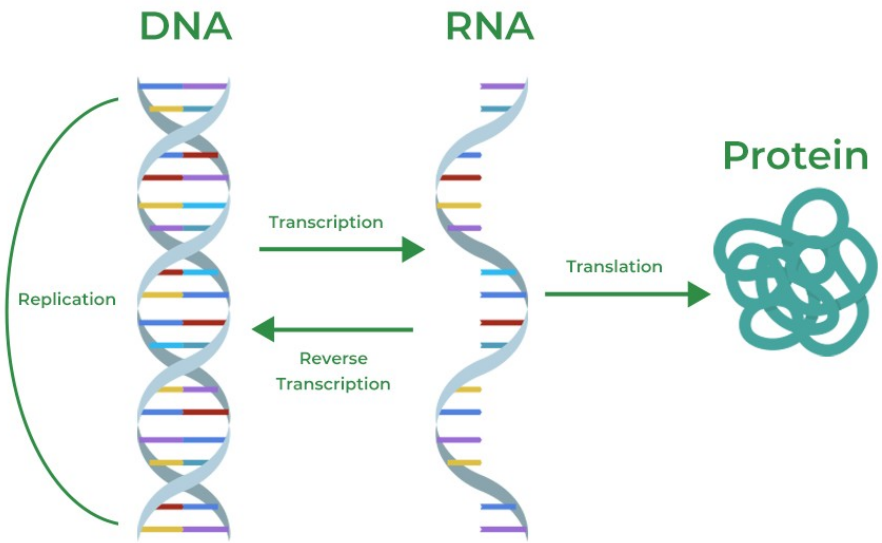
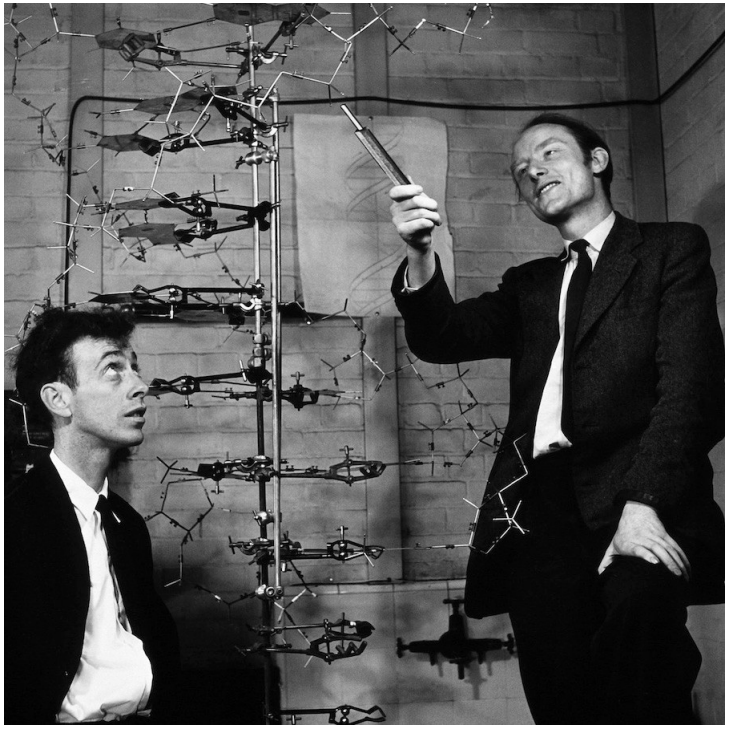
RNA Sequencing

Where do the numbers come from?

The Central Dogma of Mol Biol

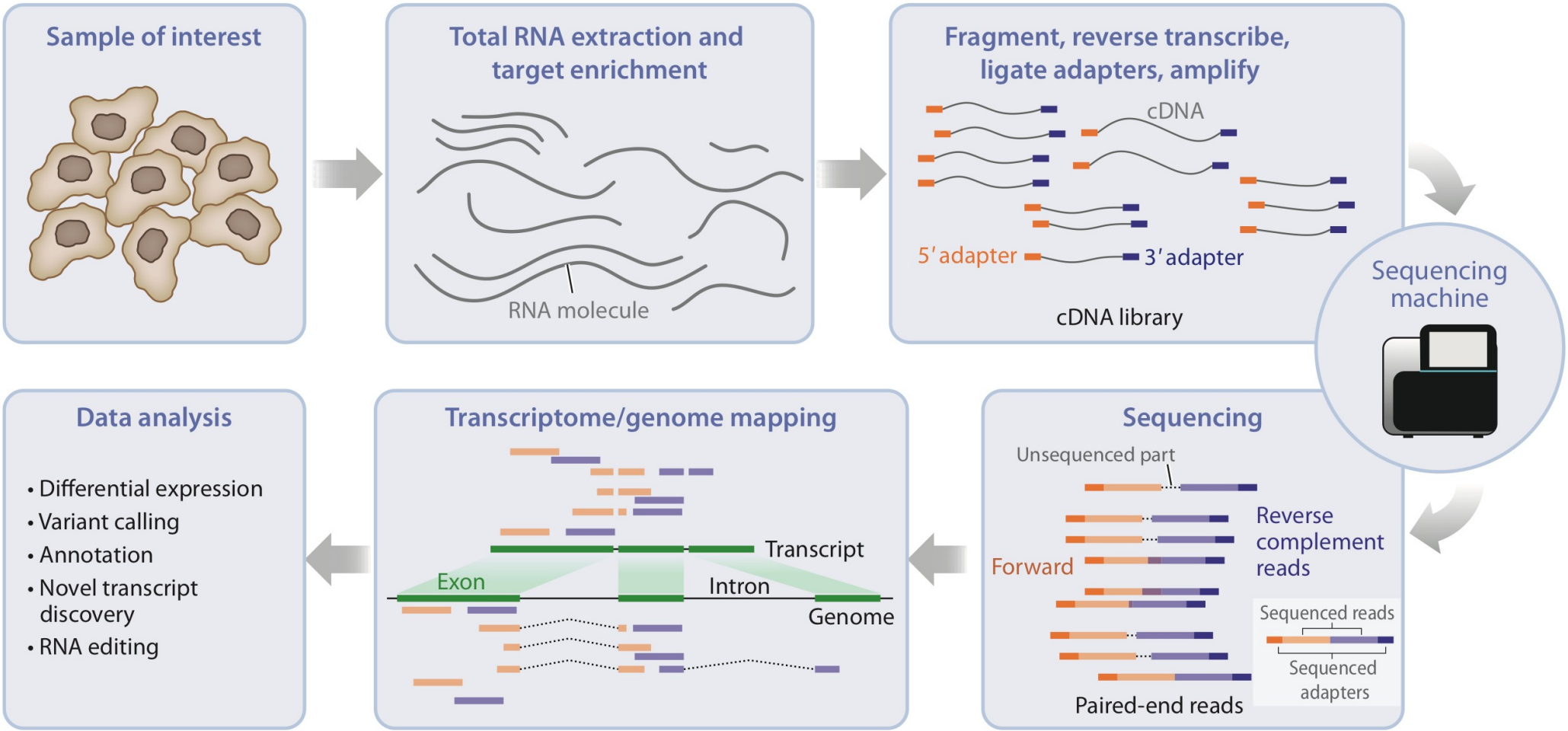


https://www.geeksforgeeks.org/biology/central-dogma-steps-guide/



https://www.sciencehistory.org/education/scientific-biographies/francis-crick-rosalind-franklin-james-watson-and-maurice-wilkins/

General Overvidew

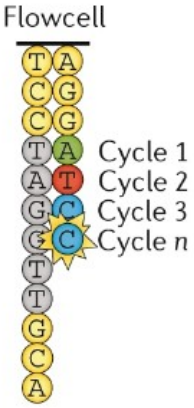


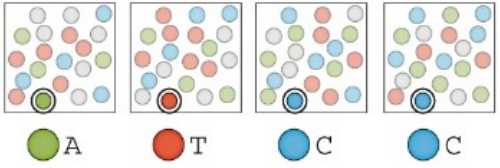
https://statomics.github.io/SGA/sequencing_intro.html

Current Technology

Short-reads

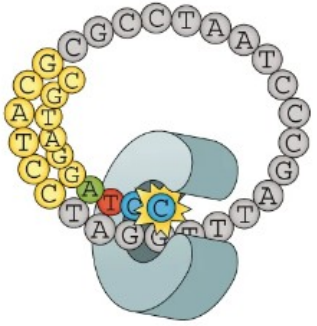
Illumina

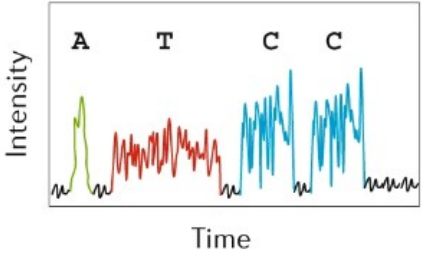




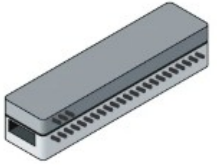
Long-reads

Pacific Biosciences

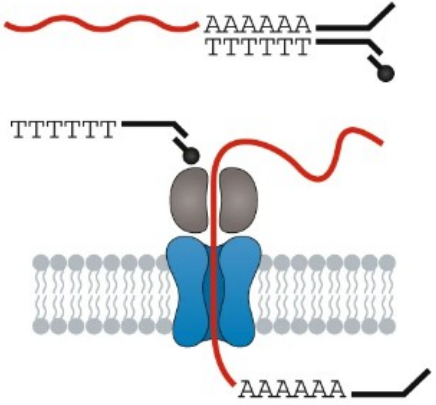


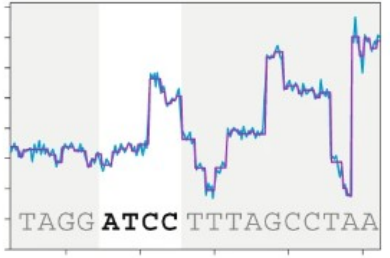


Direct-reads



Oxford Nanopore



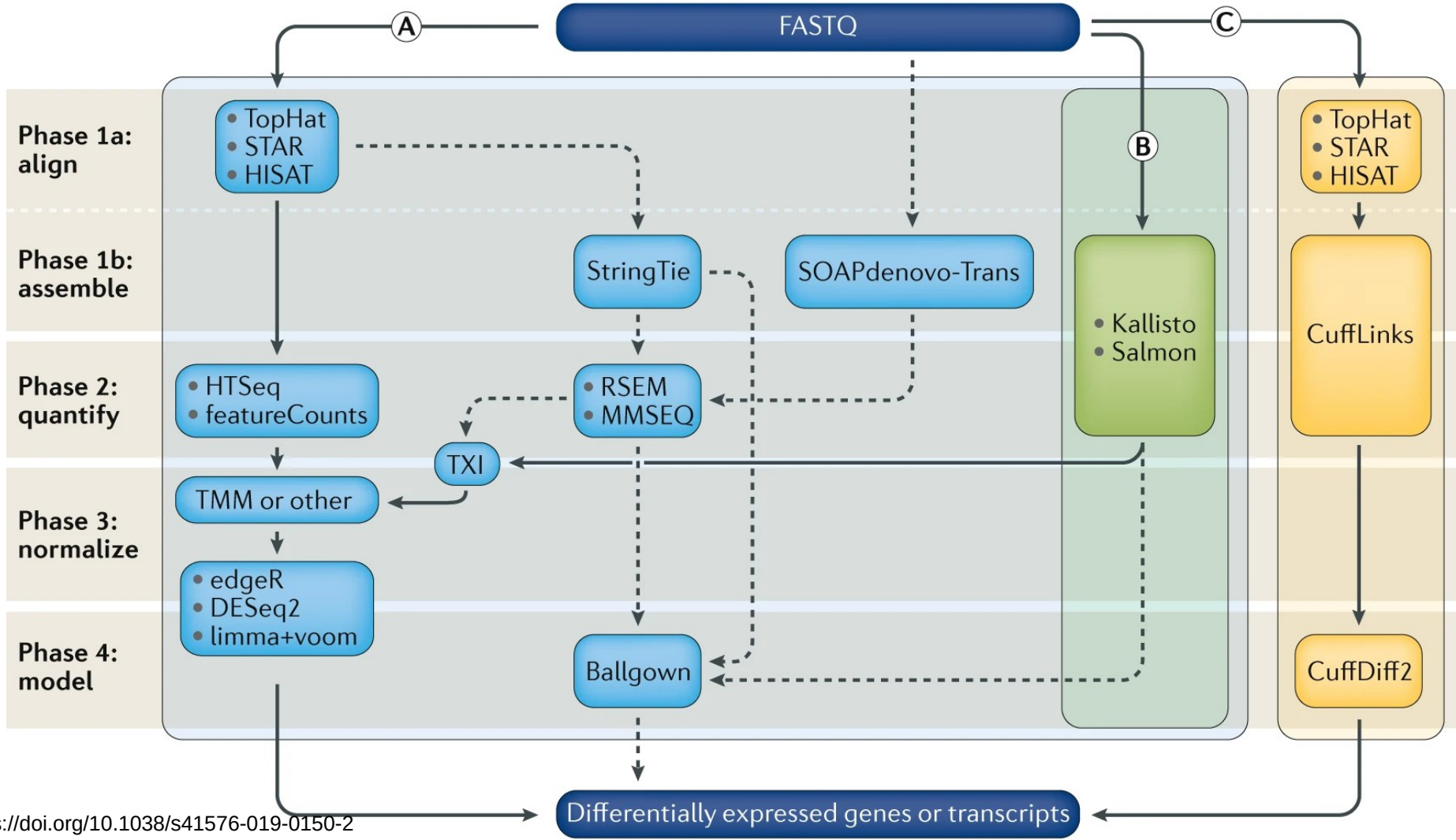


https://doi.org/10.1038/s41576-019-0150-2

FASTQ Files

```
(base) ilirsheraj@ilirsheraj:~/SRA/fastq$ zless SRR6809353_pass.fastq.gz | head -20
@SRR6809353.1.1 1 length=100
GGGGCGGGGGGTTTGTAGAGAAAGGTGACAAATTTCAGTACCTCTGGCATGCTGTCCCAGGAAACTAGGGCTCCCACTAACCTATGAGGTTTTTAAACAC
+SRR6809353.1.1 1 length=100
BBBBBFFFFFFBBFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFBFFFFFFFBB<FFFFFFFFFFFFFF
@SRR6809353.2.1 2 length=100
ATCGTATGGGGTGGTTTTTGCATTAAATCCTGGGGTCCATTTTACAATCCATTATTTTTGACCACTGCTATGTGTTCAAGTAGTATGAGAATGTGATTGT
+SRR6809353.2.1 2 length=100
BBBBBFFFFFFFFFBFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFBFFF<FFFF/FFBFBBFFFFFF/F7BBFFFFFFFFFFB
@SRR6809353.3.1 3 length=100
GGGTCCGAATCGACCATAAAGGGTGTAGGGGCCACCTCCTCCCCCTGTTCTGTTGGGGAGGGGTAGCCATGATTTGTCCCAGCCTGGGGCTCCCTCTCTG
+SRR6809353.3.1 3 length=100
BBBBBFFFFFFFFFFFFFFFFFFFFBFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFBFFFFFBBFFFFFFFFFFFFFFFFFFFFFFFFBFFB7F/7</
@SRR6809353.4.1 4 length=100
GGGGGTGACAGGGCACGAGTTCGAGGCCAGCCTGGTCCACACGGGTCGGAAAAAAAAAAAAAAAAAAAATCGGAAGAGCACACGTCTGAACCCCAGTCC
+SRR6809353.4.1 4 length=100
BBBBBFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF<FFFFFFFFFFFFFFFFFFFFF/7<77<BB///7///77//BBBBB/7BB//7/
@SRR6809353.5.1 5 length=100
ATAAAGCCACTGGGGACGAGACAGGTGCTAAAGTTGAACGAGCTGATGGATATGAACCACCAGTCCAAGAATCTGTTTAAAGTTCAGACTTCAAATAGTG
+SRR6809353.5.1 5 length=100
BBBBBFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFBFFFFFFFFFFFBFFFFFFFFFFFFFFFBFFBFFFF
(base) ilirsheraj@ilirsheraj:~/SRA/fastq$
```

Computational Pipeline



https://doi.org/10.1038/s41576-019-0150-2

STAR Sequence Alignment

https://hbctraining.github.io/Intro-to-rnaseq-hpc-O2/lessons/03_alignment.html

This requires HPC Environment! Do not try on your laptop :)

Step 1: Create Genome Index

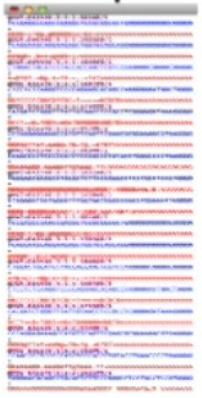
```
STAR --runThreadN 6 \
--runMode genomeGenerate \
--genomeDir chr1_hg38_index \
--genomeFastaFiles /n/groups/hbctraining/intro_rnaseq_hpc/reference_data_ensembl38/Homo_sapiens.GRCh38.dna.chromosome.1.fa \
--sjdbGTFfile /n/groups/hbctraining/intro_rnaseq_hpc/reference_data_ensembl38/Homo_sapiens.GRCh38.92.gtf \
--sjdbOverhang 99
```

Step 2: Sequence Alignment

```
STAR --genomeDir /n/groups/hbctraining/intro_rnaseq_hpc/reference_data_ensembl38/ensembl38_STAR_index/ \
--runThreadN 6 \
--readFilesIn Mov10_oe_1.subset.fq \
--outFileNamePrefix ../results/STAR/Mov10_oe_1_ \
--outSAMtype BAM SortedByCoordinate \
--outSAMunmapped Within \
--outSAMattributes Standard \
--quantMode GeneCounts
```

RNA-seq

Sample A



Align

```
GTCGCAGTANCTGTCT
|||||||||  ||||||
GTCGCAGTATCTGTCT

GGATCTGCGATATACC
||||||  |||||||||
GGATCT-CGATATACC

ATATATATATATATAT
||||||||||||||||
ATATATATATATATAT

TCTCTCCCANNAGAGC
|||||||||  |||||
TCTCTCCCAGGAGAGC
```

Aggregate

```
GTCGCAGTATCTGTCT
GTCGCAGTATCTGTCT
GTCGCAGTATCTGTCT
GTCGCAGTATCTGTCT
GTCGCAGTATCTGTCT
 TGTCGCAGTATCTGTC
  TATGTCGCAGTATCTG
  TATATCGCAGTATCTG
  TATATCGCAGTATCTG
  TATATCGCAGTATCTG
    CCCTATATCGCAGTAT
     GCACCCTATGTCGCA
     CACCCTATATCGCA
     AGCACCCTATGTCGCA
      GAGCACCCTATGTCGC
      CCGGAGCACCCTATAT
      CCGGAGCACCCTATAT
       GCCGGAGCACCCTATG
```

Statistics

Gene 1 differentially expressed?

TTTGGTATTTTCGTCTGGGGGGTATGCACGCGATAGCATTGCGAGACGCTGGAGCCGGAGCACCCTATGTCGCAGTATCTGTCTTTGATTCCTGCCTCATCCTATTATTTATCG…

Gene 1

Sample B

Align

```
GTCGCAGTANCTGTCT
|||||||||  ||||||
GTCGCAGTATCTGTCT

GGATCTGCGATATACC
||||||  |||||||||
GGATCT-CGATATACC

ATATATATATATATAT
||||||||||||||||
ATATATATATATATAT

TCTCTCCCANNAGAGC
|||||||||  |||||
TCTCTCCCAGGAGAGC
```

Aggregate

```
 TGTCGCAGTATCTGTC
  AGCACCCTATGTCGCA
GCCGGAGCACCCTATG
```

HARVARD
SCHOOL OF PUBLIC HEALTH

Slide courtesy of Ben Langmead