CSC 334: TOPICS IN COMPUTATIONAL BIOLOGY

"Algorithms for Genomic Data"

Fall 2015

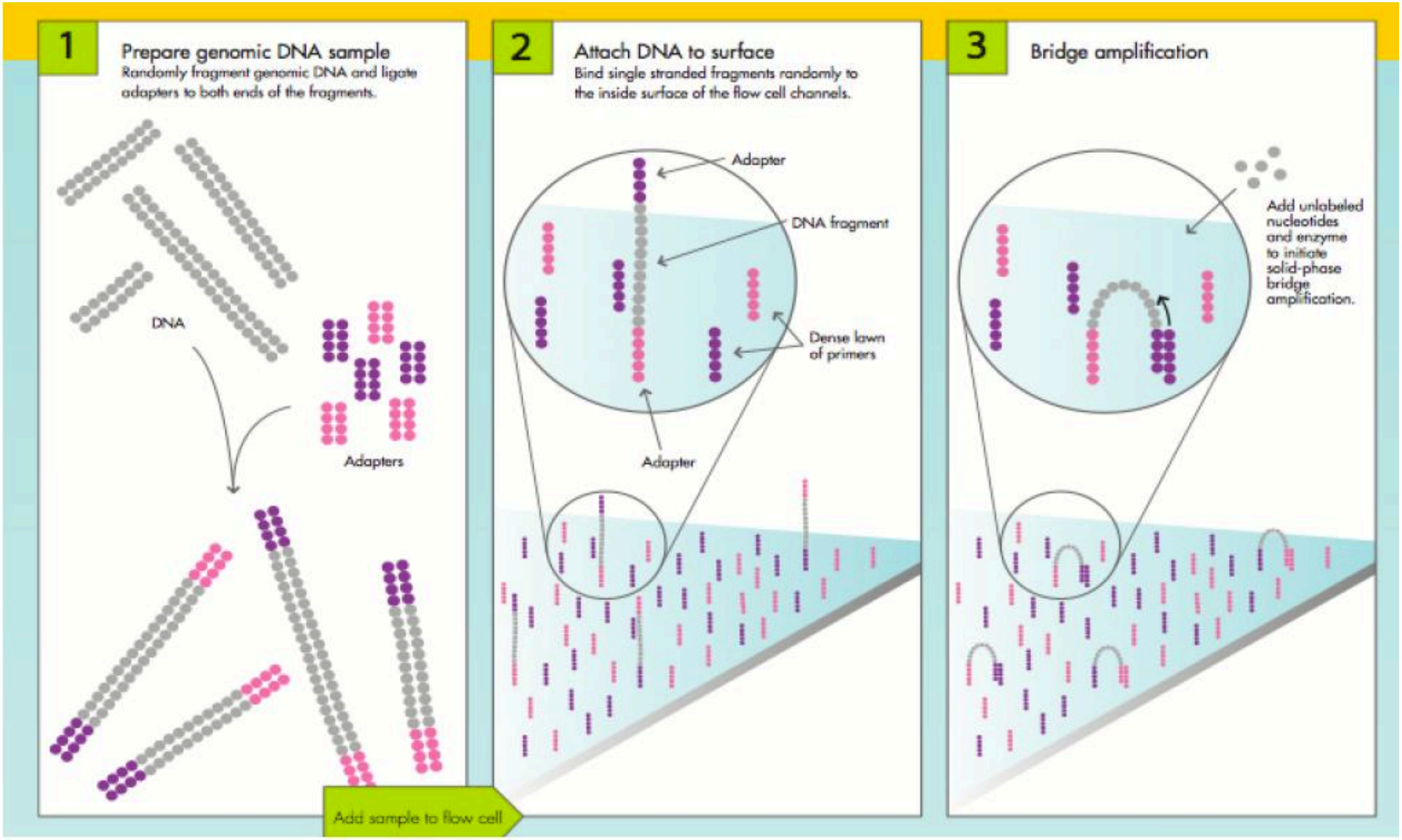
Smith College

Instructor: Prof. Sara Sheehan

Outline: 9/11

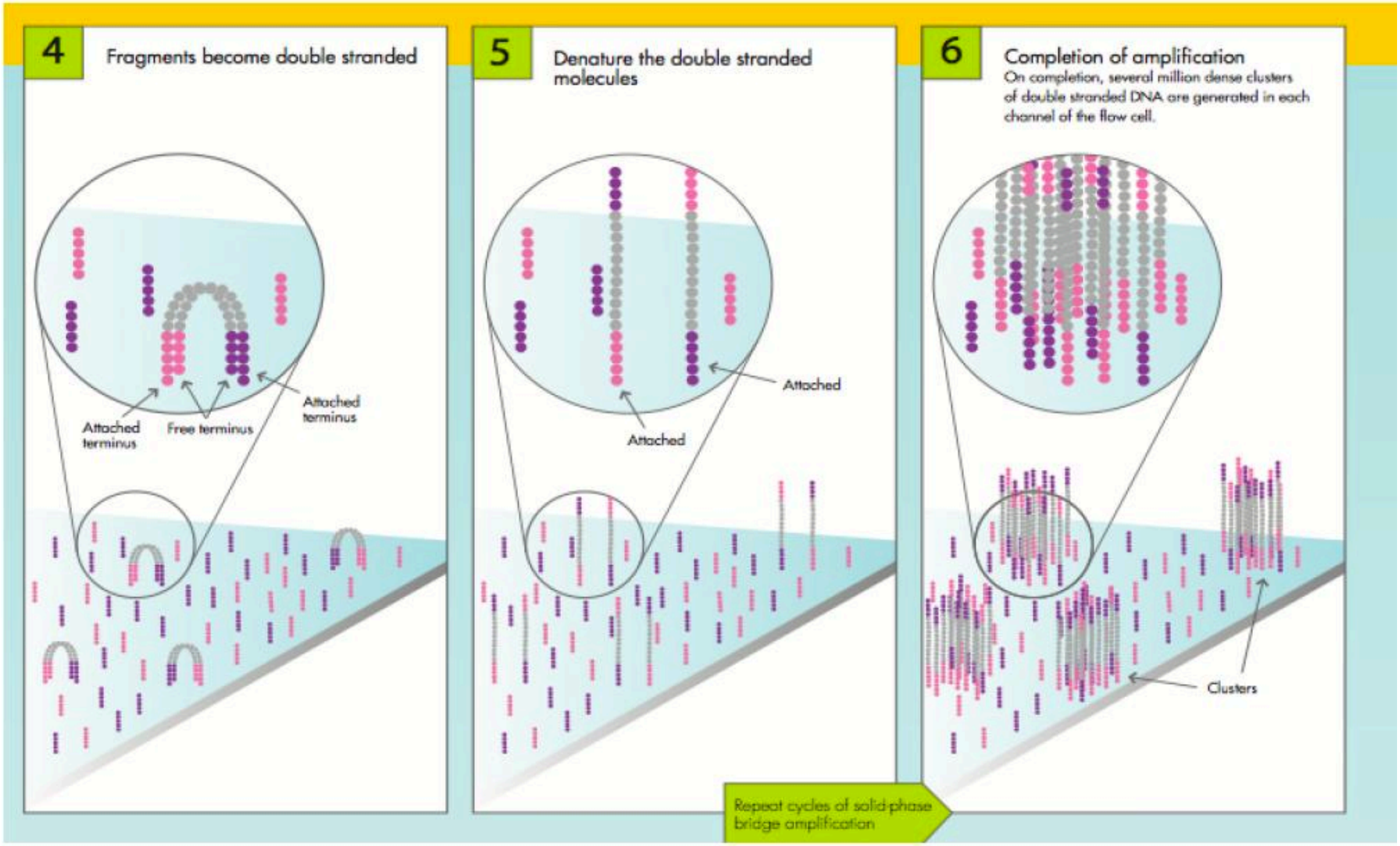
- Next-generation sequencing
- String alignment

Next-generation sequencing 1-3



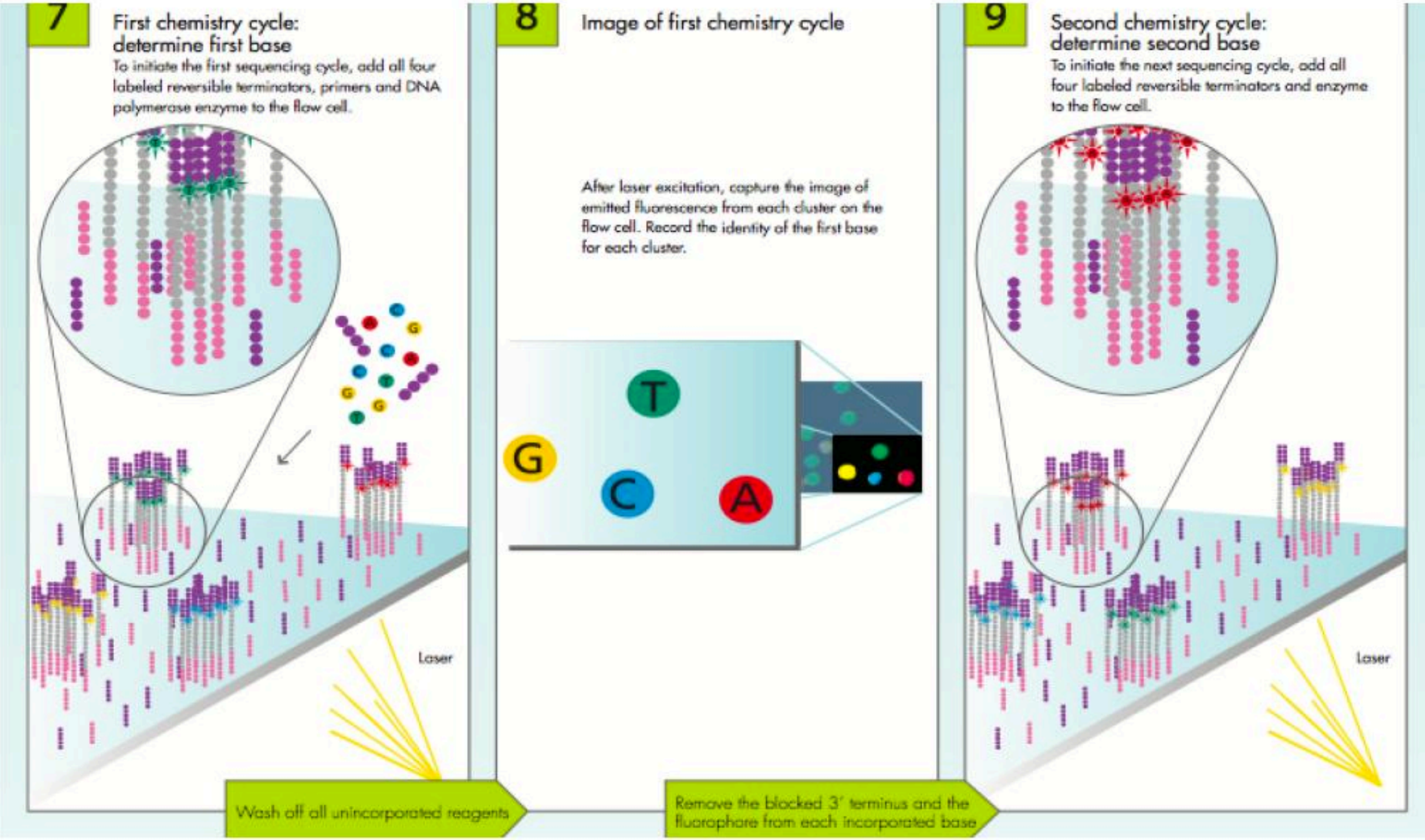
Slides by: Brian Cunningham, NanoBio Node

Next-generation sequencing 4-6



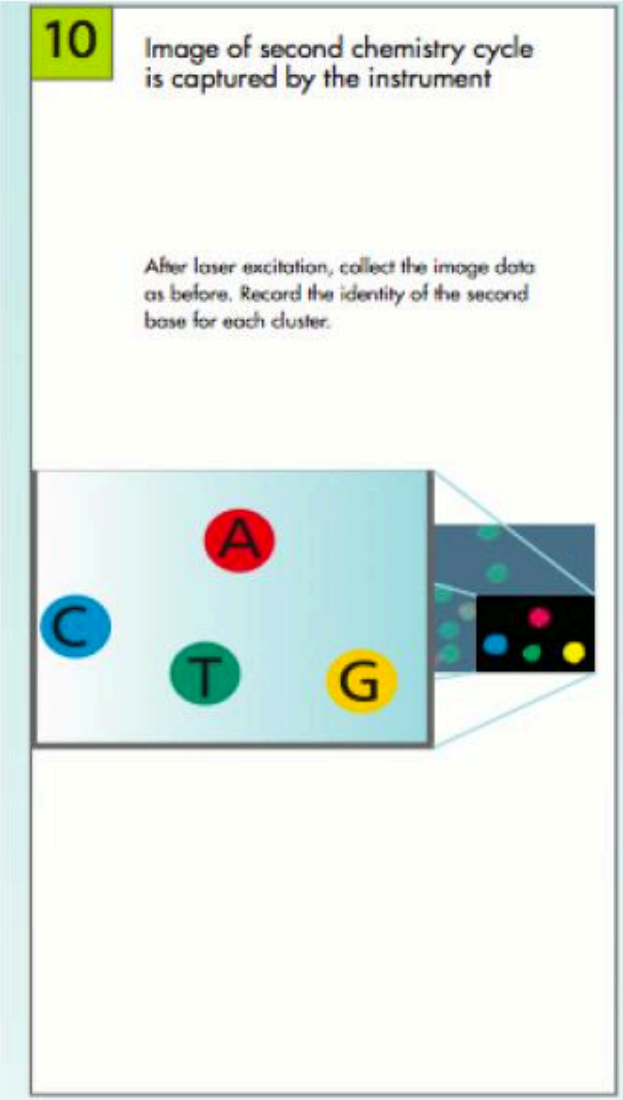
Slides by: Brian Cunningham, NanoBio Node

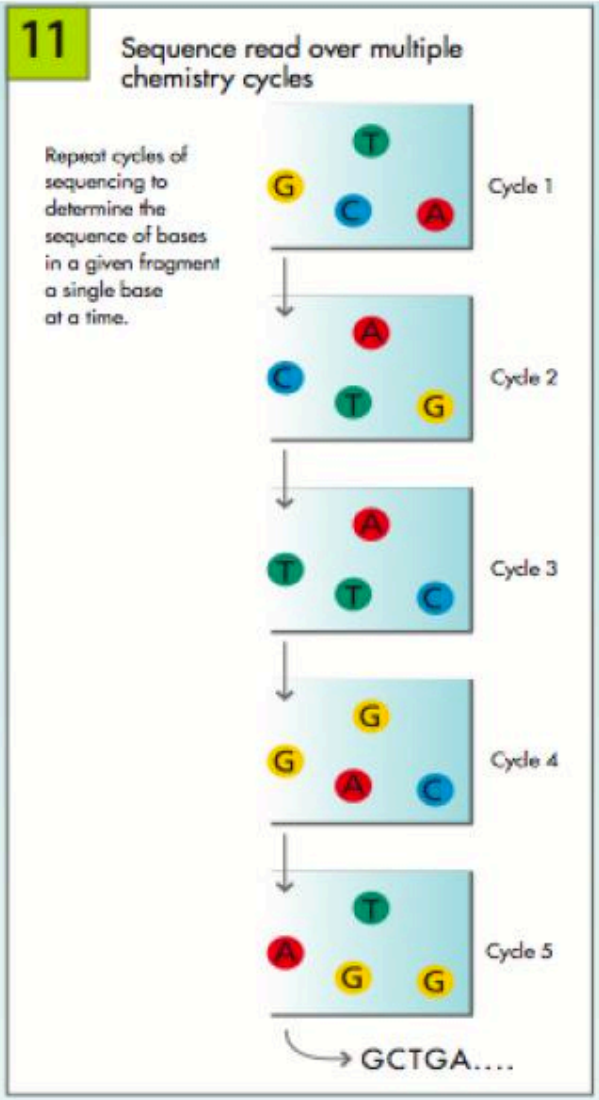
Next-generation sequencing 7-9

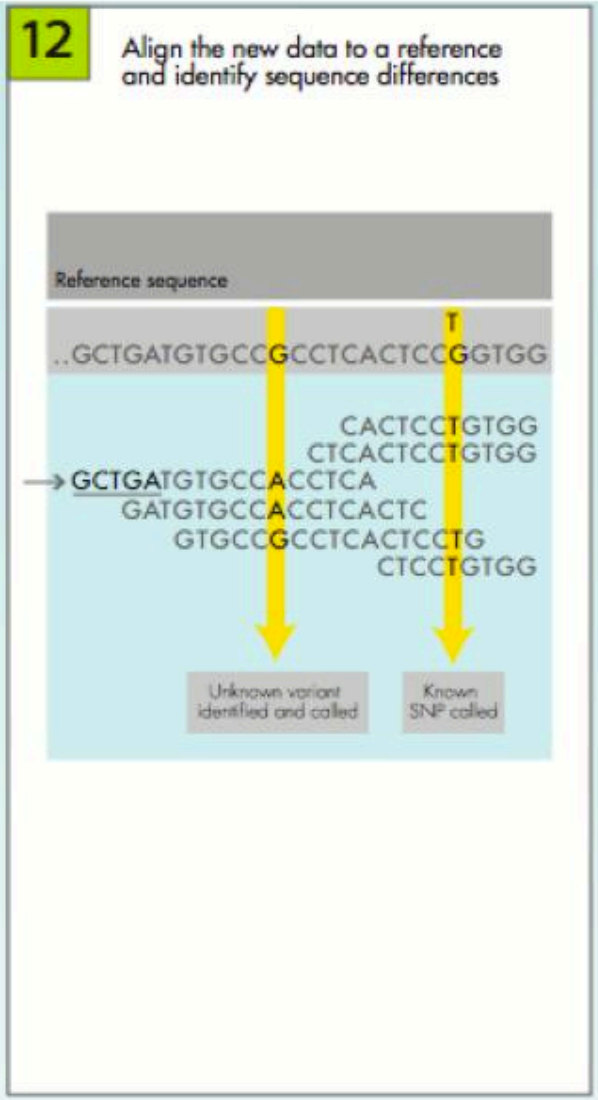


Slides by: Brian Cunningham, NanoBio Node

Next-generation sequencing 10-12







Slides by: Brian Cunningham, NanoBio Node

String Alignment

Example

- ACGGCTAGTTACG

- TCGTAGTATACCGA

- How should we line them up to get the best overlap?

Example alignments: human, chimp, macaque

```
                                Start
Human       GTTTTGAGG---ATGTTCAACAAATGCTCCTTTCATTCCTCTATTTACAGACCTGCCGCA
Chimpanzee  GTTTTGAGG---ATGTTCAATAAATGCTGCTTTCACTCCTCTATTTACAGACCTGCCGCA
Macaque     GTTTTGAGG---ATGCTCAATAAATGCTCCTTTCATTCCTCCATTTACAAACTTGCCGCA

Human       GACAATTCTGCTAGCAGCCTTTGTGCTATTATCTGTTTTCTAAACTTAGTAATTGAGTGT
Chimpanzee  GACAATTCTGCTAGCAGCCTTTGTGCTATTATCTGTTTTCTAAACTTAGTAATTGAGTGT
Macaque     GACAATTCTGCTAGCAGCCTTTGTGCTATTATCTGTTTTCTAAACTTAGTAATTGAGTGT
                                  ↓
Human       GATCTGGAGACTAA-CTCTGAAATAAATAAGCTGATTATTTATTTATTTTCTCAAAACAA
Chimpanzee  GATCTGGAGACTAAACTCTGAAATAAATAAGCTGATTATTTATTTATTTTCTCAAAACAA
Macaque     TATCTGGAGACTAAACTCTGAAATAAATAAGCTGATTATTTATTTATTTTCTCAAAACAA

Human       CAGAATACGATTTAGCAAATTACTTCTTAAGATATTATTTTACATTTCTATATTCTCCTA
Chimpanzee  CAGAATACGATTTAGCAAATTACTTCTTAAGATACTATTTTACATTTCTATATTCTCCTA
Macaque     CAGAATATGATTTAGCAAATTACCTCTTAAGATATTATTTTGCACTTCTATATTCTCCTA

Human       CCCTGAGTTGATGTGTGAGCAATATGTCACTTTCATAAAGCCAGGTATACA----TTATG
Chimpanzee  CCCTGAGTTGATGTGTGAGCCGTATGTCACTTTCATAAAGCCAGGTATACA----TTATG
Macaque     CCCTGAGTTGATGTGTGAGCAATATGTCACTTCCACAAAGCCAGGTATATATACATTACG
                                H   I   I   Y   S   T   F   L   S   K
Human       GACAGGTAAGTAAAAAACATATTATTTATTCTACGTTTTTGTCCAAAAATTTTAAATTTC
Chimpanzee  GACAGGTAAGTAAAAAACATATTATTTATTCTACGTTTTTGTCCAAGAATTTTAAATTTC
Macaque     GACAGGTAAGTAAAAA-CATATTATTTATTCTAGGTTTTTGTCCAAGAGTTTTAAATTTC

Human       AACTGTTGCGCGTGTGTTGGTAA---TGTAAAACAAACTCAGTACA
Chimpanzee  AACTGTTGCGCGTGTGTTGGTAA---TGTAAAACAAACTCAGTACA
Macaque     AACTGTTGTGCATGTGTTGGTAA---CGTAAAACAAATTCAGTACG
```

Genome Research, 2009