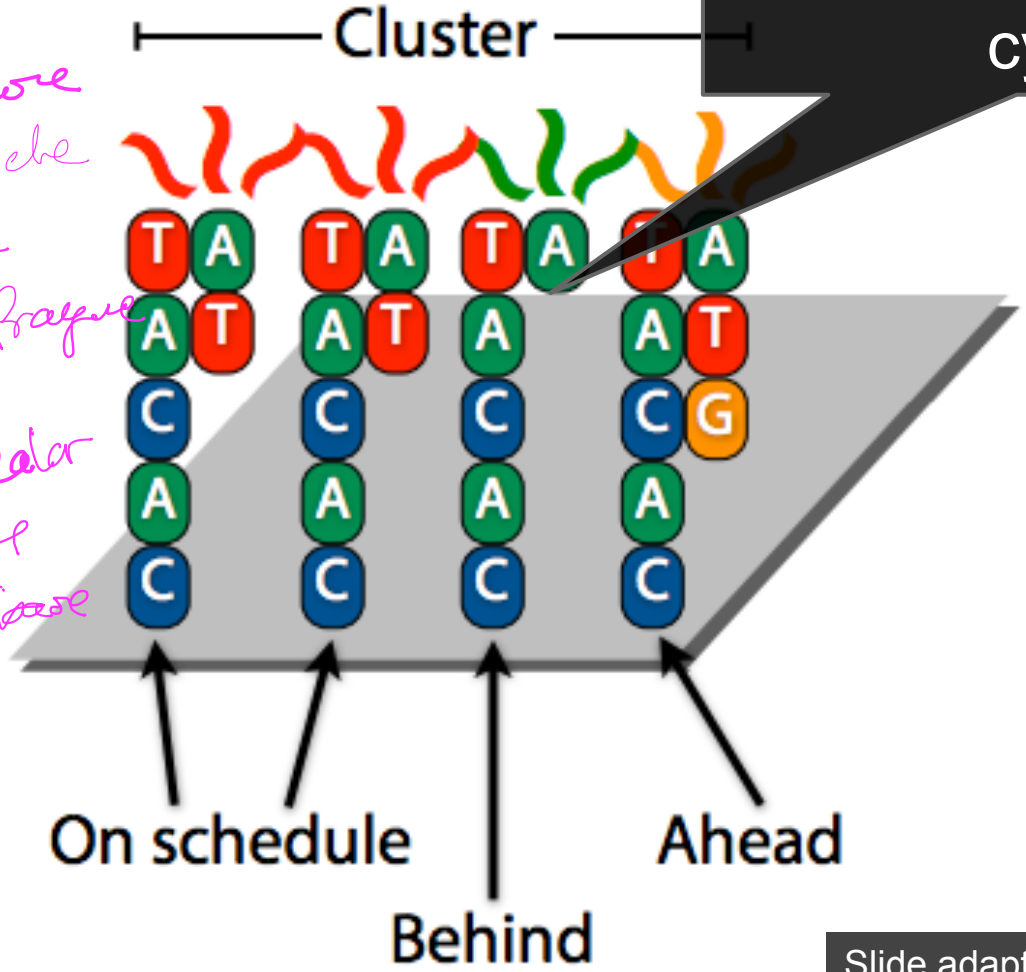


Call complementary bases

Slide adapted from Ben Langmead

error:
they increase in later cycle
polymerase can add a fragment
a few = patches
new diff color
Can't be used for 1000s of base pairs.

Errors increase in later cycles



Slide adapted from Ben Langmead

analyses base on the color intensity.

Read 1	Name
	Nucleotides
	(placeholder)
	Quality values
Read 2	Name
	Nucleotides
	(placeholder)
	Quality values
Read 3	Name
	Nucleotides
	(placeholder)
	Quality values
Read 4	Name
	Nucleotides
	(placeholder)
	Quality values
Read 5	Name
	Nucleotides
	(placeholder)
	Quality values

reads — Example — bash — 104×25

```
$ head -20 SRA_HISEQ2000_FC1.shuffle.2M.1.fastq
@509.6.64.20524.149722
AGCTCTGGTGACCCATGGGCAGCTGCTAGGGAGCCTTCTCTCCACCCTGAAAATAGCTTCTGGCTGNTGGGTGAACTATGGAGAGAAAGCGTTTTATTAT
+
HHHHHHHGHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHIIHHIHFHHFHHHIHGEHHIIFIHBC#@:@9,--541436D9?;E###############
@509.4.62.19231.2763
GTTGATAAGCAAGCATCTCATTTTGTGCATATACCTGGTCTTTCGTATTCTGGCGTGAAGTCGCCGNCTGAATGCCAGCAATCTCTTTTTGAGTCTCATT
+
HHHHHHHHHHHHHHHEHHHHHHHHHHHHHHHHHHHHHHHHDHHHHHHGHGHHHHHHHHHH=EF?DHE4#555=;===GGHEGGEGHG@C@<7<3@?F<A9@<
@509.6.47.3027.76579
CCTTTTCGACTAGAGACTGCCAAGTGCCAAAATATCCACTTGCAGATACTACAACAAGAGTGTTTCNAAACTGCTCAATCAAAAGAAATGTTCAACTCTT
+
HHHHHHHHHHHHHHHHHHHHIHHHHHHHHHHHHHHHHHHHGHHHHHHHHGHHHHHHHHHHEH?HH4#554DDADDHHHHHH@GHHFGBBFFHFHFHEHHH
@509.2.7.2951.186312
AAAGATACAACATACCACAATCTTTGAGACACACCTAAGACAATAAGGCAGTGTTAAGAGGAAAATTAATAGCACTAAATGCCCACATCAAAAAGTTAGA
+
HHHHHHHHHHHHHHHHHHHGHDHHHHHFHEHHHGHHGHHHHHHHHHGHHHHHHEHEF<?<@=BBFFFGCFFE?<;@AFG=GA;@D@D?FDFFB=B;F=>AA@
@509.6.25.8102.140546
GGACACATTCAAACCATTGCATCCATCCTCTGCATTCAGAAAGATAGTCCAACAGAAAGATCTGGANTCAAGAGACCCAGCTGATTACCAATTCCAGTTT
+
HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHIHHHHHHHIHHHIHHHIHG#FFDCDD@@GGGHHFIHEGIFIEIIIIGIIGFGF
$
```

Slide adapted from Ben Langmead