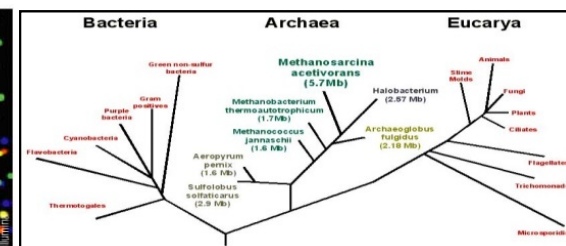
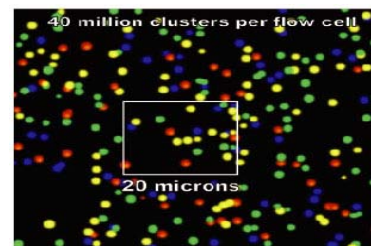




TAACCCTAACCCCTAACCCCTAACCCCTAACCCCTA  
CCTAACCCCTAACCCCTAACCCCTAACCCCTAACCC  
CCCTAACCCCTAACCCCTAACCCCTAACCCCTAAC  
AACCCCTAACCCCTAACCCCTAACCCCTAACCCCTA  
ACCCTAACCCCAACCCCAACCCCAACCCCAAC  
CTACCCTAACCCCTAACCCCTAACCCCTAACCCCTA  
ACCCTAACCCCTAACCCCTAACCCCTAACCCCTAA

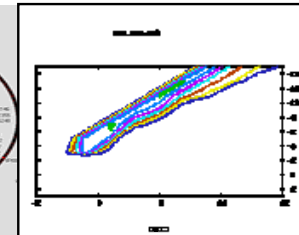
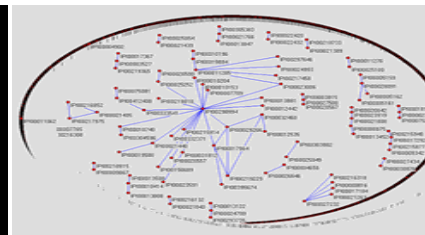
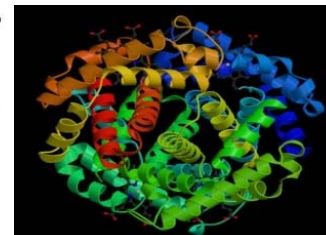
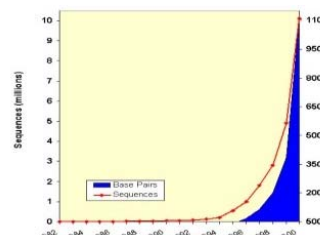
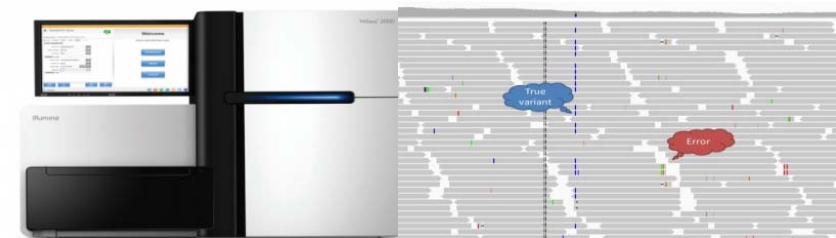


# Supplementary Learning Materials

孟宇琦(Yuqi Meng) 叶永鑫(Adam Y. Ye) 黄岳(Yue Huang)

北京大学生物信息学中心

Center for Bioinformatics, Peking University



# Outline

- BWA & BWT algorithm
- Variant caller
  - samtools
  - GATK

# BWA / BWT algorithm

- The compression algorithm used in BWA
- Lossless compression
- Sort and transform the char matrix with string rotation
- Reverse-char method was utilized for match
- Cannot handle gap

ACTACGG

A	C	T	A	C	G	G
C	T	A	C	G	G	A
T	A	C	G	G	A	C
A	C	G	G	A	C	T
C	G	G	A	C	T	A
G	G	A	C	T	A	C
G	A	C	T	A	C	G

sort  
→

A	C	G	G	A	C	T
A	C	T	A	C	G	G
C	G	G	A	C	T	A
C	T	A	C	G	G	A
G	A	C	T	A	C	G
G	G	A	C	T	A	C
T	A	C	G	G	A	C

TGAAGCC

$I=2$

Copyright © Peking University

TGAAGCC

$l=2$

L	F
T	A
G	A
A	C
A	C
G	G
C	G
C	T

sort  
→

A	C	G	G	A	C	T
A	C	T	A	C	G	G
C	G	G	A	C	T	A
C	T	A	C	G	G	A
G	A	C	T	A	C	G
G	G	A	C	T	A	C
T	A	C	G	G	A	C

F	L	
A	T	5
A	G	1
C	A	4
C	A	7
G	G	2
G	C	3
T	C	6

ACTACGG

————→ **TACG**

F L

A	T
A	G
C	A
C	A
G	G
G	C
T	C

F L

A	T	
A	G	1
C	A	
C	A	
G	G	2
G	C	
T	C	

F L

A	T	4
A	G	
C	A	3
C	A	
G	G	1
G	C	2
T	C	

**ACTACGG**

||||

**TACG**

**ACTACGG**

||

**AC**

||

**GG**

————→ **GGAC?**

ACTACGG → ACTACGG\$

F L

→ GGAC

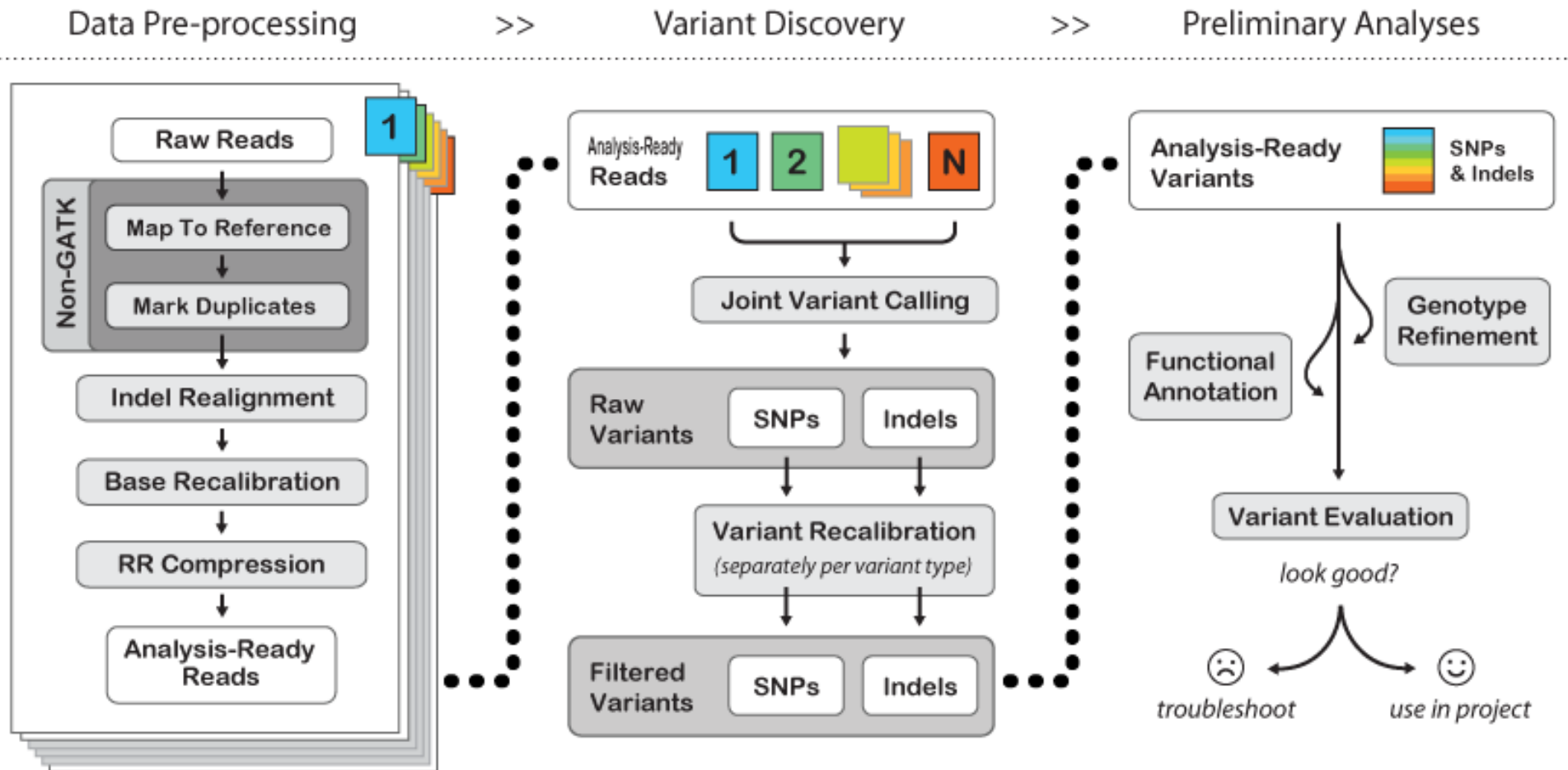
\$	G	
A	T	
A	\$	3
C	A	
C	A	2
G	G	
G	C	
T	C	1

# Variant caller

- samtools
  - mpileup + bcftools
- GATK
  - UnifiedGenotyper
  - HaplotypeCaller



# GATK



<http://www.broadinstitute.org/gatk/guide/best-practices>

# Thank you for your attention



<https://www.coursera.org/course/pkubioinfo>