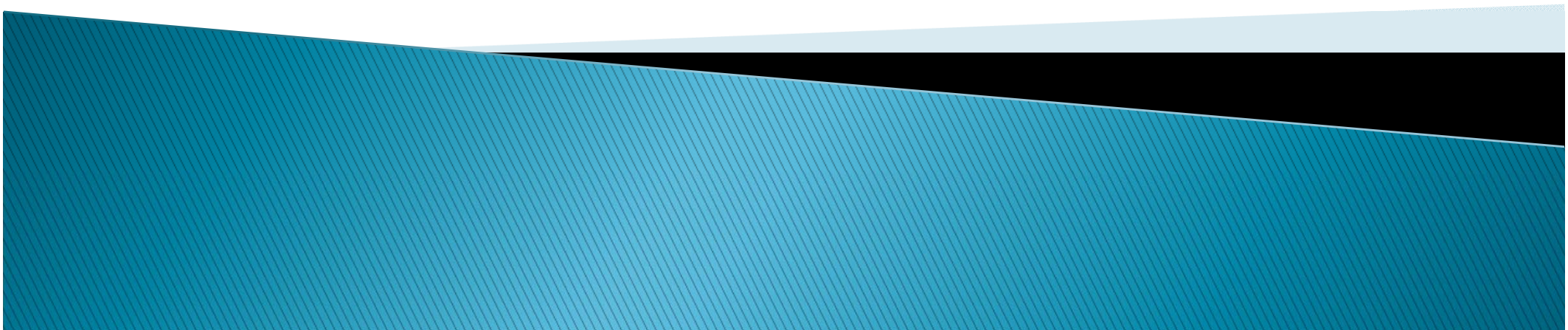
RNA seq

BoQiang Hu 胡博强
School of Life Sciences, Peking University



Before our talk:

- What can RNA-seq do?
  - Level 1:
    - Gene Expression (RPKM)
  - Level 2:
    - Transcriptome reconstruction & Alternative Splicing
    - Isoform abundance & Differential Expression
  - Also:
    - RNA-editing(DNA-SNP vs RNA-SNP)
    - eQTL(SNP-Expression)
    - Fusion
  - Using result:
    - GO & Pathway Enrichment

Topic

- 1. Read mapping
- 2. Transcriptome reconstruction
- 3. Differential Expression

1.

- 1.
  - 2
  - 3)

>NM_017900 Homo sapiens aurora kinase A interacting protein 1 (AURKAIP1), transcript variant 1, mRNA

CCCGCACCCCCTGGGATTGTGGGAAATGTAGTTTTTTGCCTCCGTAAGGGACCAGGCGGAGCTGAGGAAC
CGCGCGAGGACTGGGACCGTGATTCCACTAACCGGAAACCGTCGCCTTTCGGGCCCGGCGGGGCCTGAGC
CAATGCAGAATCGGGGGCCGCGAGGACGCCAGCGGGCGCTGTGCGTAGGAACCGCCGGGTGGCCGCTGCC
GATCGGGGCCGACTTGGGGACGGACCGGAAGTGCCCGAGGGCGGCCGCAGAACGGTCAATTTGAGCCGCG
TCGAGCTCCCCTGGGACCTGTGGCCGCCGCCCACAGACCATGCTCCTGGGGCGCCTGACTTCCCAGCTGT
TGAGGGCCGTTCCTTGGGCAGGCGGCCGCCCGCCTTGGCCCGTCTCTGGAGTGCTGGGCAGCCGGGTCTG
CGGGCCCCTTTACAGCACATCGCCGGCCGGCCCAGGTAGGGCGGCCTCTCTCCCTCGCAAGGGGGCCCAG
CTGGAGCTGGAGGAGATGCTGGTCCCCAGGAAGATGTCCGTCAGCCCCCTGGAGAGCTGGCTCACGGCCC
GCTGCTTCCTGCCCAGACTGGATACCGGGACCGCAGGGACTGTGGCTCCACCGCAATCCTACCAGTGTCC
GCCCAGCCAGATAGGGGAAGGGGCCGAGCAGGGGGATGAAGGCGTCGCGGATGCGCCTCAAATTCAGTGC
AAAAACGTGCTGAAGATCCGCCGGCGGAAGATGAACCACCACAAGTACCGGAAGCTGGTGAAGAAGACGC
GGTTCCTGCGGAGGAAGGTCCAGGAGGGACGCCTGAGACGCAAGCAGATCAAGTTCGAGAAAGACCTGAG
GCGCATCTGGCTGAAGGCGGGGCTAAAGGAAGCCCCCGAAGGCTGGCAGACCCCCAAGATCTACCTGCGG
GGCAAATGAGTCTGGCGCCGCCCTTCCCGCCCGTTGCTGCTGTGATCCGTAGTAATAAATTCTCAGAGGA
CTCAGCCTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

>NM_001127229 Homo sapiens aurora kinase A interacting protein 1 (AURKAIP1), transcript variant 2, mRNA

AAGTGCCCGAGGGCGGCCGCAGAACGGTCAATTTGAGCCGCGTCGAGGTCGGGCTTGGGAAGGGTCAGCG
GGAGGCCTGAGGGCGCCGGGCGCTGCGGCAGGCGGGCCCGGGGTCCAGCCGAGAGGGTCCGGGCGCCAGG
CAACGCGATTCGCGCGGGGGTGAACCCGGGGAGGGGGCCGGCCTCCCCGTTCTGGGACCTTTCGCTCCCC
TGGGACCTGTGGCCGCCGCCCACAGACCATGCTCCTGGGGCGCCTGACTTCCCAGCTGTTGAGGGCCGTT
CCTTGGGCAGGCGGCCGCCCGCCTTGGCCCGTCTCTGGAGTGCTGGGCAGCCGGGTCTGCGGGCCCCTTT
ACAGCACATCGCCGGCCGGCCCAGGTAGGGCGGCCTCTCTCCCTCGCAAGGGGGCCCAGCTGGAGCTGGA
GGAGATGCTGGTCCCCAGGAAGATGTCCGTCAGCCCCCTGGAGAGCTGGCTCACGGCCCGCTGCTTCCTG
CCCAGACTGGATACCGGGACCGCAGGGACTGTGGCTCCACCGCAATCCTACCAGTGTCCGCCCAGCCAGA
TAGGGGAAGGGGCCGAGCAGGGGGATGAAGGCGTCGCGGATGCGCCTCAAATTCAGTGCAAAAACGTGCT
GAAGATCCGCCGGCGGAAGATGAACCACCACAAGTACCGGAAGCTGGTGAAGAAGACGCGGTTCCTGCGG
AGGAAGGTCCAGGAGGGACGCCTGAGACGCAAGCAGATCAAGTTCGAGAAAGACCTGAGGCGCATCTGGC
TGAAGGCGGGGCTAAAGGAAGCCCCCGAAGGCTGGCAGACCCCCAAGATCTACCTGCGGGGCAAATGAGT
CTGGCGCCGCCCTTCCCGCCCGTTGCTGCTGTGATCCGTAGTAATAAATTCTCAGAGGACTCAGCCTTTA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

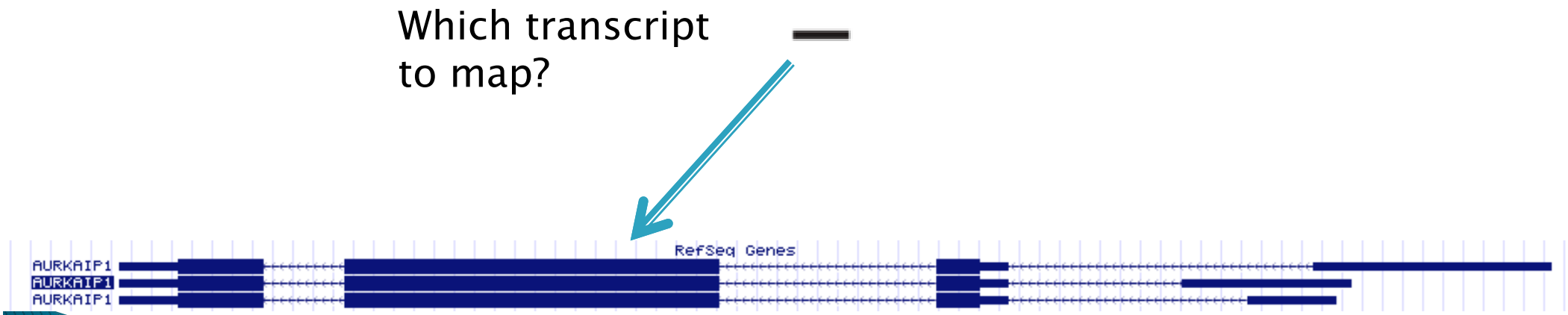
>NM_001127230 Homo sapiens aurora kinase A interacting protein 1 (AURKAIP1), transcript variant 3, mRNA

GCAGAACGGTCAATTTGAGCCGCGTCGAGGTCGGGCTTGGGAAGGGTCAGCGGGAGGCCTGAGGGCGCCG
GGCGCTGCGGCAGGCGGGCCCGGGGTCCAGCCGAGAGGCTCCCCTGGGACCTGTGGCCGCCGCCCACAGA
CCATGCTCCTGGGGCGCCTGACTTCCCAGCTGTTGAGGGCCGTTCCTTGGGCAGGCGGCCGCCCGCCTTG
GCCCGTCTCTGGAGTGCTGGGCAGCCGGGTCTGCGGGCCCCTTTACAGCACATCGCCGGCCGGCCCAGGT
AGGGCGGCCTCTCTCCCTCGCAAGGGGGCCCAGCTGGAGCTGGAGGAGATGCTGGTCCCCAGGAAGATGT
CCGTCAGCCCCCTGGAGAGCTGGCTCACGGCCCGCTGCTTCCTGCCCAGACTGGATACCGGGACCGCAGG
GACTGTGGCTCCACCGCAATCCTACCAGTGTCCGCCCAGCCAGATAGGGGAAGGGGCCGAGCAGGGGGAT

GAAGGCTGGCAGACCCCCAAGATCTACCTGCGGGGCAAATGAGTCTGGCGCCGCCCTTCCCGCCCGTTGC
TGCTGTGATCCGTAGTAATAAATTCTCAGAGGACTCAGCCTTAAAAAAAAAAAAAA

1.Reads mapping

- Disadvantage:
  - For Level1 (Gene-Expression) is OK.
  - However, we cannot get transcripts expression because of potential multi-match of reads in transcripts of a single gene.



Which transcript to map?

Read:
CGGGGCTAAAGGAAGCCCCCGAAGGCTGGCAGACCCCCAAGATCTACCTGCGG
GGCAAATGAGTCTGGCGCCGCCCTTCCCGCCCGTTGCTGCTGTGATCCGTAGT
AATAAATTCTCAGAGGACTCAGCCT

```
>NM_017900     Homo sapiens aurora kinase A interacting protein 1 (AURKAIP1), transcript variant 1, mRNA
CCCGCACCCCCTGGGATTGTGGGAAATGTAGTTTTTTGCCTCCGTAAGGGACCAGGCGGAGCTGAGGAACCGCGCGAGGACTGGGACCGTGATTCCACTAACCGGA
AACCGTCGCCTTTCGGGCCCGGCGGGGCCTGAGCCAATGCAGAATCGGGGGCCGCGAGGACGCCAGCGGGCGCTGTGCGTAGGAACCGCCGGGTGGCCGCTGCCGA
TCGGGGCCGACTTGGGGACGGACCGGAAGTGCCCGAGGGCGGCCGCAGAACGGTCAATTTGAGCCGCGTCGAGCTCCCCTGGGACCTGTGGCCGCCGCCCACAGAC
CATGCTCCTGGGGCGCCTGACTTCCCAGCTGTTGAGGGCCGTTCCTTGGGCAGGCGGCCGCCCGCCTTGGCCCGTCTCTGGAGTGCTGGGCAGCCGGGTCTGCGGG
CCCCTTTACAGCACATCGCCGGCCGGCCCAGGTAGGGCGGCCTCTCTCCCTCGCAAGGGGGCCCAGCTGGAGCTGGAGGAGATGCTGGTCCCCAGGAAGATGTCCG
TCAGCCCCCTGGAGAGCTGGCTCACGGCCCGCTGCTTCCTGCCCAGACTGGATACCGGGACCGCAGGGACTGTGGCTCCACCGCAATCCTACCAGTGTCCGCCCAG
CCAGATAGGGGAAGGGGCCGAGCAGGGGGATGAAGGCGTCGCGGATGCGCCTCAAATTCAGTGCAAAAACGTGCTGAAGATCCGCCGGCGGAAGATGAACCACCAC
AAGTACCGGAAGCTGGTGAAGAAGACGCGGTTCCTGCGGAGGAAGGTCCAGGAGGGACGCCTGAGACGCAAGCAGATCAAGTTCGAGAAAGACCTGAGGCGCATCT
GGCTGAAGGCGGGGCTAAAGGAAGCCCCCGAAGGCTGGCAGACCCCCAAGATCTACCTGCGGGGCAAATGAGTCTGGCGCCGCCCTTCCCGCCCGTTGCTGCTGTG
ATCCGTAGTAATAAATTCTCAGAGGACTCAGCCTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
>NM_001127229   Homo sapiens aurora kinase A interacting protein 1 (AURKAIP1), transcript variant 2, mRNA
AAGTGCCCGAGGGCGGCCGCAGAACGGTCAATTTGAGCCGCGTCGAGGTCGGGCTTGGGAAGGGTCAGCGGGAGGCCTGAGGGCGCCGGGCGCTGCGGCAGGCGGG
CCCGGGGTCCAGCCGAGAGGGTCCGGGCGCCAGGCAACGCGATTCGCGCGGGGGTGAACCCGGGGAGGGGGCCGGCCTCCCCGTTCTGGGACCTTTCGCTCCCCTG
GGACCTGTGGCCGCCGCCCACAGACCATGCTCCTGGGGCGCCTGACTTCCCAGCTGTTGAGGGCCGTTCCTTGGGCAGGCGGCCGCCCGCCTTGGCCCGTCTCTGG
AGTGCTGGGCAGCCGGGTCTGCGGGCCCCTTTACAGCACATCGCCGGCCGGCCCAGGTAGGGCGGCCTCTCTCCCTCGCAAGGGGGCCCAGCTGGAGCTGGAGGAG
ATGCTGGTCCCCAGGAAGATGTCCGTCAGCCCCCTGGAGAGCTGGCTCACGGCCCGCTGCTTCCTGCCCAGACTGGATACCGGGACCGCAGGGACTGTGGCTCCAC
CGCAATCCTACCAGTGTCCGCCCAGCCAGATAGGGGAAGGGGCCGAGCAGGGGGATGAAGGCGTCGCGGATGCGCCTCAAATTCAGTGCAAAAACGTGCTGAAGAT
CCGCCGGCGGAAGATGAACCACCACAAGTACCGGAAGCTGGTGAAGAAGACGCGGTTCCTGCGGAGGAAGGTCCAGGAGGGACGCCTGAGACGCAAGCAGATCAAG
TTCGAGAAAGACCTGAGGCGCATCTGGCTGAAGGCGGGGCTAAAGGAAGCCCCCGAAGGCTGGCAGACCCCCAAGATCTACCTGCGGGGCAAATGAGTCTGGCGCC
GCCCTTCCCGCCCGTTGCTGCTGTGATCCGTAGTAATAAATTCTCAGAGGACTCAGCCTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
>NM_001127230   Homo sapiens aurora kinase A interacting protein 1 (AURKAIP1), transcript variant 3, mRNA
GCAGAACGGTCAATTTGAGCCGCGTCGAGGTCGGGCTTGGGAAGGGTCAGCGGGAGGCCTGAGGGCGCCGGGCGCTGCGGCAGGCGGGCCCGGGGTCCAGCCGAGA
GGCTCCCCTGGGACCTGTGGCCGCCGCCCACAGACCATGCTCCTGGGGCGCCTGACTTCCCAGCTGTTGAGGGCCGTTCCTTGGGCAGGCGGCCGCCCGCCTTGGC
CCGTCTCTGGAGTGCTGGGCAGCCGGGTCTGCGGGCCCCTTTACAGCACATCGCCGGCCGGCCCAGGTAGGGCGGCCTCTCTCCCTCGCAAGGGGGCCCAGCTGGA
GCTGGAGGAGATGCTGGTCCCCAGGAAGATGTCCGTCAGCCCCCTGGAGAGCTGGCTCACGGCCCGCTGCTTCCTGCCCAGACTGGATACCGGGACCGCAGGGACT
GTGGCTCCACCGCAATCCTACCAGTGTCCGCCCAGCCAGATAGGGGAAGGGGCCGAGCAGGGGGATGAAGGCGTCGCGGATGCGCCTCAAATTCAGTGCAAAAACG
TGCTGAAGATCCGCCGGCGGAAGATGAACCACCACAAGTACCGGAAGCTGGTGAAGAAGACGCGGTTCCTGCGGAGGAAGGTCCAGGAGGGACGCCTGAGACGCAA
GCAGATCAAGTTCGAGAAAGACCTGAGGCGCATCTGGCTGAAGGCGGGGCTAAAGGAAGCCCCCGAAGGCTGGCAGACCCCCAAGATCTACCTGCGGGGCAAATGA
GTCTGGCGCCGCCCTTCCCGCCCGTTGCTGCTGTGATCCGTAGTAATAAATTCTCAGAGGACTCAGCCTTAAAAAAAAAAAAAAA
```

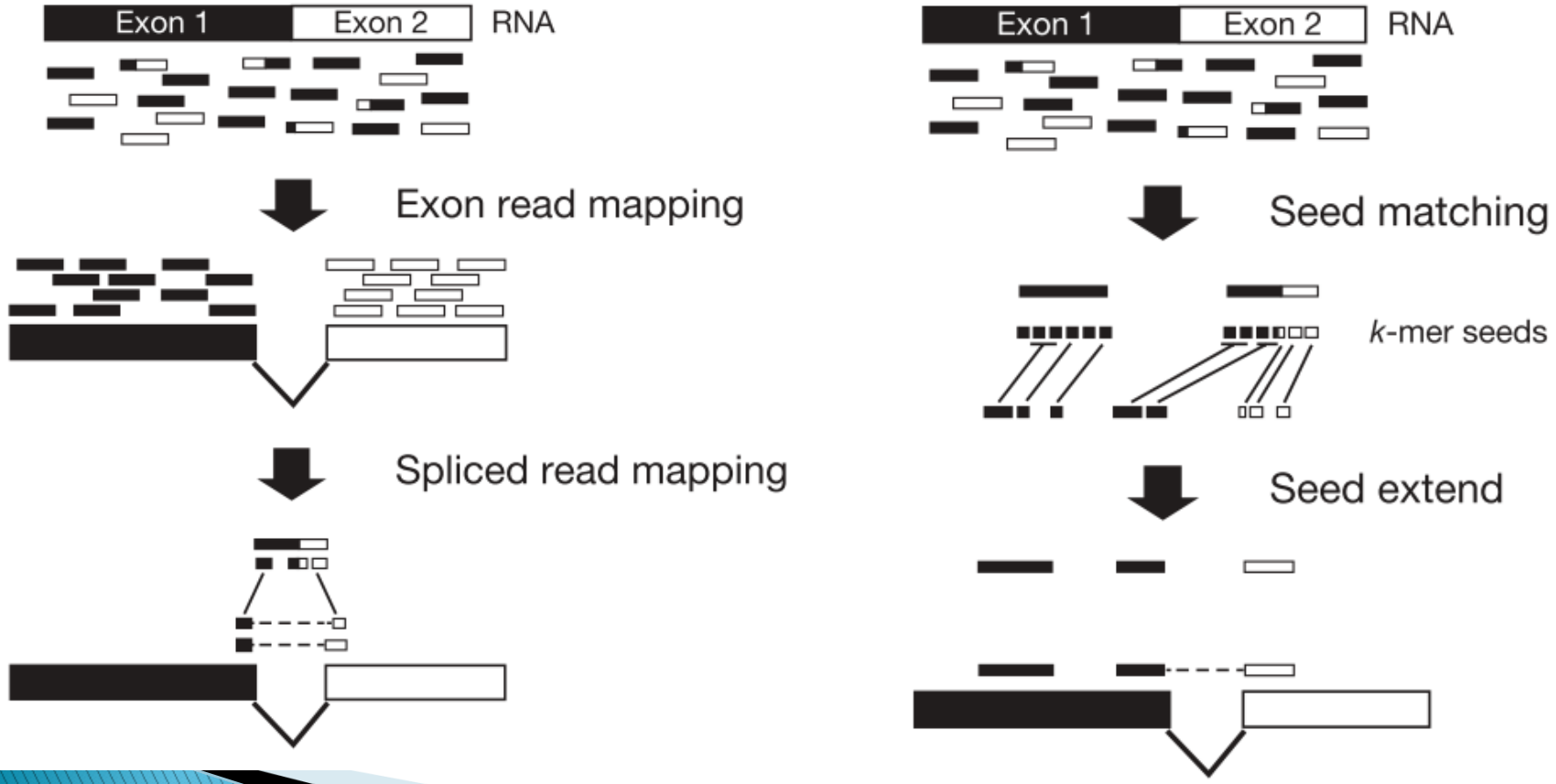
1.Reads mapping

- 1. Spliced aligners:
  - Align reads to genome, noval splice junctions were OK.
  - Using junctions and expression of each exons to estimate expression of each transcript.
  - So reference is genome.
  - Exon-first vs Seed-extend

1.Reads mapping

- 1. Spliced aligners

Exon-first vs Seed-extend

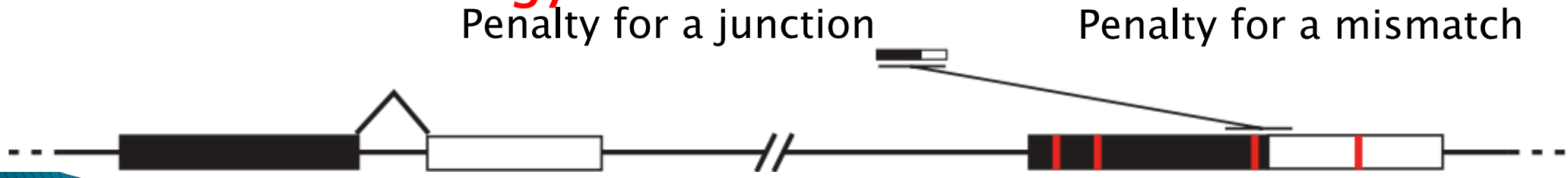


Garber, Manuel, et al. *Nature methods,* 2011

1.Reads mapping

- Exon-first's disadvantage:
  - When there is a retrotransposed pseudogene, a spliced read could be incorrectly assigned to the pseudogene as it appears to be exonic.

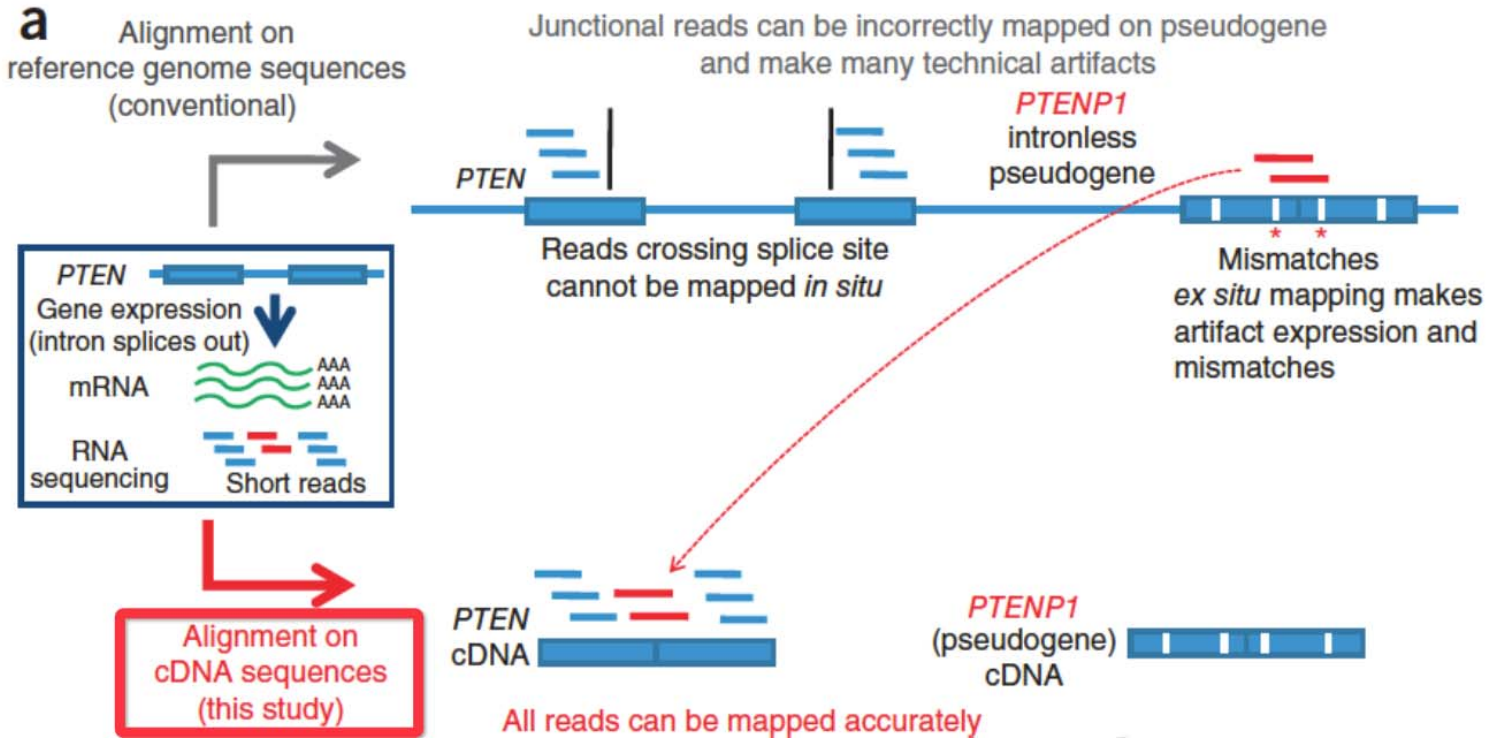
However, this method is still recommended because of its balance of resource usage and accuracy.
So the solution is to make a adjustment of this strategy



Garber, Manuel, et al. *Nature methods,* 2011

1.Reads mapping

- Exon-first's disadvantage:
  - An alternative strategy : map to cDNA first.



Young Seok Ju *et al, Nature Genetics*, 2011

1.Reads mapping

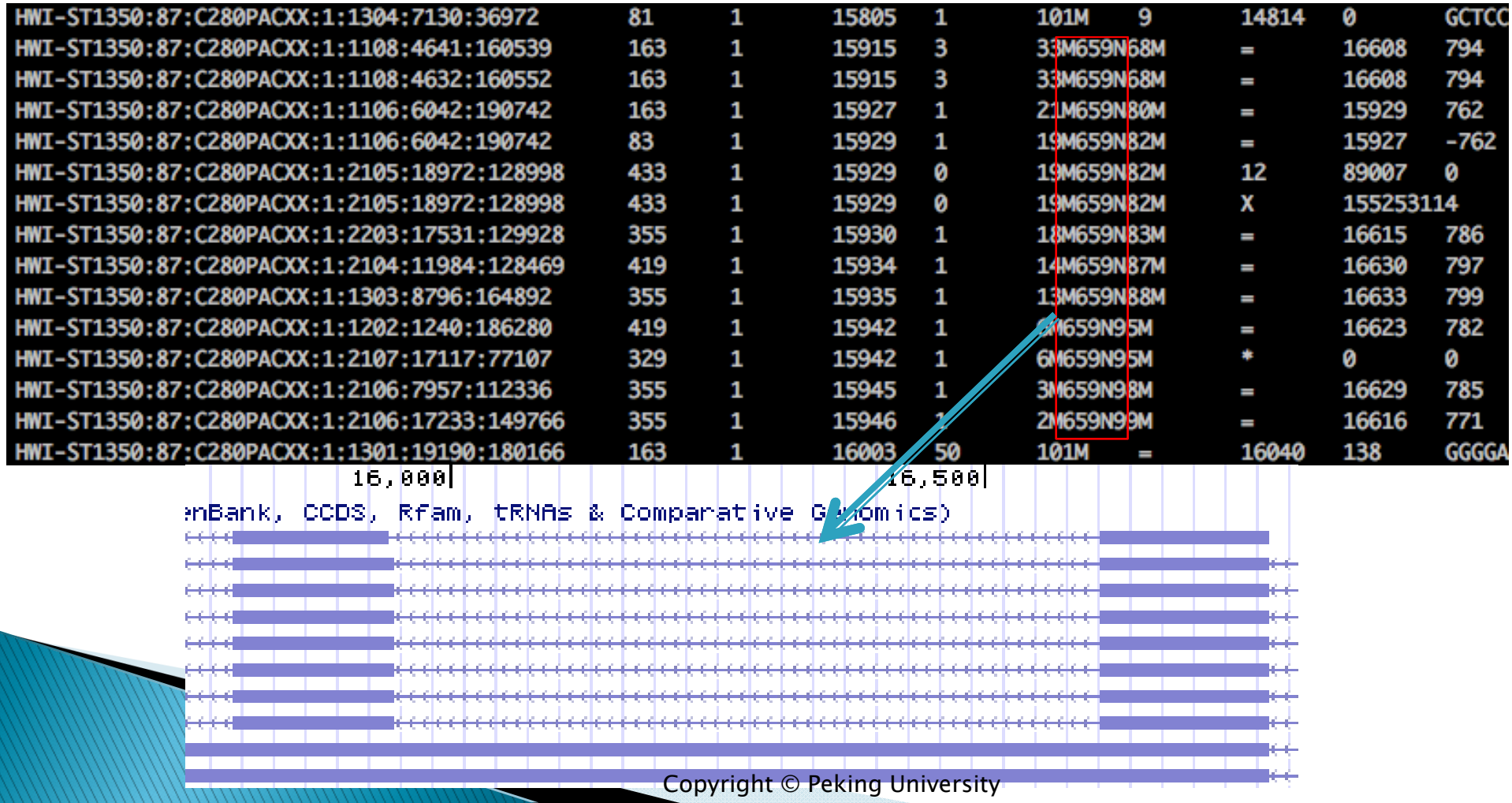
- TopHat recommand

Table 1 | Selected list of RNA-seq analysis programs

Class	Category	Package	Notes	Uses	Input
Read mapping					
Unspliced aligners[a]	Seed methods	Short-read mapping package (SHRiMP)[41]	Smith-Waterman extension	Aligning reads to a reference transcriptome	Reads and reference transcriptome
		Stampy[39]	Probabilistic model		
	Burrows-Wheeler transform methods	Bowtie[43]			
		BWA[44]	Incorporates quality scores		
Spliced aligners	Exon-first methods	MapSplice[52]	Works with multiple unspliced aligners	Aligning reads to a reference genome. Allows for the identification of novel splice junctions	Reads and reference genome
		SpliceMap[50]			
		TopHat[51]	Uses Bowtie alignments		
	Seed-extend methods	GSNAP[53]	Can use SNP databases		
		QPALMA[54]	Smith-Waterman for large gaps		

Garber, Manuel, et al. *Nature methods* 2011

Spliced-reads in tophat

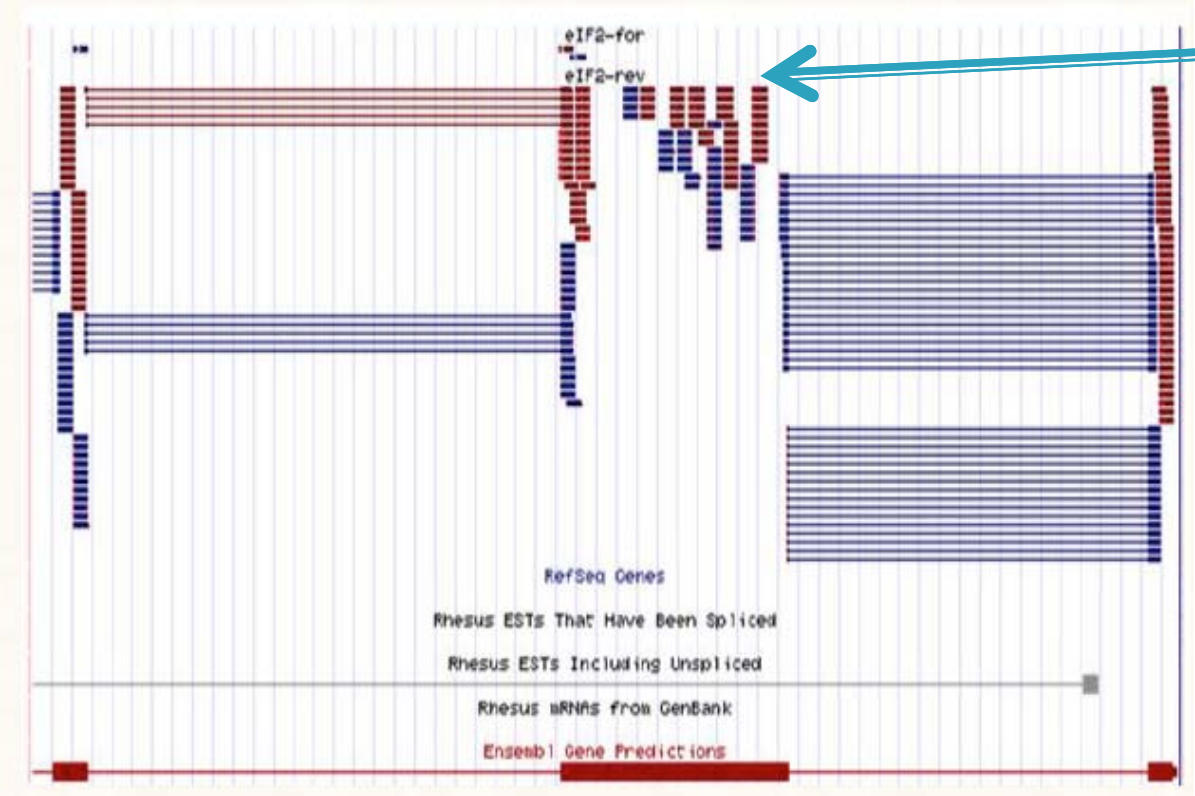


Topic

- 1. Read mapping
- 2. Transcriptome reconstruction
- 3. Differential Expression

2. Transcriptome reconstruction

- 2 approaches:
  - 1. Genome-guided
    - Map genome reads to Exon region.



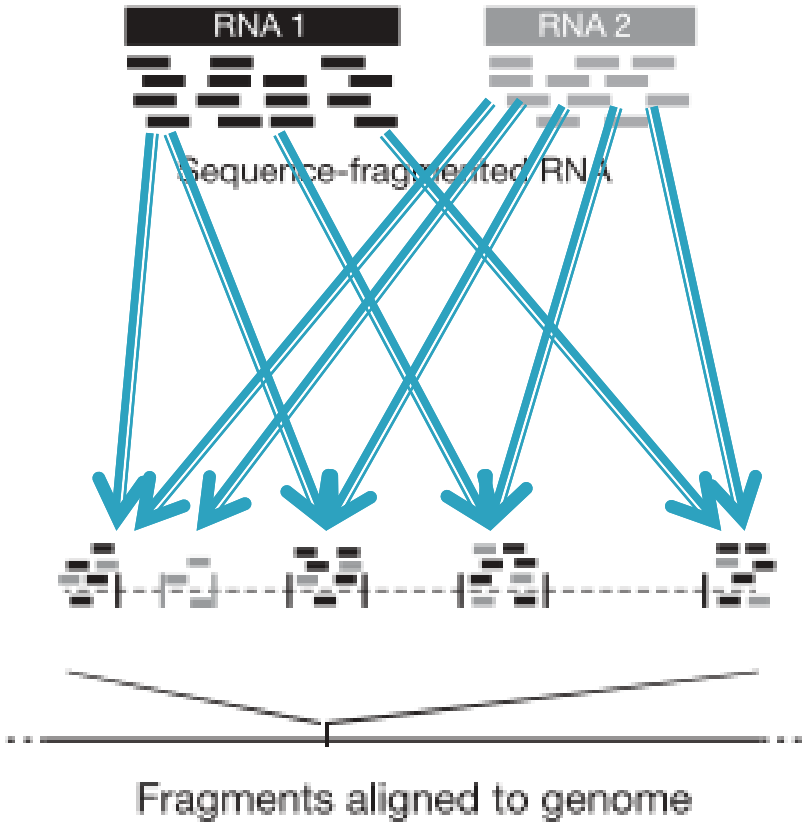
Genome-reads

Introns Exons Introns

Gene-region

2. Transcriptome reconstruction

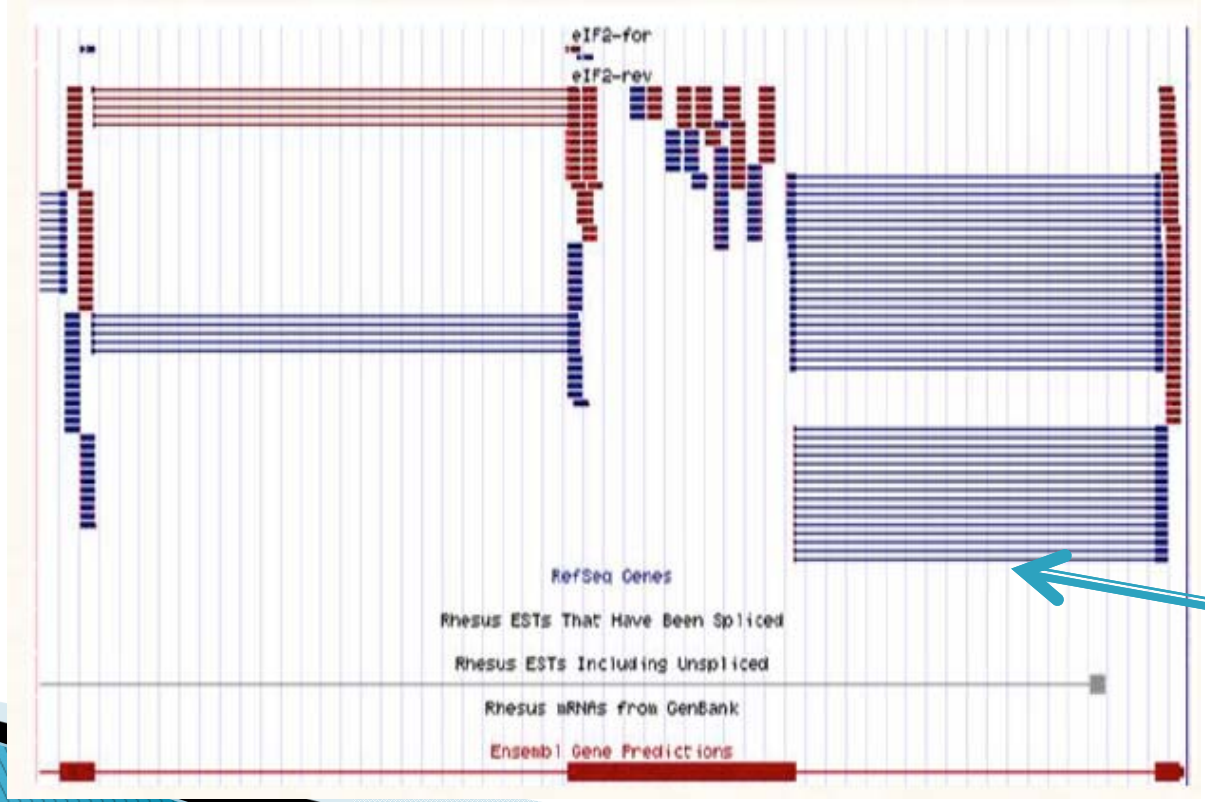
- 1



Trapnell, Cole, et al. *Nature biotechnology*, 2010

2. Transcriptome reconstruction

- 2 approaches:
  - 1. Genome-guided
    - Map Splicing-reads as junctions.



Genome-reads

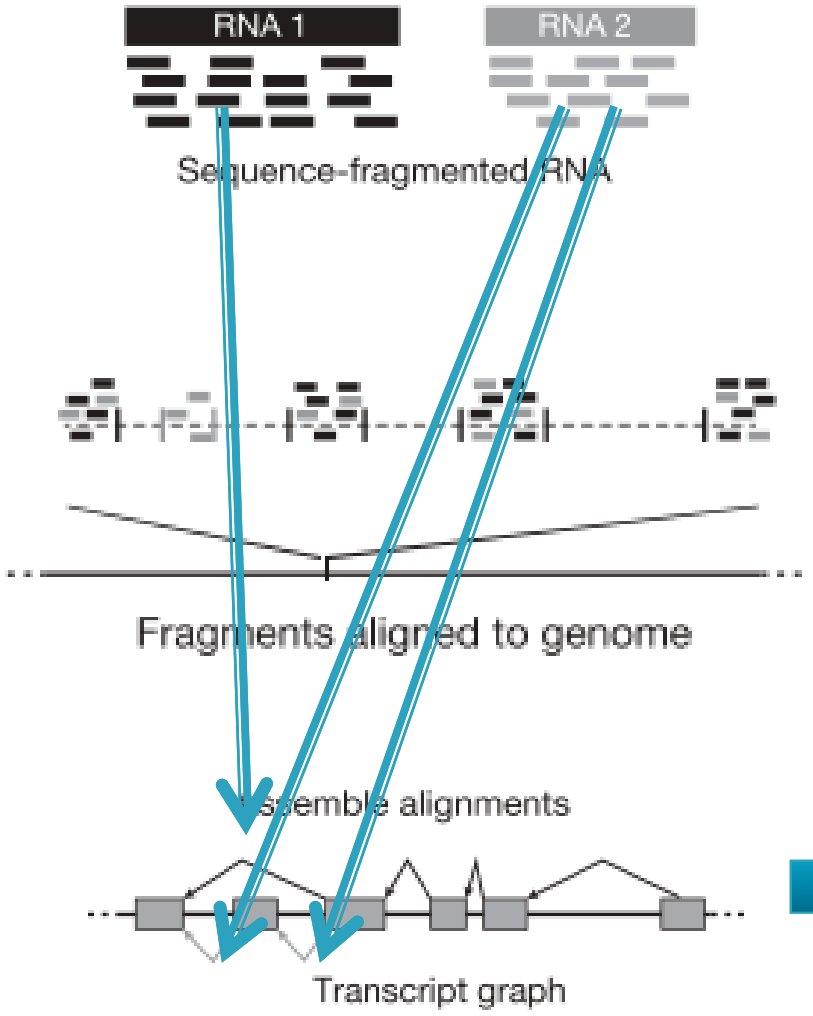
Splicing-reads

Gene-region

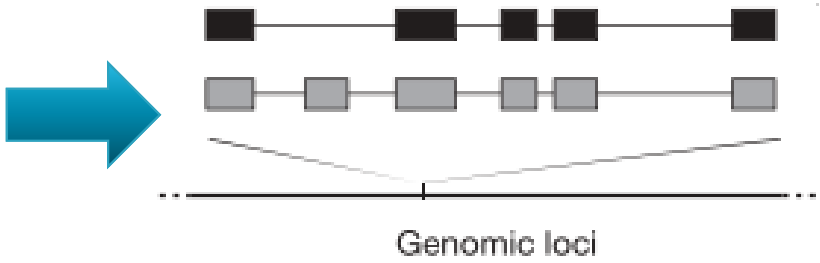
Exons

Introns

2. Transcriptome reconstruction



- Splice-reads mapping



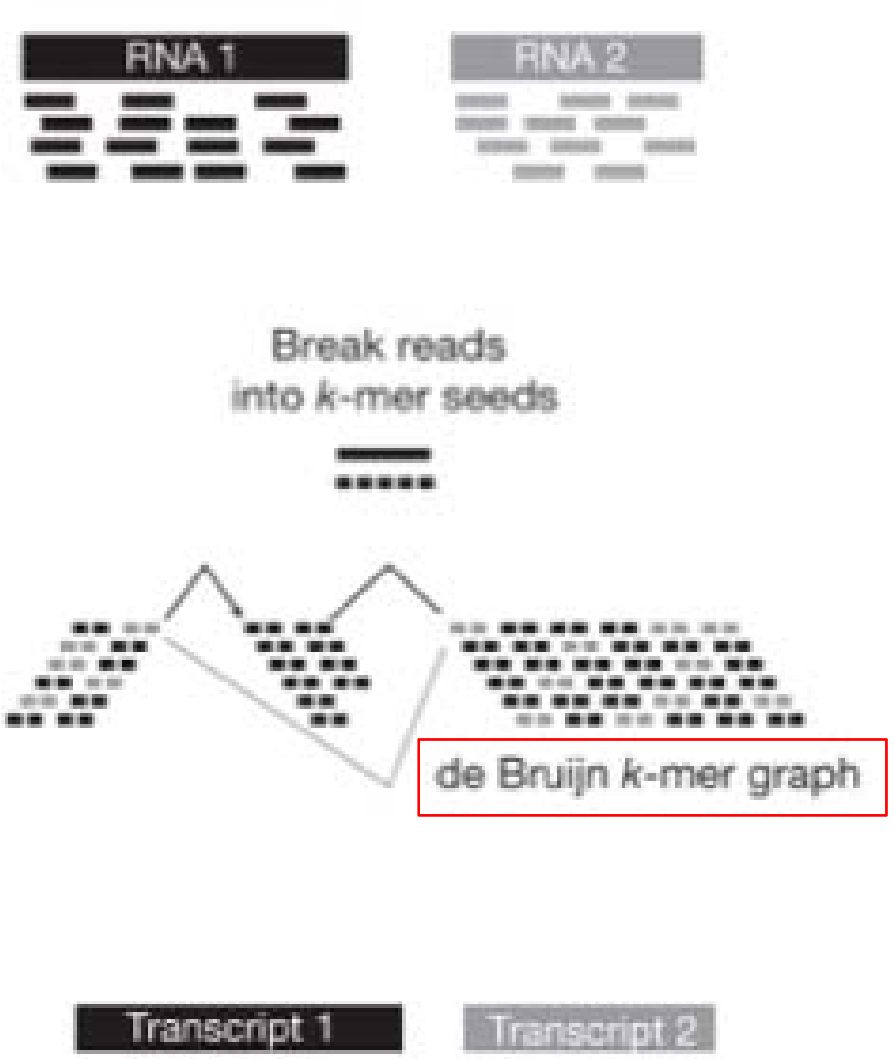
Trapnell, Cole, et al. *Nature biotechnology*, 2010

2. Transcriptome reconstruction

- 2 approaches:
  - 2. *De novo* Assemble
- Usually for species without transcriptomes

2. Transcriptome reconstruction

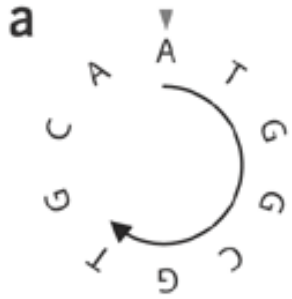
- Approach2:
  - Assemble



Trapnell, Cole, et al. *Nature biotechnology*, 2010

About *de Brujin* graph

- A example




```
$ cat a.fasta
>1
ATGGCGT
>2
GGCGTGC
>3
CGTGCAA
>4
TGCAATG
>5
CAATGGC
```

Input fasta

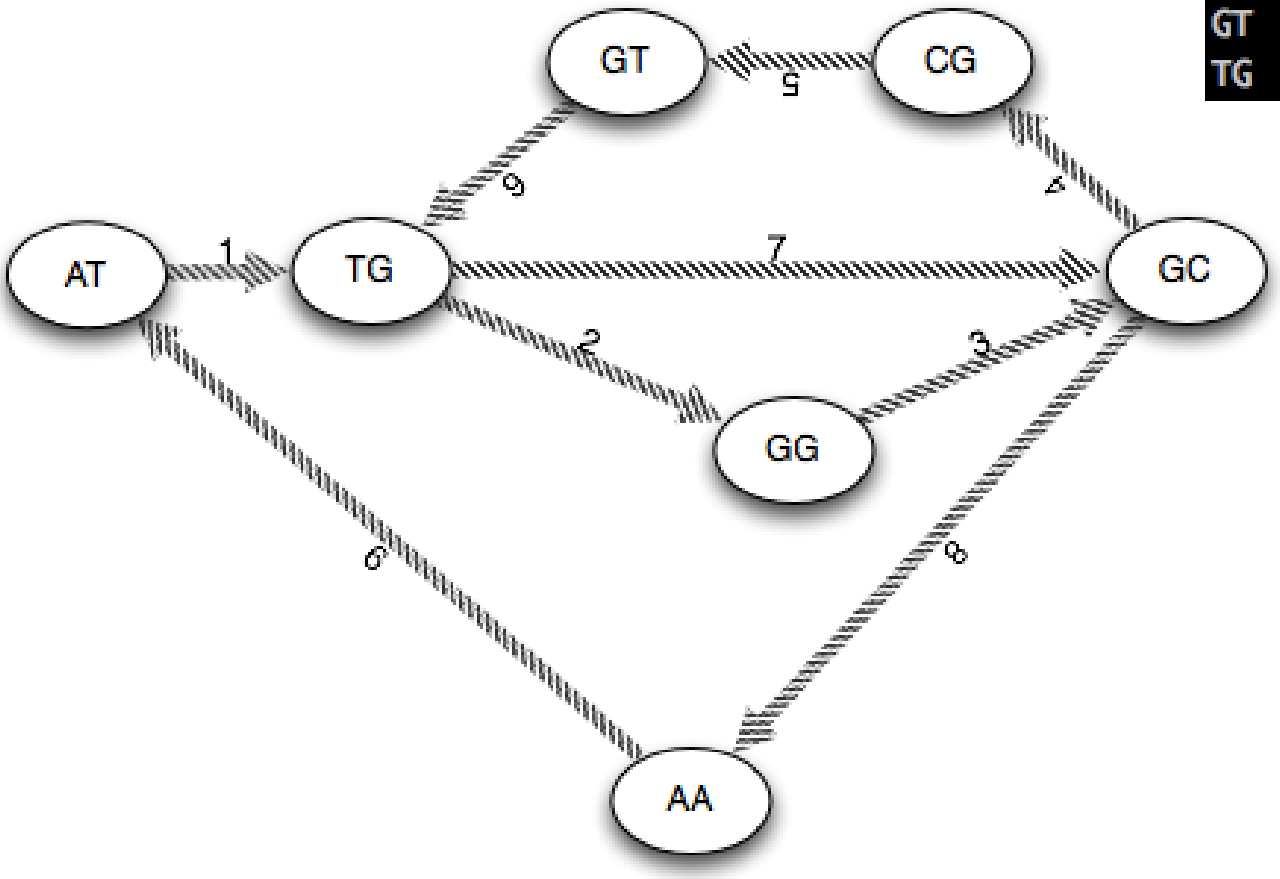
Former dimer => Next dimer

```
Possible combination of 2 consist bases in all sequence
Relationship of 2-bases comb and 3-bases comb
AA      enlonaate with 1 base    AAT
AT      enlon AA  =>    AT       G
              AT  =>    TG
CA      enlon CA  =>    AA       A
CG      enlon CG  =>    GT       T
GC      enlon GC  =>    CA or CG A or GCG
GG      enlon GG  =>    GC       C
              GT  =>    TG
GT      enlon TG  =>    GC or GG G
TG      enlongate with 1 base    TGC or TGG
```

Compeau et al . *Nature biotechnology*, 2011

About de Brujin graph

```
AA  =>  AT
AT  =>  TG
CA  =>  AA
CG  =>  GT
GC  =>  CA or CG
GG  =>  GC
GT  =>  TG
TG  =>  GC or GG
```

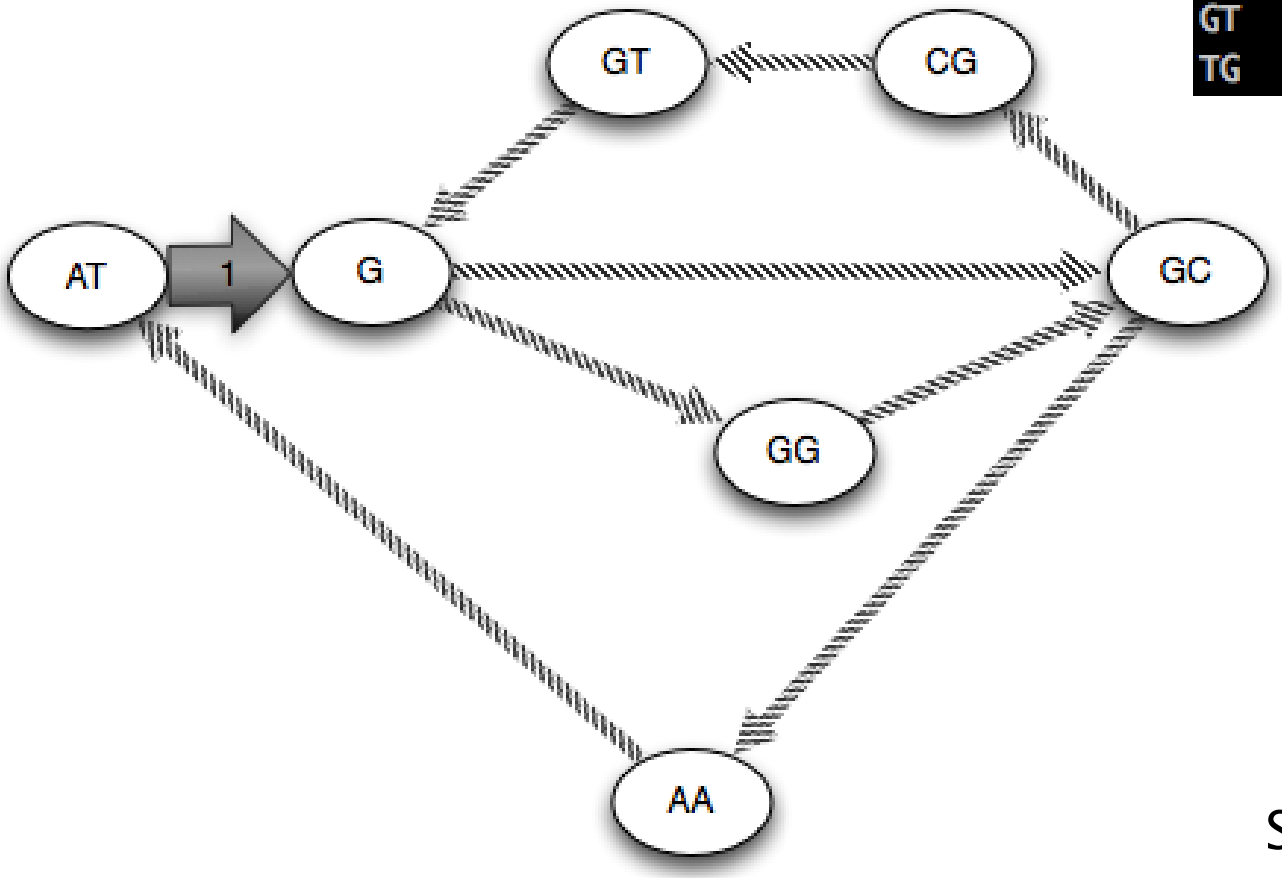



```
$ cat a.fasta
>1
ATGGCGT
>2
GGCGTGC
>3
CGTGCAA
>4
TGCAATG
>5
CAATGGC
```

Compeau et al . *Nature biotechnology*, 2011

About de Brujin graph

```
AA  =>  AT
AT  =>  TG
CA  =>  AA
CG  =>  GT
GC  =>  CA or CG
GG  =>  GC
GT  =>  TG
TG  =>  GC or GG
```



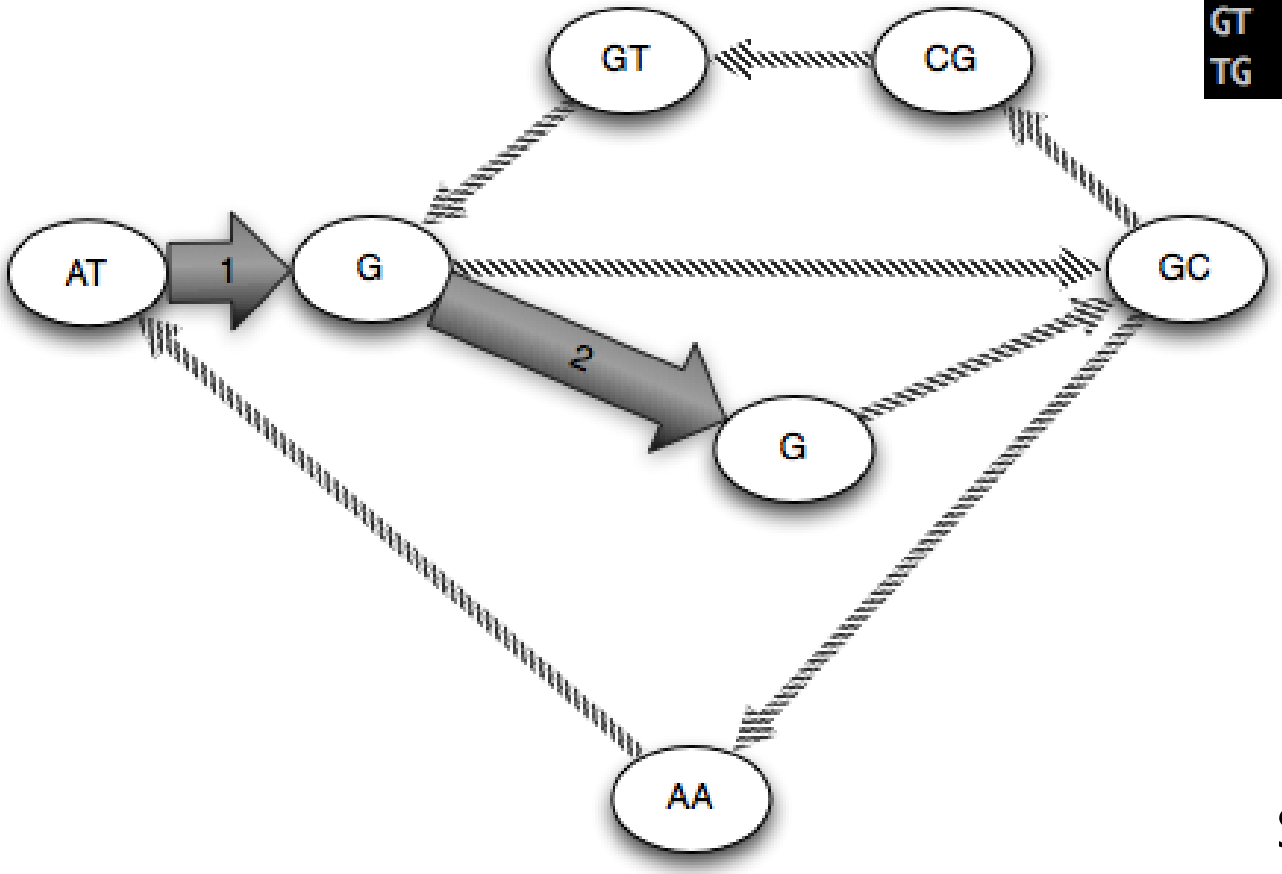

```
$ cat a.fasta
>1
ATGGCGT
>2
GGCGTGC
>3
CGTGCAA
>4
TGCAATG
>5
CAATGGC
```

Sequence: ATG

Compeau et al . *Nature biotechnology*, 2011

About de Brujin graph

```
AA  =>  AT
AT  =>  TG
CA  =>  AA
CG  =>  GT
GC  =>  CA or CG
GG  =>  GC
GT  =>  TG
TG  =>  GC or GG
```



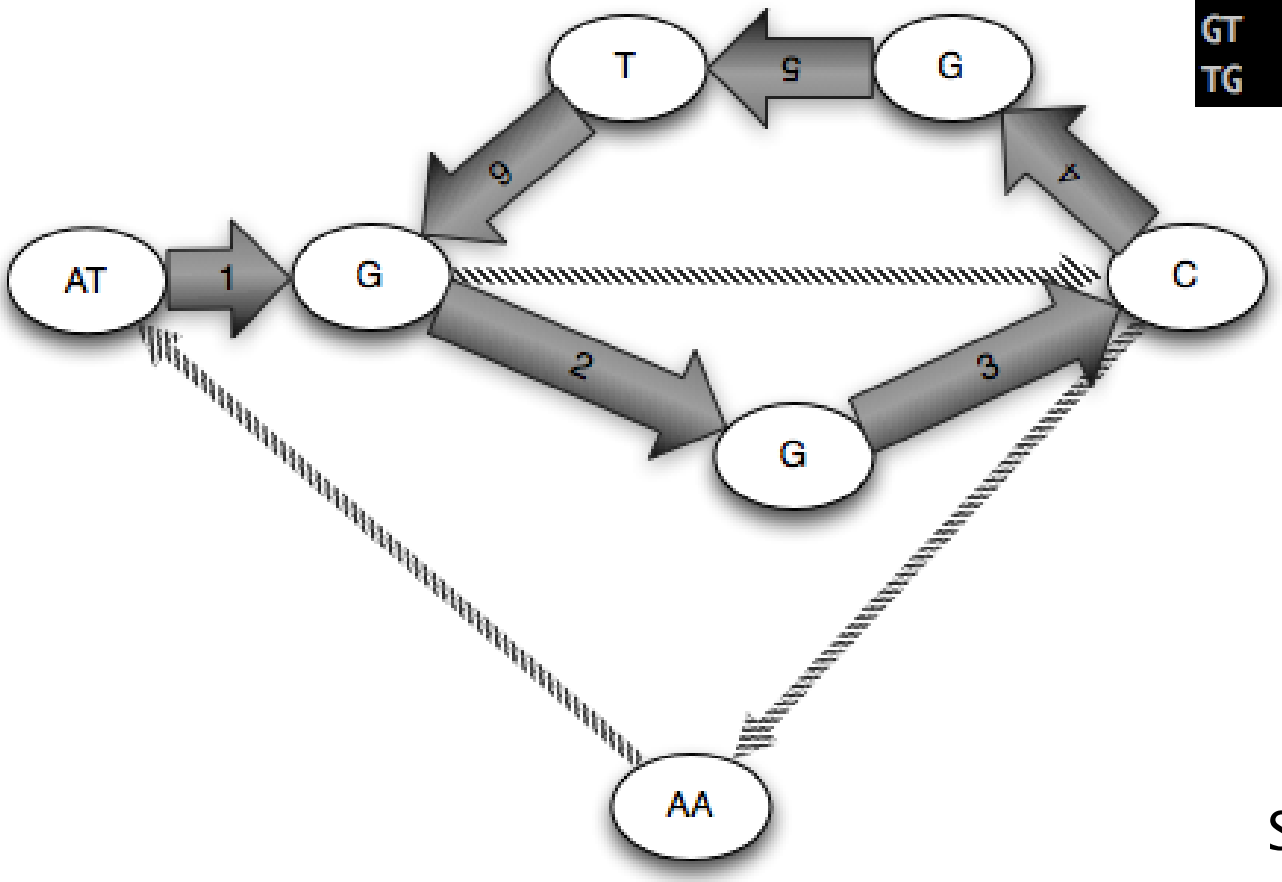

```
$ cat a.fasta
>1
ATGGCGT
>2
GGCGTGC
>3
CGTGCAA
>4
TGCAATG
>5
CAATGGC
```

Sequence: ATGG

Compeau et al . *Nature biotechnology*, 2011

About de Brujin graph

```
AA  =>  AT
AT  =>  TG
CA  =>  AA
CG  =>  GT
GC  =>  CA or CG
GG  =>  GC
GT  =>  TG
TG  =>  GC or GG
```



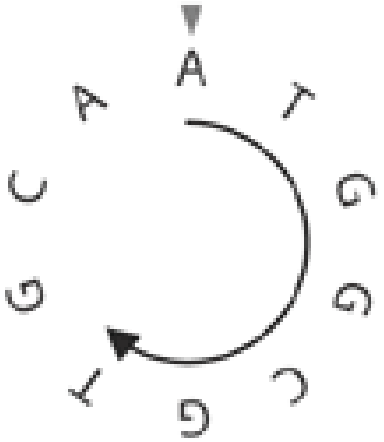

```
$ cat a.fasta
>1
ATGGCGT
>2
GGCGTGC
>3
CGTGCAA
>4
TGCAATG
>5
CAATGGC
```

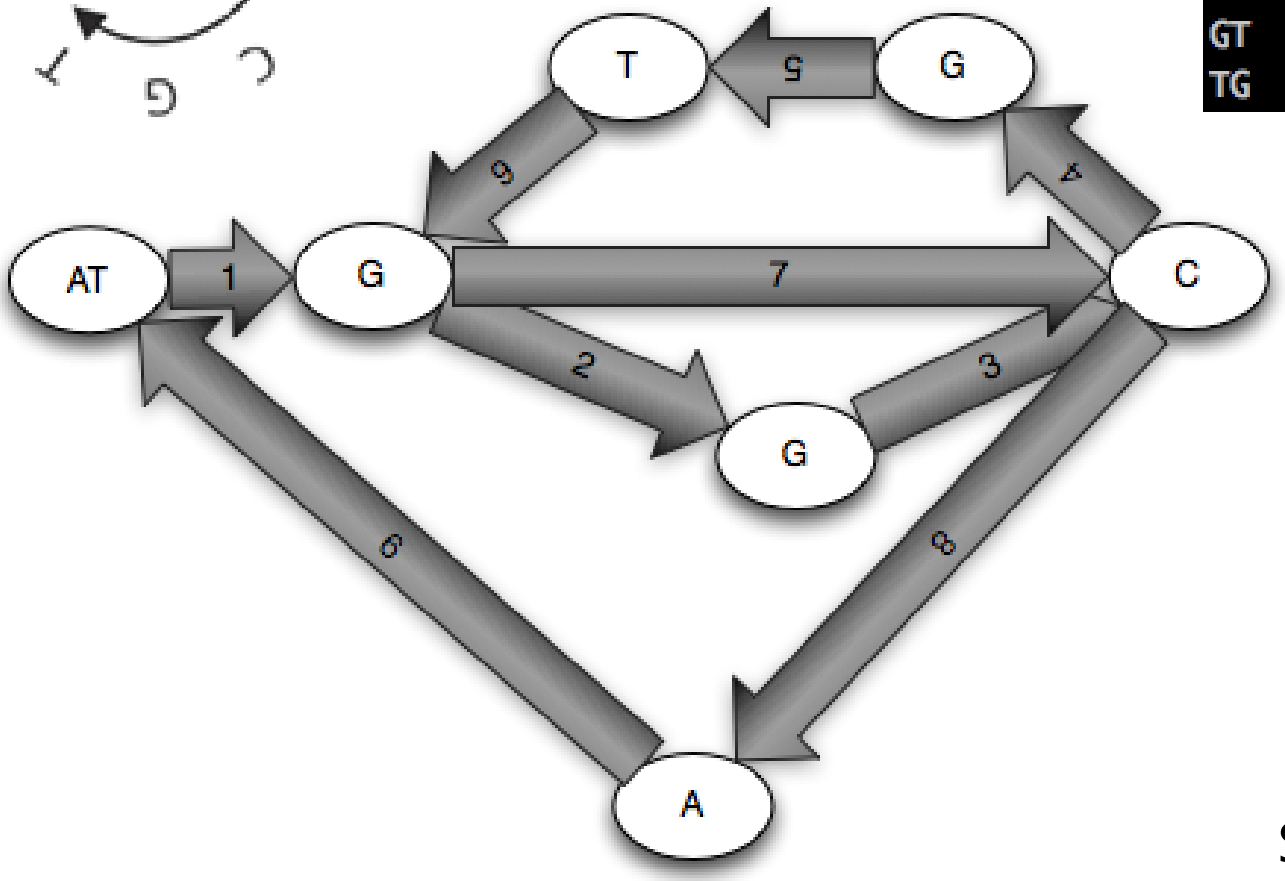
Sequence: ATGGCGTG

Compeau et al . *Nature biotechnology*, 2011

About de Brujin graph

a





AA	=>	AT
AT	=>	TG
CA	=>	AA
CG	=>	GT
GC	=>	CA or CG
GG	=>	GC
GT	=>	TG
TG	=>	GC or GG

```
$ cat a.fasta
>1
ATGGCGT
>2
GGCGTGC
>3
CGTGCAA
>4
TGCAATG
>5
CAATGGC
```

Sequence: ATGGCGTGCA

Compeau et al . *Nature biotechnology*, 2011

2. Transcriptome reconstruction

For gene-isoform detection:

Transcriptome reconstruction

Genome-guided reconstruction	Exon identification Genome-guided assembly	G.Mor.Se Scripture[28] Cufflinks[29]	Assembles exons Reports all isoforms Reports a minimal set of isoforms	Identifying novel transcripts using a known reference genome	Alignments to reference genome
Genome-independent reconstruction	Genome-independent assembly	Velvet[61] TransABySS[56]	Reports all isoforms	Identifying novel genes and transcript isoforms without a known reference genome	Reads

For splicing type (SE, A3SS,A5SS,MXE,TandemUTR,RI,AFE,ALE) detection or Percent Splicing In (PSI) caculation:

Analysis and design of RNA sequencing experiments for identifying isoform regulation

Yarden Katz, Eric T Wang, Edoardo M Airoldi & Christopher B Burge

Affiliations | Contributions | Corresponding authors

Nature Methods 7, 1009–1015 (2010) | doi:10.1038/nmeth.1528

Received 18 August 2010 | Accepted 08 October 2010 | Published online 07 November 2010

Nature **456**, 470-476 (27 November 2008) | doi:10.1038/nature07509; Received 23 June 2008; Accepted 3 October 2008; Published online 2 November 2008; Corrected 27 November 2008

Alternative isoform regulation in human tissue transcriptomes

Eric T. Wang[1,2,7], Rickard Sandberg[1,3,7], Shujun Luo[4], Irina Khrebtukova[4], Lu Zhang[4], Christine Mayr[5], Stephen F. Kingsmore[6], Gary P. Schroth[4] & Christopher B. Burge[1]

3. Differential Expression

- 1. FPKM, the indicator of expression:
  - FPKM, Fragments(pair-reads) Per Kilo-bases per Million-reads total

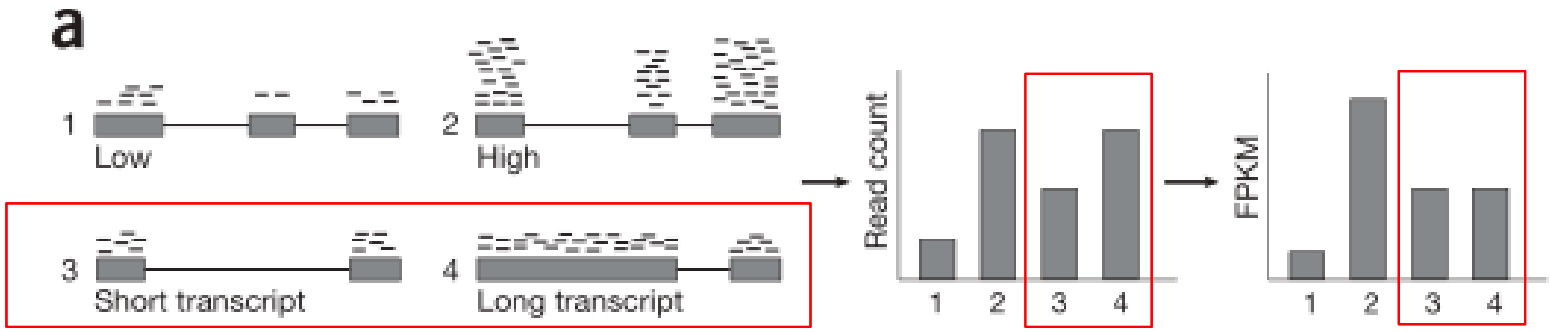
$$FPKM = 10^9 \times \frac{C}{NL}$$

C= the number of reads mapped onto the gene's exons
N= total number of reads in the experiment
L= the sum of the exons in base pairs.

Mortazavi, Ali, et al. *Nature methods* ,2008

3. Differential Expression

- 1. FPKM, the indicator of expression:
- Normalize the effect of length.



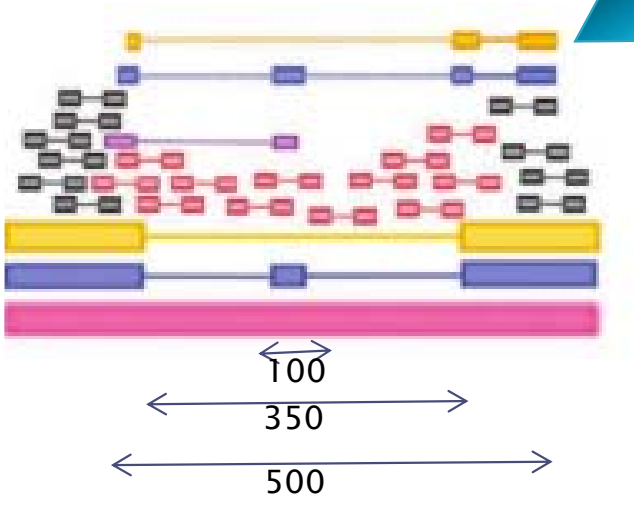
Normalize the effect of different Total Reads in different experiment.

Garber, Manuel, et al. *Nature methods* 2011

3. Differential Expression

- 2. Abundance of different transcripts in a single gene.

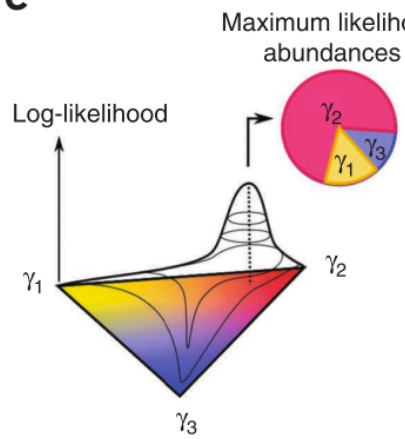
Cufflinks Pair-End sequence



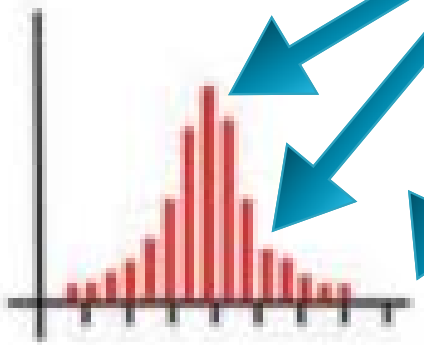
If this reads belong to C, length_in_C = 500,likehood=l1

If this reads belong to B, length_in_B = 250,likehood=l2

If this reads belong to A, length_in_A = 150,likehood=l3



All reads' likehoold of possible transcripts



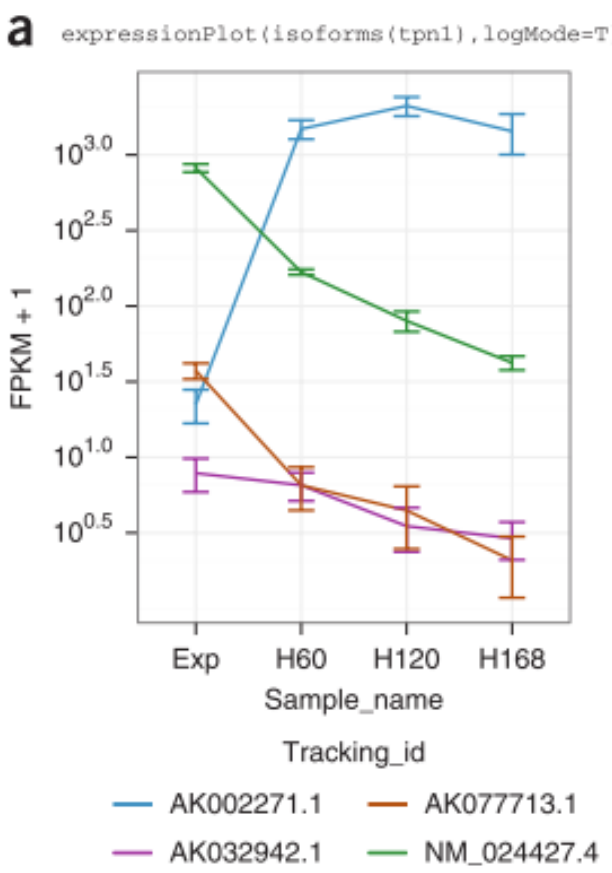
1 150 500

Fragment_length distribution (Normal)

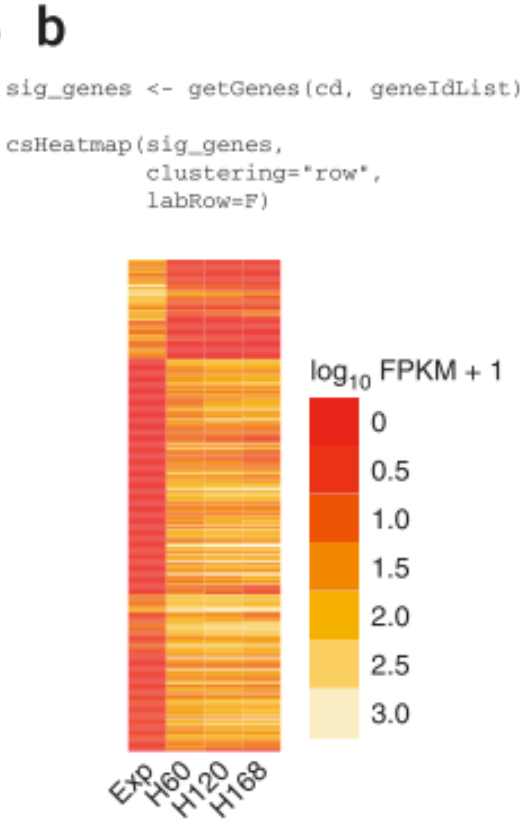
Trapnell, Cole, et al. *Nature biotechnology*, 2010

3. Differential Expression

- 3.Transcripts difference in different samples:



Isoform change for a single gene



Heatmap for Expression difference for a group gene

Trapnell, Cole, et al. *Nature protocols*, 2012

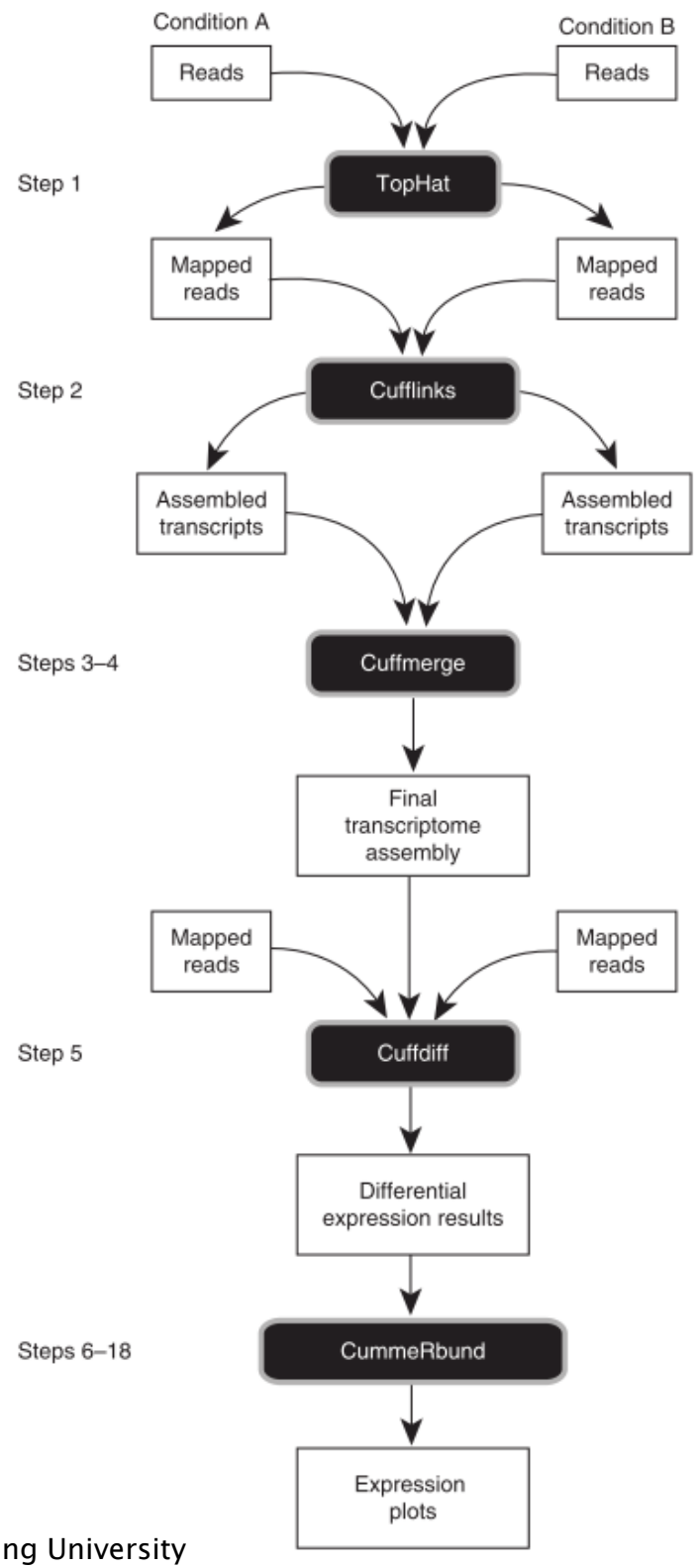
Want have a try?

- A recommend pipeline f

PROTOCOL

Differential gene and transcr of RNA-seq experiments with

Cole Trapnell[1,2], Adam Roberts[3], Loyal Goff[1,2,4], Geo Pertea[5,6], D
Steven L Salzberg[5,6], John L Rinn[1,2] & Lior Pachter[3,8,9]



Trapnell

Reference

- Garber, Manuel, et al. "Computational methods for transcriptome annotation and quantification using RNA-seq." *Nature methods* 8.6 (2011): 469-477.
- Trapnell, Cole, et al. "Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation." *Nature biotechnology* 28.5 (2010): 511-515.
- Compeau, Phillip EC, Pavel A. Pevzner, and Glenn Tesler. "How to apply de Bruijn graphs to genome assembly." Nature biotechnology 29.11 (2011): 987-991.
- Trapnell, Cole, et al. "Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks." *Nature protocols* 7.3 (2012): 562-578.
- Mortazavi, Ali, et al. "Mapping and quantifying mammalian transcriptomes by RNA-Seq." *Nature methods* 5.7 (2008): 621-628.
- Katz, Yarden, et al. "Analysis and design of RNA sequencing experiments for identifying isoform regulation." Nature methods 7.12 (2010): 1009-1015.
- Wang, Eric T., et al. "Alternative isoform regulation in human tissue transcriptomes." Nature 456.7221 (2008): 470-476.

Thank you!

- Tingting Lu陆婷婷
- Rongfeng Zhu祝融峰
- Boqiang Hu胡博强
- Mingshan Liu刘明珊
- Yongzhegn Li李永正
- Yirong Wang王奕蓉
- Chuze Shen沈初泽
- Xiao Wang王潇