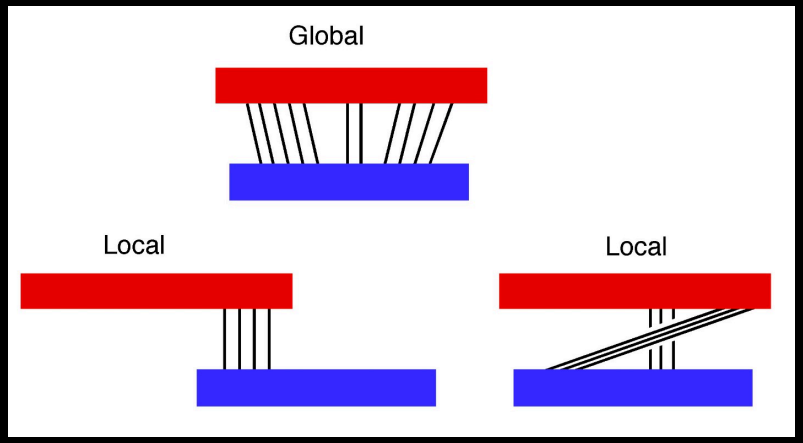
Blast and NGS mapping

Sequence alignment

- One of the core technique in bioinformatics

- It can perform in two ways in general
  - Global alignment (Needleman-Wunsch algorithm)
    - Align every residue in two or multiple sequences
  - Local alignment (Smith-Waterman algorithm)
    - Align local fragments with high score



http://www.pitt.edu/~mcs2/teaching/biocomp/tutorials/global.html

Global alignment, Needleman–Wunsch algorithm

- A end-to-end alignment
- Dynamic programming
- Example,
  - match = +5
  - mismatch = -2
  - insertion = -6
  - TGCTCGTA
  - TTCATA

$$H_{ij} = \max \begin{cases} H_{i-1,j-1} + s(a_i, b_j), \\ \max_{k \geq 1}\{H_{i-k,j} - W_k\}, \\ \max_{l \geq 1}\{H_{i,j-l} - W_l\}, \\ 0 \end{cases} \qquad (1 \leq i \leq n, 1 \leq j \leq m)$$

Go right: insert a gap in the column sequence
Go down: insert a gap in the row sequence

		T	G	C	T	C	G	T	A
	0	-6	-12	-18	-24	-30	-36	-42	-48
T	-6	5	-1	-7	-13	-19	-25	-31	-37
T	-12	-1	3	-3	-2	-8	-14	-20	-26
C	-18	-7	-3	8	2	3	-3	-9	-15
A	-24	-13	-9	2	6	0	1	-5	-4
T	-30	-19	-15	-4	7	4	-2	6	0
A	-36	-25	-20	-10	1	5	2	0	11

Traceback: Begin with the **cell at the lower right of the matrix**, end at top left cell

```
T G C T C G T A
|     | |   | |
T - - T C A T A
```

http://telliott99.blogspot.hk/2009/08/alignment-needleman-wunsch.html

Local alignment, Smith-Waterman algorithm

- Dynamic programming

- Assumed we have two sequences
  - AGACTAGTTAC and
  - CGAGACGT
  - mismatch = -3
  - insertion = -5
  - Match: GG=7;AA=10;CC=9;TT=8

	_	A	G	A	C	T	A	G	T	T	A	C
_	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-50	-55
C	-5	-3	-8	-13	-6	-11	-16	-21	-26	-31	-36	-41
G	-10	-6	4	-1	-6	-9	-12	-9	-14	-19	-24	-29
A	-15	0	-1	14	9	4	1	-4	-9	-14	-9	-14
G	-20	-5	7	9	9	6	3	8	3	-2	-7	-12
A	-25	-10	2	17	12	7	16	11	6	1	8	3
C	-30	-15	-3	12	26	21	16	11	11	6	3	17
G	-35	-20	-8	7	21	23	20	25	20	15	10	12
T	-40	-25	-13	2	16	29	24	20	33	28	23	18

Traceback: Begin with the **highest score**, end when 0 is encountered

NCBI-BLAST

- Why using BLAST (Basic Local Alignment Search Tool)
  - Universal and flexible
  - Highly sensitive
  - Most frequently used homology search software
  - Well tuned empirical parameters
  - Golden standard in local-alignment
  - Web-accessible
  - Still workable for NGS
- Limitations of BLAST
  - Time consuming, huge computational resources needed for big project
  - Relatively complicated usage or result manipulation

Applications of BLAST

- Search for the homologous genes/sequences in the genomes of same or different species
- Deduce the gene function via approximate ortholog (or paralog) identification
- Screen out potential coding sequences or to define ORF (open reading frame)

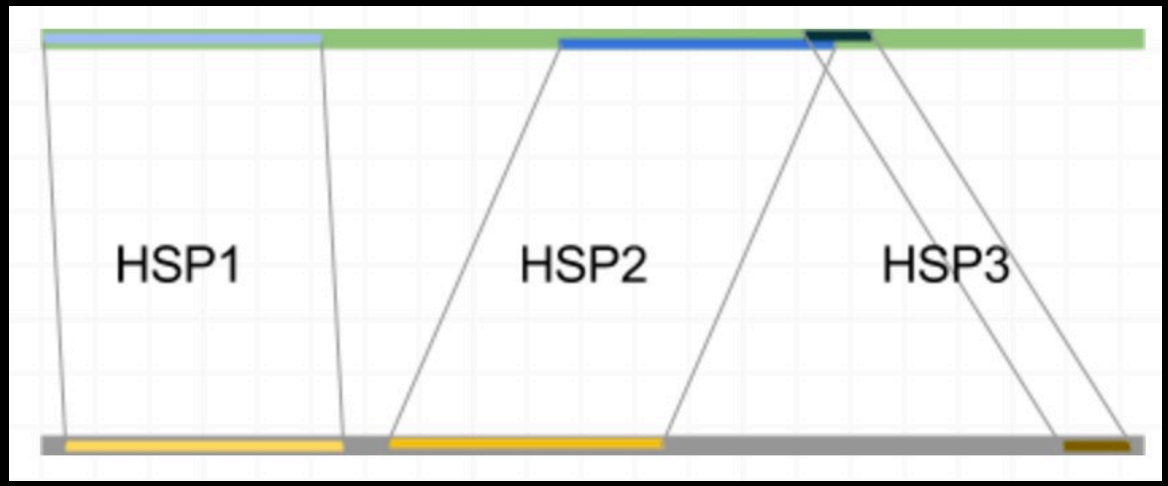
Component of NCBI-blast package

- Blastn (DNA to DNA)
- Blastp (protein to protein)
- Tblastn (protein to translated DNA)
- Blastx (translated DNA to protein)
- Tblastx (translated DNA to translated DNA)
- Psi-blast (blastpgp) (protein to protein)
- …

Central idea of BLAST

- “The central idea of the BLAST algorithm is that a statistically significant alignment is likely to contain a high-scoring pair (HSP) of aligned words. ”

-- Altschul et al. (1997)



http://www.genomequest.com/docs/section/sequence-comparison-algorithms/

Concepts in BLAST, *database index*

- Why?
  - Speed up searching speed by using indexed binary database
- 3 binary files created
  - Index, information about the database
  - Sequence, all the residues
  - Header, information for each sequence

Concepts in BLAST, *substitution matrix*

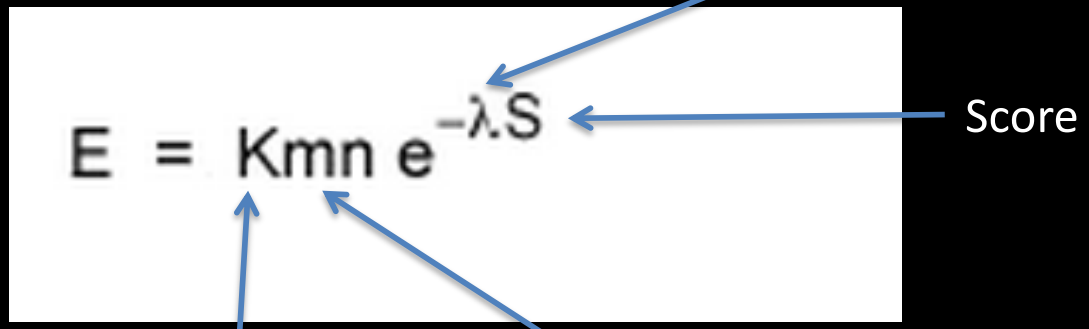
- BLOSUM (BLOcks SUbstitution Matrix) matrix, substitution score matrix
- BLOSUM62

```
# Entries for the BLOSUM62 matrix at a scale of ln(2)/2.0.
   A  R  N  D  C  Q  E  G  H  I  L  K  M  F  P  S  T  W  Y  V  B  J  Z  X  *
A  4 -1 -2 -2  0 -1 -1  0 -2 -1 -1 -1 -1 -2 -1  1  0 -3 -2  0 -2 -1 -1 -1 -4
R -1  5  0 -2 -3  1  0 -2  0 -3 -2  2 -1 -3 -2 -1 -1 -3 -2 -3 -1 -2  0 -1 -4
N -2  0  6  1 -3  0  0  0  1 -3 -3  0 -2 -3 -2  1  0 -4 -2 -3  4 -3  0 -1 -4
D -2 -2  1  6 -3  0  2 -1 -1 -3 -4 -1 -3 -3 -1  0 -1 -4 -3 -3  4 -3  1 -1 -4
C  0 -3 -3 -3  9 -3 -4 -3 -3 -1 -1 -3 -1 -2 -3 -1 -1 -2 -2 -1 -3 -1 -3 -1 -4
Q -1  1  0  0 -3  5  2 -2  0 -3 -2  1  0 -3 -1  0 -1 -2 -1 -2  0 -2  4 -1 -4
E -1  0  0  2 -4  2  5 -2  0 -3 -3  1 -2 -3 -1  0 -1 -3 -2 -2  1 -3  4 -1 -4
G  0 -2  0 -1 -3 -2 -2  6 -2 -4 -4 -2 -3 -3 -2  0 -2 -2 -3 -3 -1 -4 -2 -1 -4
H -2  0  1 -1 -3  0  0 -2  8 -3 -3 -1 -2 -1 -2 -1 -2 -2  2 -3  0 -3  0 -1 -4
I -1 -3 -3 -3 -1 -3 -3 -4 -3  4  2 -3  1  0 -3 -2 -1 -3 -1  3 -3  3 -3 -1 -4
L -1 -2 -3 -4 -1 -2 -3 -4 -3  2  4 -2  2  0 -3 -2 -1 -2 -1  1 -4  3 -3 -1 -4
K -1  2  0 -1 -3  1  1 -2 -1 -3 -2  5 -1 -3 -1  0 -1 -3 -2 -2  0 -3  1 -1 -4
M -1 -1 -2 -3 -1  0 -2 -3 -2  1  2 -1  5  0 -2 -1 -1 -1 -1  1 -3  2 -1 -1 -4
F -2 -3 -3 -3 -2 -3 -3 -3 -1  0  0 -3  0  6 -4 -2 -2  1  3 -1 -3  0 -3 -1 -4
P -1 -2 -2 -1 -3 -1 -1 -2 -2 -3 -3 -1 -2 -4  7 -1 -1 -4 -3 -2 -2 -3 -1 -1 -4
S  1 -1  1  0 -1  0  0  0 -1 -2 -2  0 -1 -2 -1  4  1 -3 -2 -2  0 -2  0 -1 -4
T  0 -1  0 -1 -1 -1 -1 -2 -2 -1 -1 -1 -1 -2 -1  1  5 -2 -2  0 -1 -1 -1 -1 -4
W -3 -3 -4 -4 -2 -2 -3 -2 -2 -3 -2 -3 -1  1 -4 -3 -2 11  2 -3 -4 -2 -2 -1 -4
Y -2 -2 -2 -3 -2 -1 -2 -3  2 -1 -1 -2 -1  3 -3 -2 -2  2  7 -1 -3 -1 -2 -1 -4
V  0 -3 -3 -3 -1 -2 -2 -3 -3  3  1 -2  1 -1 -2 -2  0 -3 -1  4 -3  2 -2 -1 -4
B -2 -1  4  4 -3  0  1 -1  0 -3 -4  0 -3 -3 -2  0 -1 -4 -3 -3  4 -3  0 -1 -4
J -1 -2 -3 -3 -1 -2 -3 -4 -3  3  3 -3  2  0 -3 -2 -1 -2 -1  2 -3  3 -3 -1 -4
Z -1  0  0  1 -3  4  4 -2  0 -3 -3  1 -1 -3 -1  0 -1 -2 -2 -2  0 -3  4 -1 -4
X -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -4
* -4 -4 -4 -4 -4 -4 -4 -4 -4 -4 -4 -4 -4 -4 -4 -4 -4 -4 -4 -4 -4 -4 -4 -4  1
```

Concepts in local-alignment tools, *score and e-value*

high-scoring segment pair (HSP)

```
RPPPPELFELKNTLIVGLITSPDRLIQIRRNRLLSLKENRESDYVDADAVRQEIIAARRL
+      L   LKN L+VGL   SPDRLIQIRRNRL   L      S YV+ DAVR E I A+R
QEDGERLVGLKNPLVVGLTVSPDRLIQIRRNRLDGLNATHTSSYVEPDAVRDETIKAKRA
```

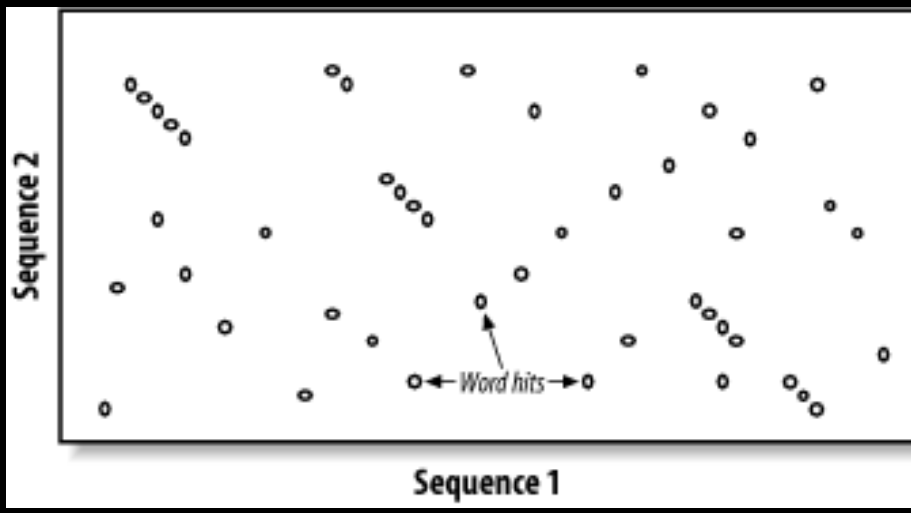



```
# Entries for the BLOSUM62 matrix at a scale of ln(2)/2.0.
   A  R  N  D  C  Q  E  G  H  I  L  K  M  F  P  S  T  W  Y  V  B  J
A  4 -1 -2 -2  0 -1 -1  0 -2 -1 -1 -1 -1 -2 -1  1  0 -3 -2  0 -2 -1
R -1  5  0 -2 -3  1  0 -2  0 -3 -2  2 -1 -3 -2 -1 -1 -3 -2 -3 -1 -2
N -2  0  6  1 -3  0  0  0  1 -3 -3  0 -2 -3 -2  1  0 -4 -2 -3  4 -3
D -2 -2  1  6 -3  0  2 -1 -1 -3 -4 -1 -3 -3 -1  0 -1 -4 -3 -3  4 -3
C  0 -3 -3 -3  9 -3 -4 -3 -3 -1 -1 -3 -1 -2 -3 -1 -1 -2 -2 -1 -3 -1
Q -1  1  0  0 -3  5  2 -2  0 -3 -2  1  0 -3 -1  0 -1 -2 -1 -2  0 -2
E -1  0  0  2 -4  2  5 -2  0 -3 -3  1 -2 -3 -1  0 -1 -3 -2 -2  1 -3
G  0 -2  0 -1 -3 -2 -2  6 -2 -4 -4 -2 -3 -3 -2  0 -2 -2 -3 -3 -1 -4
H -2  0  1 -1 -3  0  0 -2  8 -3 -3 -1 -2 -1 -2 -1 -2 -2  2 -3  0 -3
I -1 -3 -3 -3 -1 -3 -3 -4 -3  4  2 -3  1  0 -3 -2 -1 -3 -1  3 -3  3
L -1 -2 -3 -4 -1 -2 -3 -4 -3  2  4 -2  2  0 -3 -2 -1 -2 -1  1 -4  3
K -1  2  0 -1 -3  1  1 -2 -1 -3 -2  5 -1 -3 -1  0 -1 -3 -2 -2  0 -3
M -1 -1 -2 -3 -1  0 -2 -3 -2  1  2 -1  5  0 -2 -1 -1 -1 -1  1 -3  2
F -2 -3 -3 -3 -2 -3 -3 -3 -1  0  0 -3  0  6 -4 -2 -2  1  3 -1 -3  0
P -1 -2 -2 -1 -3 -1 -1 -2 -2 -3 -3 -1 -2 -4  7 -1 -1 -4 -3 -2 -2 -3
S  1 -1  1  0 -1  0  0  0 -1 -2 -2  0 -1 -2 -1  4  1 -3 -2 -2  0 -2
T  0 -1  0 -1 -1 -1 -1 -2 -2 -1 -1 -1 -1 -2 -1  1  5 -2 -2  0 -1 -1
W -3 -3 -4 -4 -2 -2 -3 -2 -2 -3 -2 -3 -1  1 -4 -3 -2 11  2 -3 -4 -2
Y -2 -2 -2 -3 -2 -1 -2 -3  2 -1 -1 -2 -1  3 -3 -2 -2  2  7 -1 -3 -1
V  0 -3 -3 -3 -1 -2 -2 -3 -3  3  1 -2  1 -1 -2 -2  0 -3 -1  4 -3  2
B -2 -1  4  4 -3  0  1 -1  0 -3 -4  0 -3 -3 -2  0 -1 -4 -3 -3  4 -3
J -1 -2 -3 -3 -1 -2 -3 -4 -3  3  3 -3  2  0 -3 -2 -1 -2 -1  2 -3  3
Z -1  0  0  1 -3  4  4 -2  0 -3 -3  1 -1 -3 -1  0 -1 -2 -2 -2  0 -3
X -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1
* -4 -4 -4 -4 -4 -4 -4 -4 -4 -4 -4 -4 -4 -4 -4 -4 -4 -4 -4 -4 -4 -4
```

Concepts in BLAST, *seed and word size*

- Seed: the small words used for matching
- K-mer dissection in query sequence

http://etutorials.org/



Query: FDRIEA

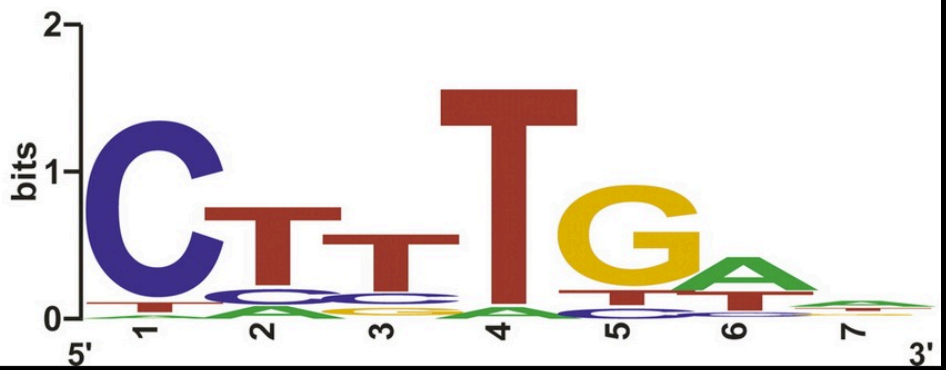
Words (3-mer): FDR, DRI, RIE, IEA

- Dynamic programing to extend the seed match to the neighbor regions using dynamics programing
- Why not direct Smith-waterman search?
  - Gain speed by scarifying a certain amount of sensitivity

PSI-BLAST

- PSI-- Position-Specific Iterative
- PSSM-- position-specific scoring matrix $Score_{ij} = log(\frac{f'_{ij}}{q_i})$
- More sensitive than blastp

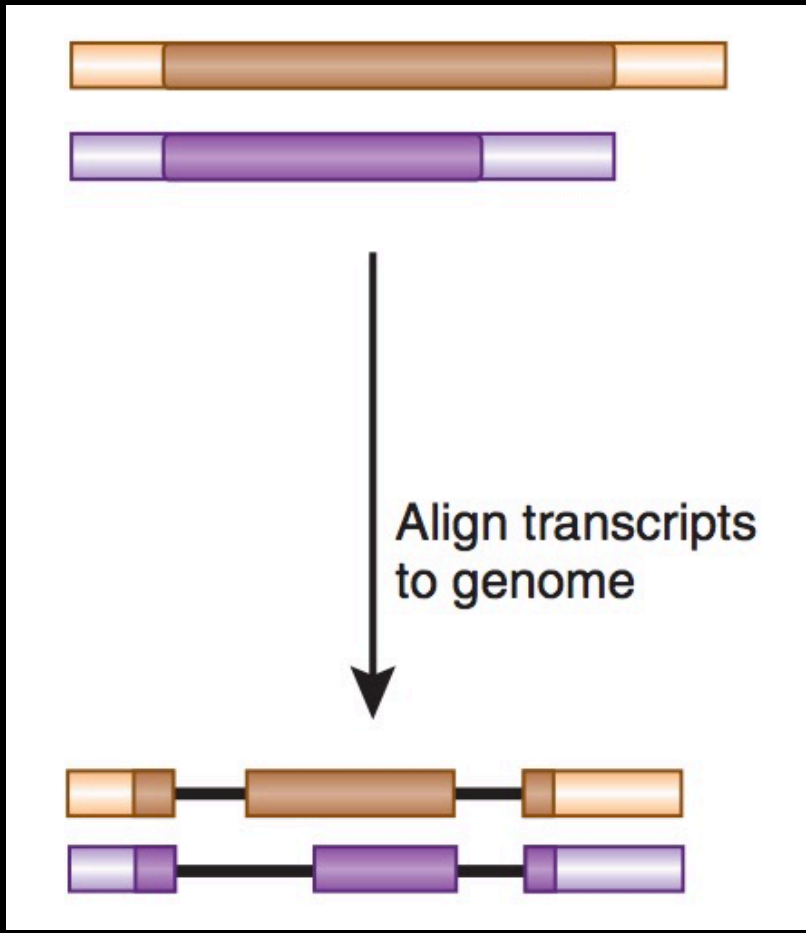
	1	2	3	4	5	6	7
A	1	4	1	2	0	17	13
C	28	5	5	0	3	3	2
G	0	0	4	0	25	1	7
T	2	22	21	29	4	10	9



BLAT, the BLAST-like alignment tool

Features

- Search of DNA fragments in other longer DNA fragments
- The database is usually a genome or an genome assembly
- Less sensitive but faster search than Blast
  - perfect word matches of size 11
- More suitable for cDNA/transcript mapping against genome



http://www.nature.com/nbt/journal/v28/n5/full/nbt0510-421.html

Softwares for NGS mapping
-- a primer for NGS data analysis

- BWA
- Bowtie2
- Soap2
- MOSAIK
- SMALT
- Diamond
- …

Burrows-Wheeler transform

Transformation				
Input	All Rotations	Sorting All Rows in Alphabetical Order by their first letters	Taking Last Column	Output Last Column
^BANANA|	^BANANA| |^BANANA A|^BANAN NA|^BANA ANA|^BAN NANA|^BA ANANA|^B BANANA|^	ANANA|^B ANA|^BAN A|^BANAN BANANA|^ NANA|^BA NA|^BANA ^BANANA| |^BANANA	ANANA|^B ANA|^BAN A|^BANAN BANANA|^ NANA|^BA NA|^BANA ^BANANA| |^BANANA	BNN^AA|A

Input	SIX.MIXED.PIXIES.SIFT.SIXTY.PIXIE.DUST.BOXES
Output	TEXYDST.E.IXIXIXXSSMPPS.B..E.S.UESFXDIIOIIIT

Practice

- Format a database and run blastn and tblastx between DNA sequences
- Re-align the sequences using Smith-Waterman and Needleman–Wunsch algorithm
- Run protein to protein alignment using blastp and blastpgp (PSI-blast)
- Use megablast to search highly homologous sequences
- Use blat to map coding sequence to the genome
- Process management: kill process and background running

Practice

- Use BWA and bowtie2 to map NGS reads to a database
- Examine the raw results with SAM format and manipulate the results using samtools
- Convert SAM to BAM and sort it
- Merge BAM files
- Visualize the alignment

Format of m8 tabular output

```
gi|16124257|ref|NP_418821.1|    gi|16124257|ref|NP_418821.1|    100.00  315    0      0      1      315    1      315    0.0      616
gi|16124257|ref|NP_418821.1|    gi|302381476|ref|YP_003817299.1|        71.06  273    78     1      39     310    11     283    4e-133   377
gi|16124257|ref|NP_418821.1|    gi|403529934|ref|YP_006664463.1|        52.11  261    124    1      48     308    1      260    2e-79    239
gi|16124257|ref|NP_418821.1|    gi|348025276|ref|YP_004765080.1|        32.96  267    171    2      46     308    6      268    5e-46    152
gi|16124257|ref|NP_418821.1|    gi|550918313|ref|YP_008674905.1|        35.11  262    165    4      46     304    4      263    4e-44    147
gi|16124257|ref|NP_418821.1|    gi|550890559|ref|YP_008664651.1|        36.36  275    154    5      46     301    16     288    4e-40    137
gi|16124257|ref|NP_418821.1|    gi|550920762|ref|YP_008677354.1|        31.25  256    173    3      49     302    7      261    1e-38    133
gi|16124257|ref|NP_418821.1|    gi|328955470|ref|YP_004372803.1|        33.85  260    167    4      45     300    11     269    4e-37    129
gi|16124257|ref|NP_418821.1|    gi|374336230|ref|YP_005092917.1|        31.54  279    165    9      46     311    4      269    2e-29    107
gi|16124257|ref|NP_418821.1|    gi|157145973|ref|YP_001453292.1|        31.41  277    167    8      45     308    8      274    2e-27    101
gi|16124257|ref|NP_418821.1|    gi|544576130|ref|YP_008571654.1|        31.77  277    166    8      45     308    8      274    5e-27    100
gi|16124257|ref|NP_418821.1|    gi|407713918|ref|YP_006834483.1|        31.72  268    160    8      46     301    5      261    6e-27    100
gi|16124257|ref|NP_418821.1|    gi|312796261|ref|YP_004029183.1|        31.00  271    158    7      46     301    6      262    7e-27    100
gi|16124258|ref|NP_418822.1|    gi|16124258|ref|NP_418822.1|    100.00  199    0      0      1      199    1      199    2e-122   342
gi|16124258|ref|NP_418822.1|    gi|302381475|ref|YP_003817298.1|        53.71  175    78     2      6      179    10     182    5e-52    162
gi|16124258|ref|NP_418822.1|    gi|403529935|ref|YP_006664464.1|        45.00  180    96     2      1      179    1      178    1e-40    132
gi|16124258|ref|NP_418822.1|    gi|374335786|ref|YP_005092473.1|        32.56  172    104    6      6      174    4      166    2e-14    60.8
gi|16124258|ref|NP_418822.1|    gi|479172553|ref|YP_007800740.1|        28.73  181    113    5      3      179    1      169    6e-14    59.7
gi|16124258|ref|NP_418822.1|    gi|479159506|ref|YP_007788775.1|        26.70  176    117    5      6      179    4      169    6e-14    59.7
gi|16124258|ref|NP_418822.1|    gi|157146214|ref|YP_001453533.1|        26.90  171    115    4      6      174    4      166    4e-12    54.7
gi|16124258|ref|NP_418822.1|    gi|160880072|ref|YP_001559040.1|        24.04  183    124    5      3      179    1      174    2e-11    53.1
gi|16124258|ref|NP_418822.1|    gi|157148823|ref|YP_001456142.1|        28.25  177    120    6      3      179    1      170    1e-10    50.8
gi|16124258|ref|NP_418822.1|    gi|550890932|ref|YP_008665024.1|        31.75  189    110    7      6      179    38     222    1e-10    51.2
gi|16124258|ref|NP_418822.1|    gi|16126581|ref|NP_421145.1|    31.11  180    106    7      6      179    7      174    2e-10    50.4
gi|16124258|ref|NP_418822.1|    gi|479144379|ref|YP_007775242.1|        27.67  159    105    6      22     179    4      153    2e-09    47.4
gi|16124258|ref|NP_418822.1|    gi|326791183|ref|YP_004309004.1|        24.72  178    124    5      3      179    1      169    2e-09    47.0
gi|16124258|ref|NP_418822.1|    gi|348026821|ref|YP_004766626.1|        27.43  175    117    5      6      179    2      167    4e-09    46.6
gi|16124258|ref|NP_418822.1|    gi|544577788|ref|YP_008573312.1|        27.12  177    122    5      3      179    1      170    7e-09    46.2
gi|16124258|ref|NP_418822.1|    gi|479176873|ref|YP_007804474.1|        25.54  184    121    6      3      179    1      175    1e-08    45.1
gi|16124258|ref|NP_418822.1|    gi|479151127|ref|YP_007781303.1|        25.99  177    118    6      6      179    2      168    3e-08    44.3
gi|16124258|ref|NP_418822.1|    gi|302387552|ref|YP_003823374.1|        25.14  179    120    6      6      179    7      176    4e-08    44.3
gi|16124258|ref|NP_418822.1|    gi|302382821|ref|YP_003818644.1|        28.57  182    112    7      1      176    1      170    5e-08    43.5
gi|16124258|ref|NP_418822.1|    gi|479168315|ref|YP_007796814.1|        26.02  196    119    6      1      179    1      187    5e-08    43.9
gi|16124258|ref|NP_418822.1|    gi|317054836|ref|YP_004103303.1|        25.57  176    120    6      6      179    12     178    6e-08    43.5
gi|16124258|ref|NP_418822.1|    gi|407712758|ref|YP_006833323.1|        31.91  188    114    7      2      179    7      190    6e-08    43.5
```

array position	field
0	query name
1	subject name
2	percent identities
3	aligned length
4	number of mismatched positions
5	number of gap positions
6	query sequence start
7	query sequence end
8	subject sequence start
9	subject sequence end
10	e-value
11	bit score

Format of SAM file, standard of NGS alignment

- SAM stands for Sequence Alignment/Map format.

```
M02023:86:000000000-AAC6P:1:1101:12596:1778    83    gb|BK006935.2|  40320   60    151M    =    40142   -329    ACCTAAA
M02023:86:000000000-AAC6P:1:1101:12596:1778    163   gb|BK006935.2|  40142   60    151M    =    40320   329     TTGGGGG
M02023:86:000000000-AAC6P:1:1101:13273:1799    99    gb|BK006935.2|  134064  60    151M    =    134324  411     ACACAAA
M02023:86:000000000-AAC6P:1:1101:13273:1799    147   gb|BK006935.2|  134324  60    151M    =    134064  -411    ATACCAG
M02023:86:000000000-AAC6P:1:1101:15378:1814    83    gb|BK006935.2|  177727  60    151M    =    177318  -560    AAGAGAA
M02023:86:000000000-AAC6P:1:1101:15378:1814    163   gb|BK006935.2|  177318  60    150M    =    177727  560     CACAAGA
M02023:86:000000000-AAC6P:1:1101:12303:1818    99    gb|BK006935.2|  96931   60    150M    =    97108   328     CTTTACA
M02023:86:000000000-AAC6P:1:1101:12303:1818    147   gb|BK006935.2|  97108   60    151M    =    96931   -328    CCTCGTT
M02023:86:000000000-AAC6P:1:1101:14768:1820    83    gb|BK006935.2|  432     60    22M1I27M5D22M2D79M      =       251
M02023:86:000000000-AAC6P:1:1101:14768:1820    163   gb|BK006935.2|  251     60    5S122M1I22M             =       432     338
M02023:86:000000000-AAC6P:1:1101:16682:1858    99    gb|BK006935.2|  163276  60    150M    =    163649  524     CGGATGT
M02023:86:000000000-AAC6P:1:1101:16682:1858    147   gb|BK006935.2|  163649  60    151M    =    163276  -524    GAACCAC
M02023:86:000000000-AAC6P:1:1101:18337:1910    83    gb|BK006935.2|  57556   60    151M    =    57389   -318    ACGTACC
M02023:86:000000000-AAC6P:1:1101:18337:1910    163   gb|BK006935.2|  57389   60    150M    =    57556   318     TATGTTT
M02023:86:000000000-AAC6P:1:1101:17019:1934    99    gb|BK006935.2|  160923  60    151M    =    161101  328     TTGAAGT
```

Col	Field	Description
1	QNAME	Query (pair) NAME
2	FLAG	bitwise FLAG
3	RNAME	Reference sequence NAME
4	POS	1-based leftmost POSition/coordinate of clipped sequence
5	MAPQ	MAPping Quality (Phred-scaled)
6	CIAGR	extended CIGAR string
7	MRNM	Mate Reference sequence NaMe ('=' if same as RNAME)
8	MPOS	1-based Mate POSistion
9	ISIZE	Inferred insert SIZE
10	SEQ	query SEQuence on the same strand as the reference
11	QUAL	query QUALity (ASCII-33 gives the Phred base quality)
12	OPT	variable OPTional fields in the format TAG:VTYPE:VALUE

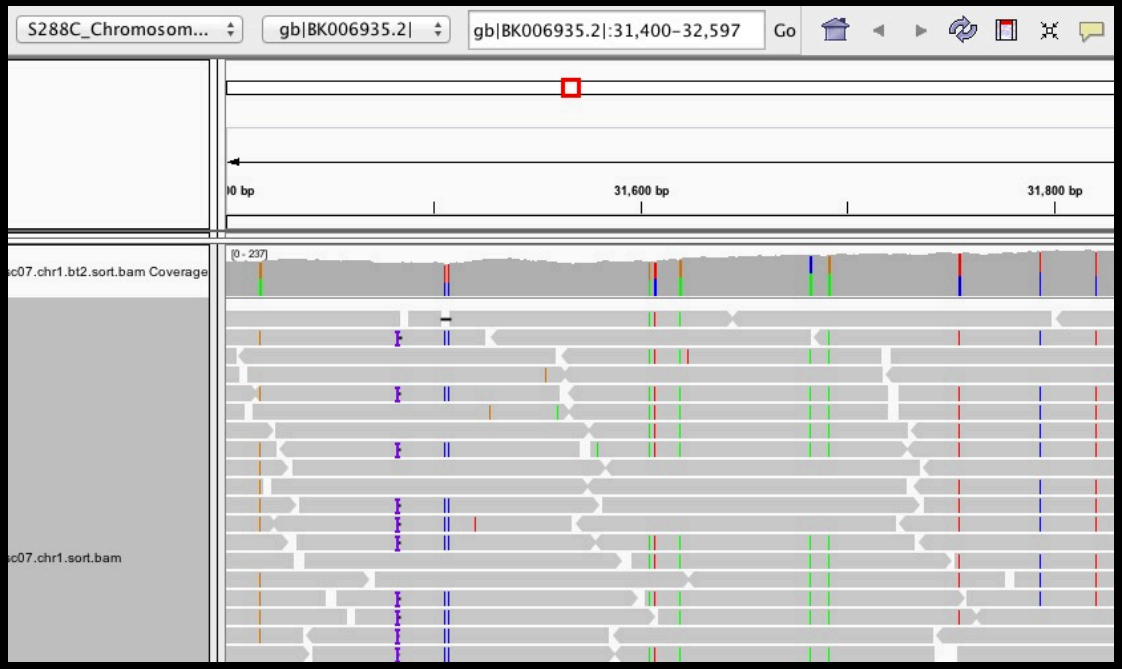
Bit		Description
1	0x1	template having multiple segments in sequencing
2	0x2	each segment properly aligned according to the aligner
4	0x4	segment unmapped
8	0x8	next segment in the template unmapped
16	0x10	SEQ being reverse complemented
32	0x20	SEQ of the next segment in the template being reverse complemented
64	0x40	the first segment in the template
128	0x80	the last segment in the template
256	0x100	secondary alignment
512	0x200	not passing quality controls
1024	0x400	PCR or optical duplicate
2048	0x800	supplementary alignment

Op	BAM	Description
M	0	alignment match (can be a sequence match or mismatch)
I	1	insertion to the reference
D	2	deletion from the reference
N	3	skipped region from the reference
S	4	soft clipping (clipped sequences present in SEQ)
H	5	hard clipping (clipped sequences NOT present in SEQ)
P	6	padding (silent deletion from padded reference)
=	7	sequence match
X	8	sequence mismatch

https://samtools.github.io/hts-specs/SAMv1.pdf

Visualize the alignment

- Open IGV (Start-Documents) and go to “Genome”-> “load genome from file” , then select “S288C_Chromosome_I.fsa” in your pc
- Load bam alignment “sc07.chr1.sort.bam” from “File”-> “Load from File”



“sc07.chr1.sort.bam.bai” should also be transferred back to your PC and put under the same fold with another two files

Go to IGV official website for more info

http://www.broadinstitute.org/igv/book/export/html/6