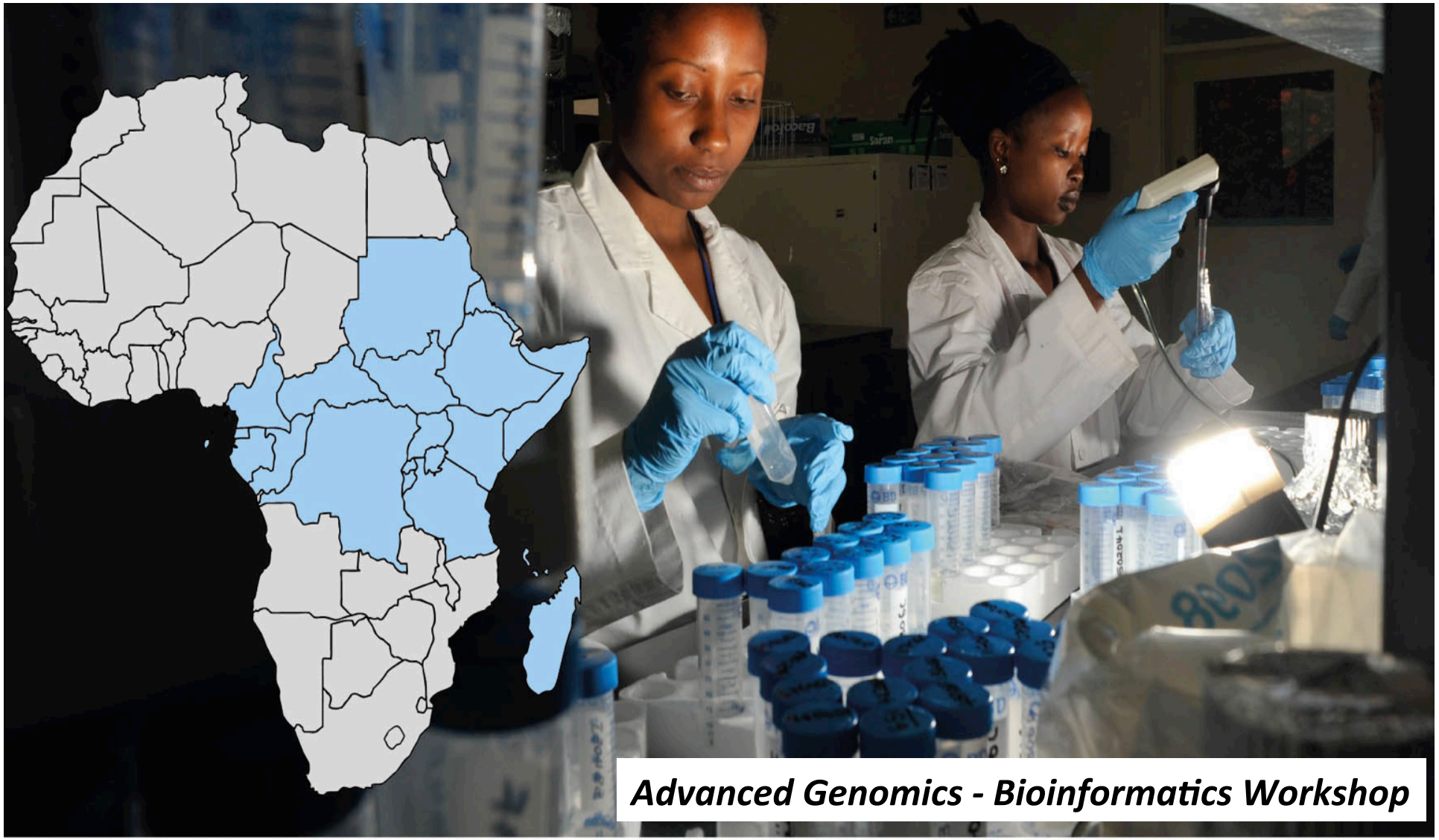


Advanced Genomics - Bioinformatics Workshop

Mark Wamalwa
BecA-ILRI Hub, Nairobi, Kenya
http://hub.africabiosciences.org/
m.wamalwa@cgiar.org

7th – 18th September 2015

ILRI INTERNATIONAL LIVESTOCK RESEARCH INSTITUTE

African Union

NEPAD A PROGRAMME OF THE AFRICAN UNION

biosciences eastern and central africa

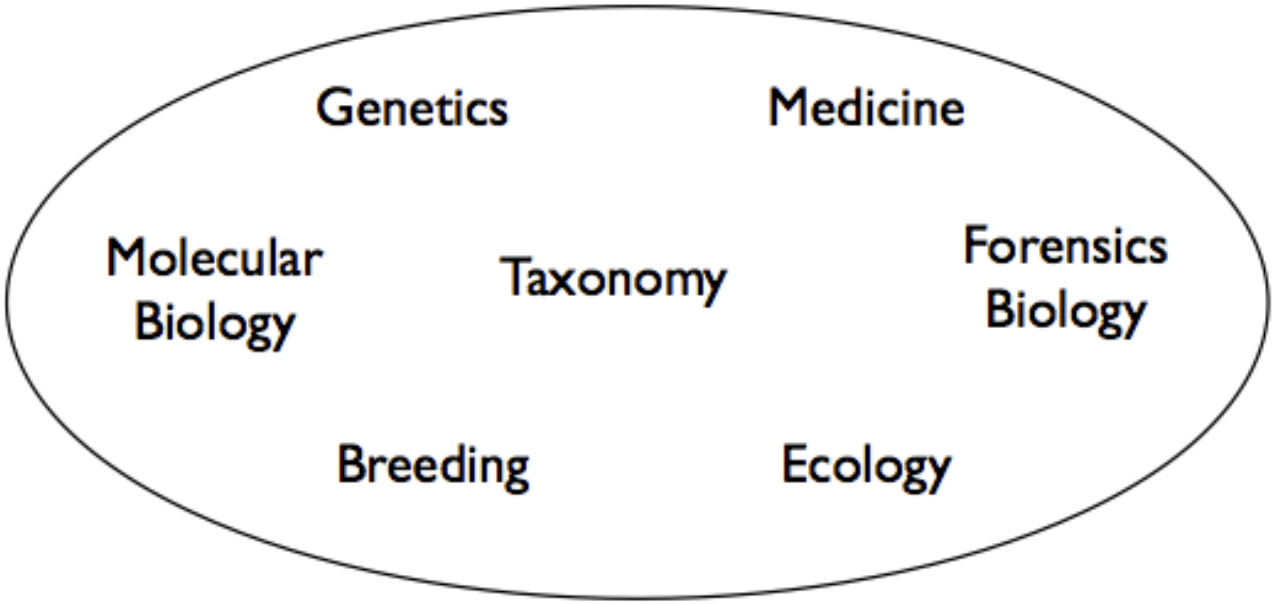
RNA sequencing, transcriptome and expression quantification

Lecture Overview

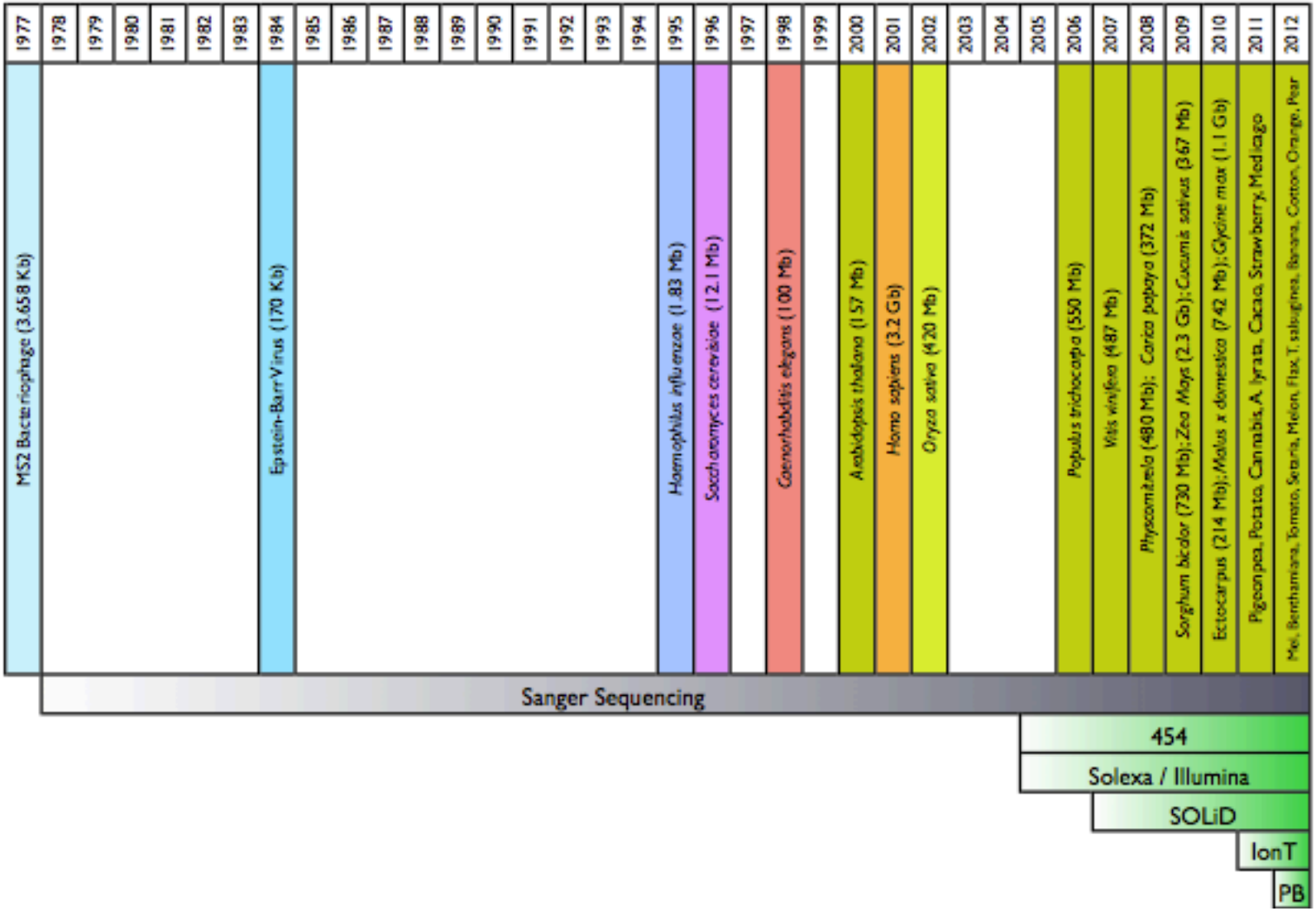
- What is RNA-seq?
- Basic concepts
- Mapping-based transcriptomics (genome - based)
- De novo based transcriptomics (genome-free)
- Expression counts and differential expression
- Transcript annotation

Recap - Basics of the Next Generation Sequencing (NGS)

- ***DNA Sequencing:*** "Process of determining the precise order of nucleotides within a DNA molecule." -Wikipedia



Recap - Basics of the Next Generation Sequencing (NGS)



Features of Next Generation Sequencing.

1. Massive sequence production (from 0.1 to 300 Gb).

2. Wide range of sequence lengths (from 50 to 3,000 bp).

3. Same or bigger error rate that the traditional sequencing (from 87 to 99.9%).

4. Cheap price per base.

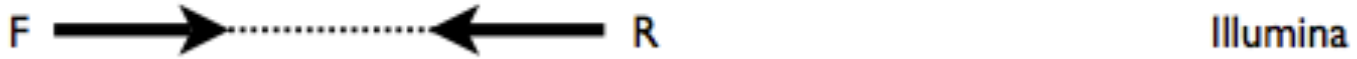
Features of Next Generation Sequencing.

	Inputs	Outputs
454 Pyrosequencing (GS FLX Titanium XL+)	– Single Reads Library. – Pair End Library (3 to 20 Kb insert size). – Multiplexed sample.	– sff files – (fasta and fastq files)
Illumina (HiSeq 2500)	– Single Reads Library. – Pair End Library (170-800 bp insert size). – Mate Pair Library (2 to 10 Kb insert Size) – Multiplexed sample.	– fastq files (Phred+64) – fastq files (Phred+33, Illumina 1.8+)
Illumina (MiSeq)		
SOLID (5500xl system)	– Single Reads Library. – Mate Pairs Library (0.6 to 6 Kb insert size). – Multiplexed sample.	– fastq files (Phred+33)
Ion Torrent (Ion Proton I)	– Single Reads Library. – Pair End Library (0.6 to 6 Kb insert size). – Multiplexed sample.	
PacBio (PacBioRS)	– Single Reads Library.	

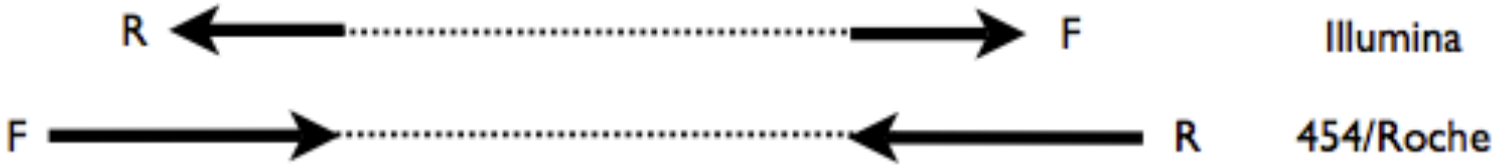
Library types (orientations)

- Single reads

- Pair ends (PE) (150-800 bp insert size)



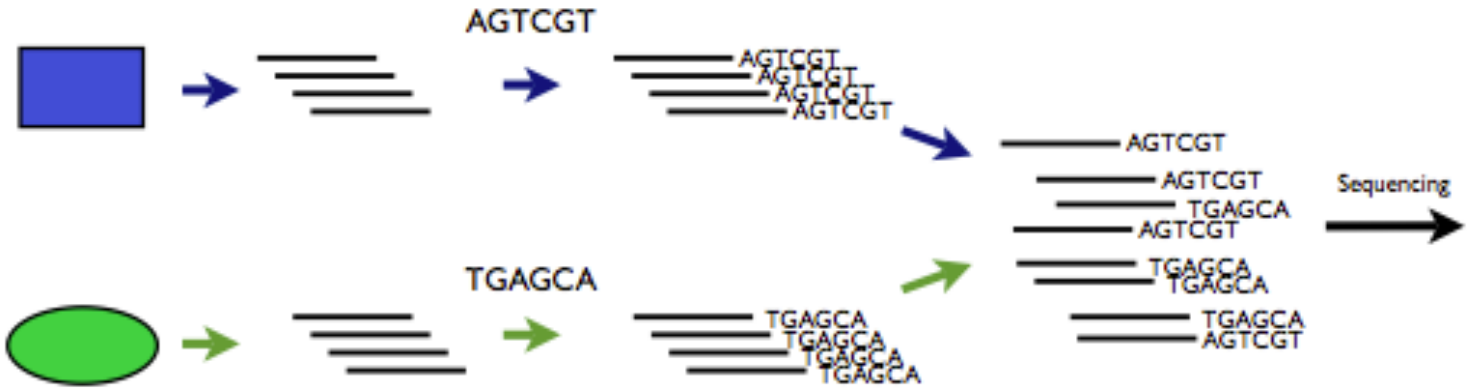
- Mate pairs (MP) (2-20 Kb insert size)



Library types (orientations)

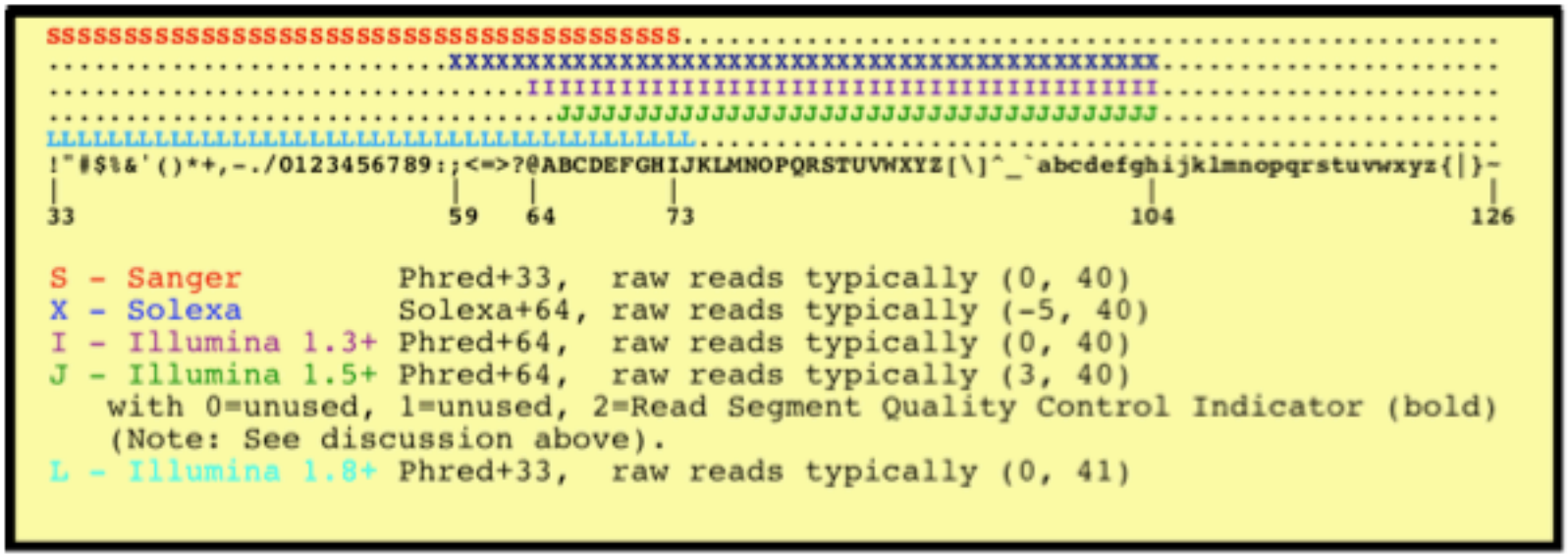
★ Multiplexing:

Use of different tags (4-6 nucleotides) to identify different samples in the same lane/ sector.



File Formats

Fastq files:



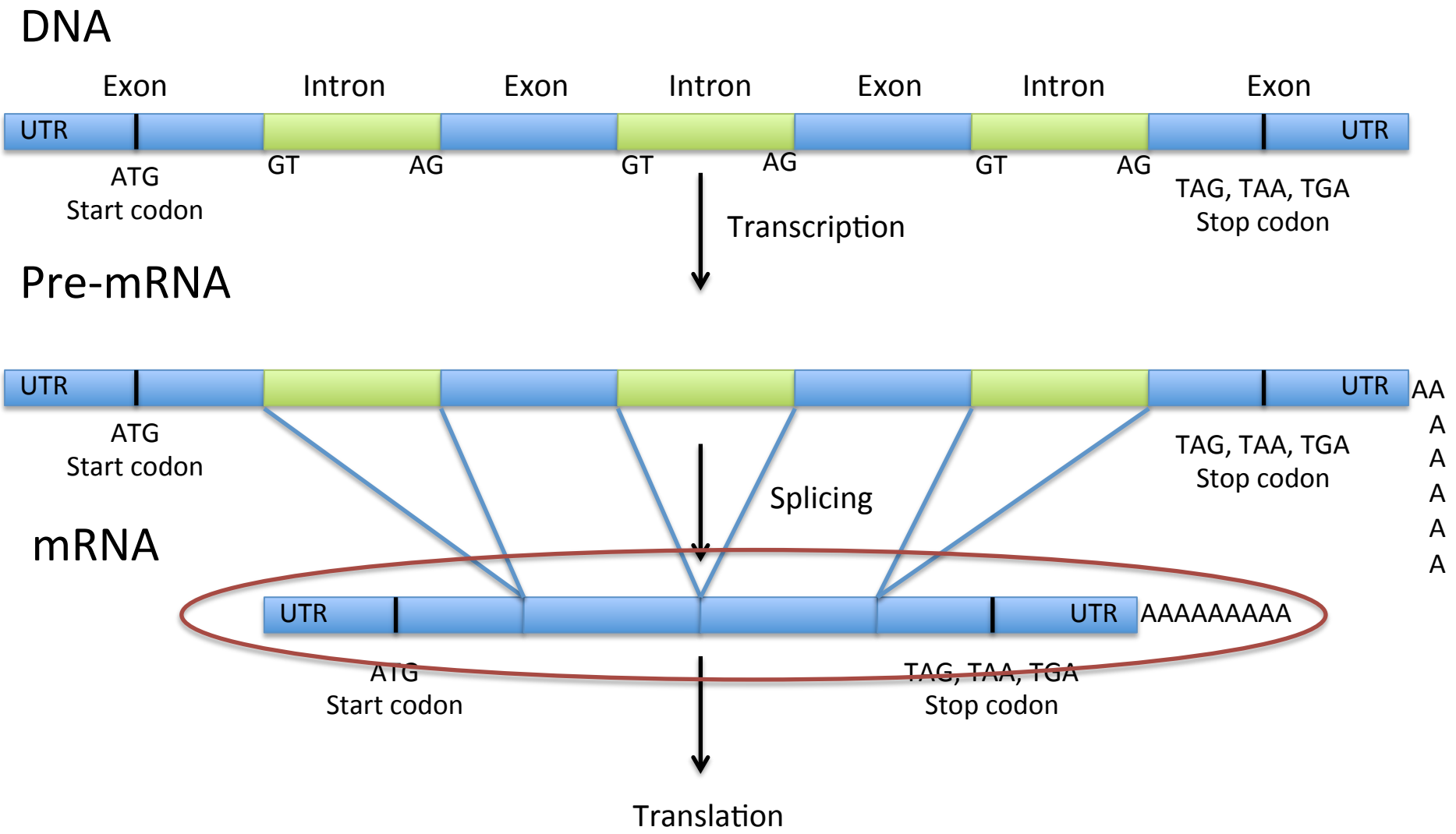
Phred score of a base is: **Qphred=-10 log10 (e)**

Q=15	e=0.03	(min. used Sanger)
Q=20	e=0.01	(min. used 454 and Illumina)
Q=30	e=0.001	(standard used 454)

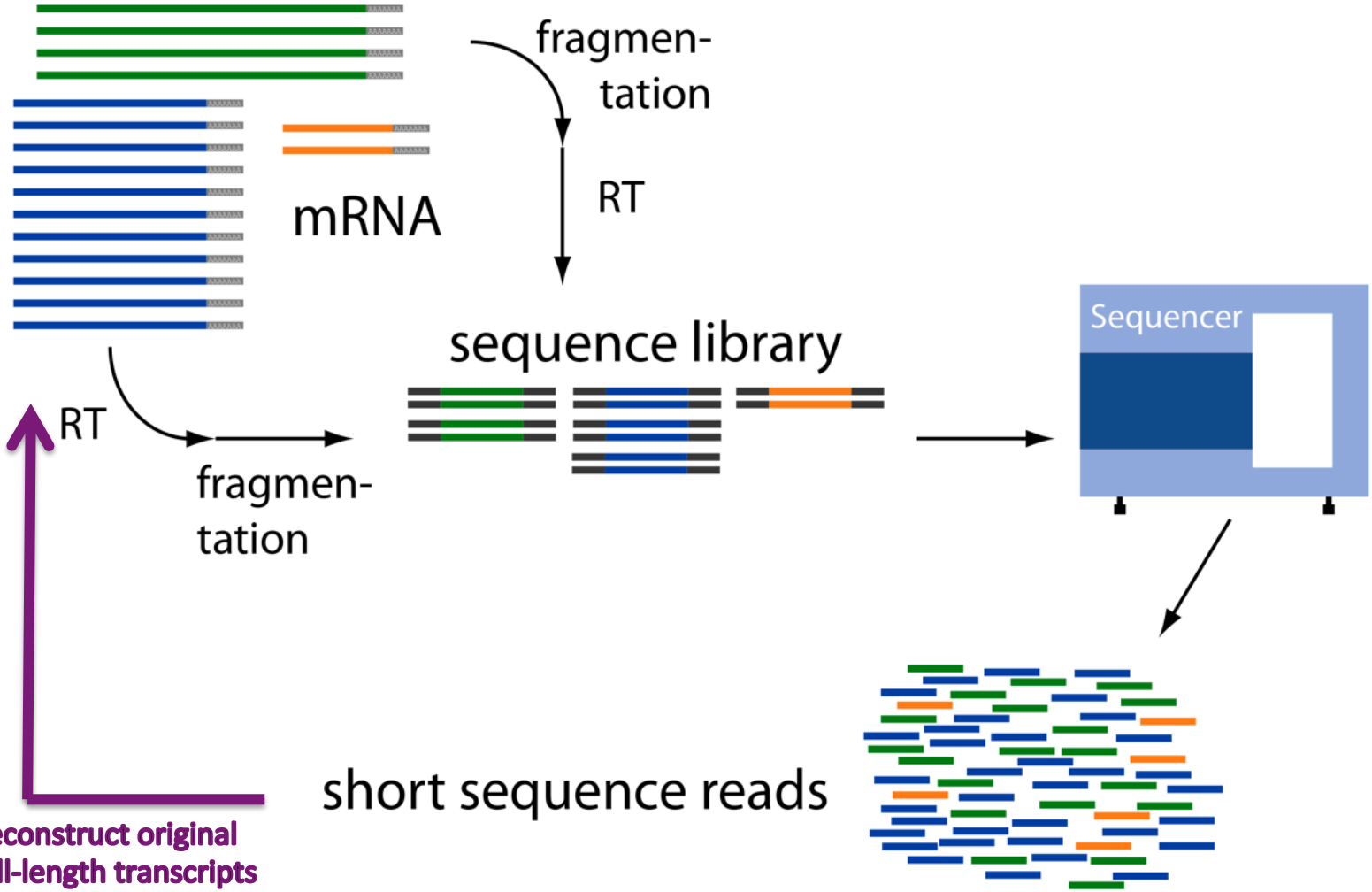
Lecture Overview

- What is RNA-seq?
- Basic concepts
- Mapping-based transcriptomics (genome - based)
- De novo based transcriptomics (genome-free)
- Expression counts and differential expression
- Transcript annotation

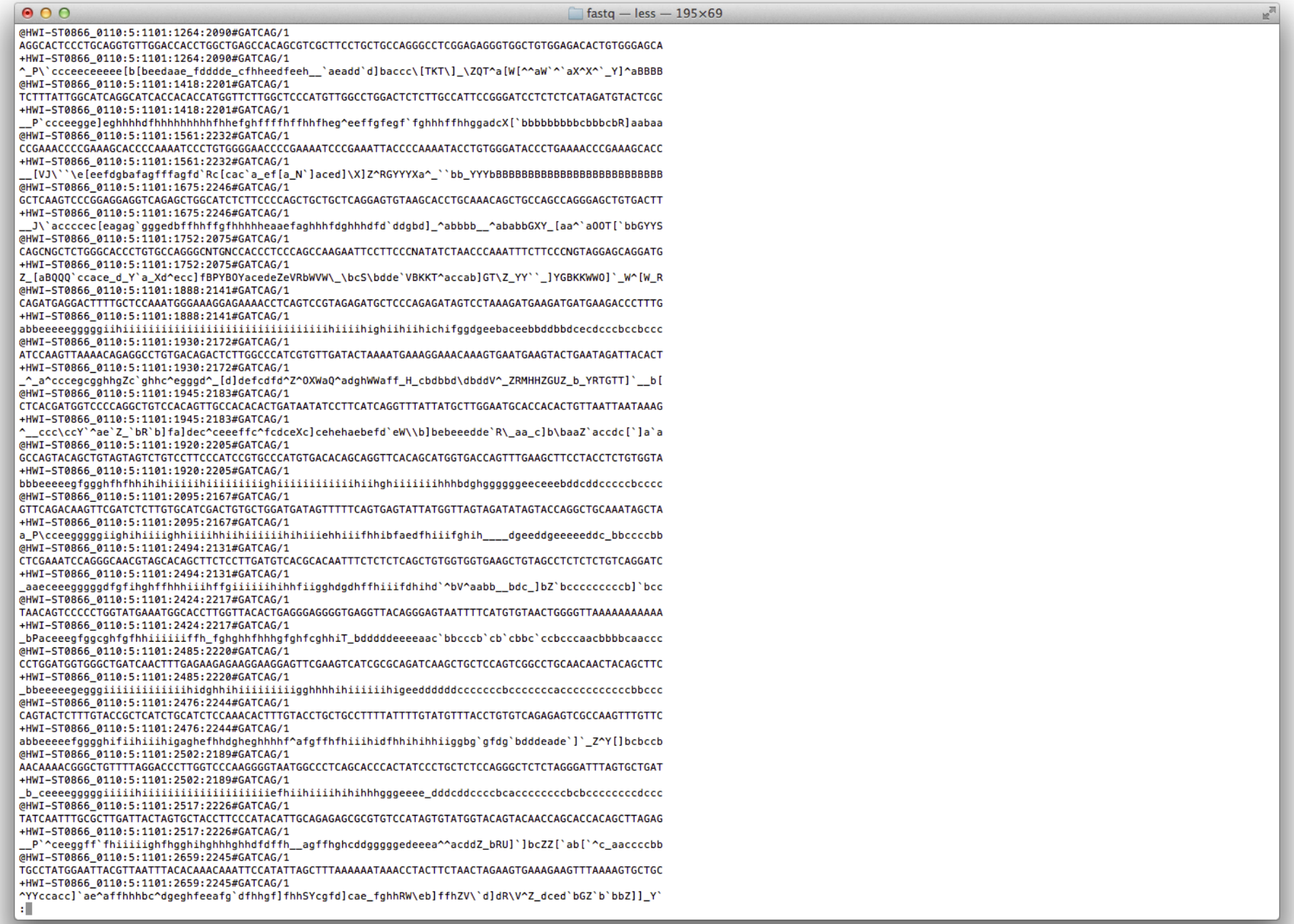
RNA-seq


DNA
Exon
Intron
Exon
Intron
Exon
Intron
Exon
UTR
UTR
ATG
Start codon
GT
AG
GT
AG
GT
AG
TAG, TAA, TGA
Stop codon
Transcription
Pre-mRNA
UTR
UTR
AA
A
A
A
A
A
ATG
Start codon
TAG, TAA, TGA
Stop codon
Splicing
mRNA
UTR
UTR
AAAAAAAAA
ATG
Start codon
TAG, TAA, TGA
Stop codon
Translation

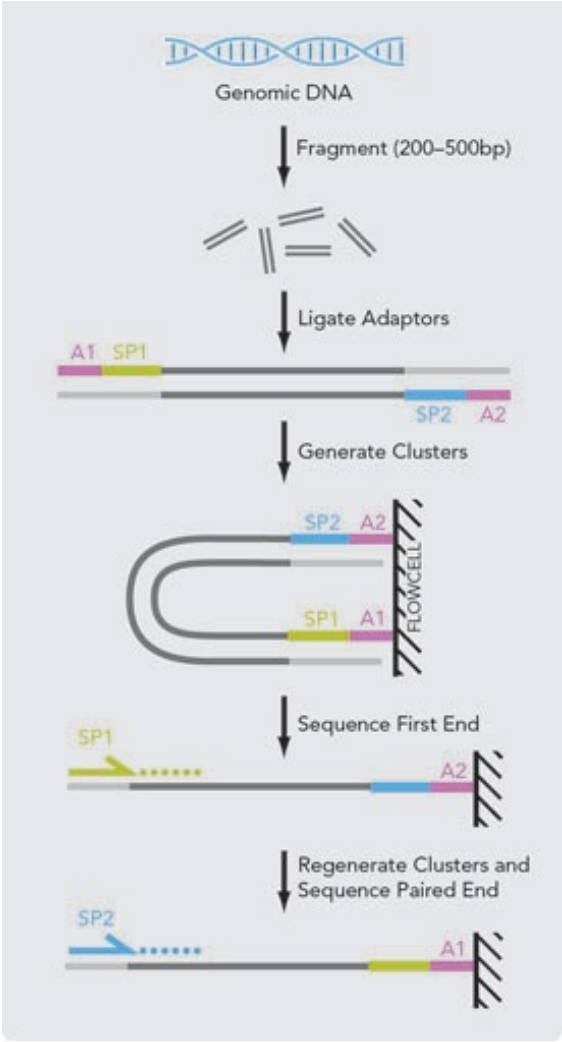
Overview of RNA-Seq



From: http://www2.fml.tuebingen.mpg.de/raetsch/members/research/transcriptomics.html

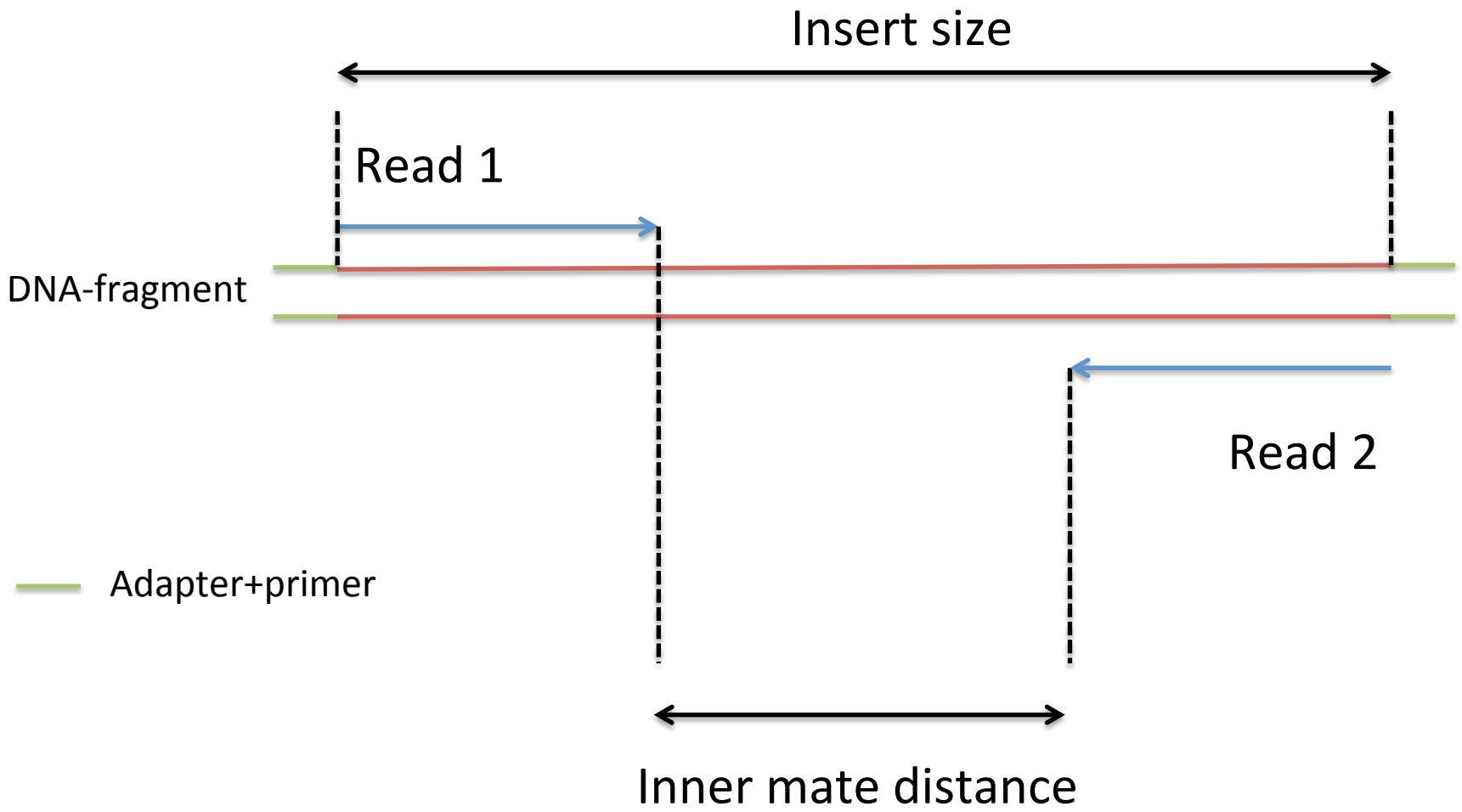

fastq — less — 195×69
@HWI-ST0866_0110:5:1101:1264:2090#GATCAG/1
AGGCACTCCCTGCAGGTGTTGGACCACCTGGCTGAGCCACAGCGTCGCTTCCTGCTGCCAGGGCCTCGGAGAGGGTGGCTGTGGAGACACTGTGGGAGCA
+HWI-ST0866_0110:5:1101:1264:2090#GATCAG/1
^_P\`ccceeceeeee[b[beedaae_fdddde_cfhheedfeeh__`aeadd`d]baccc\[TKT\]_\ZQT^a[W[^^aW`^`aX^X^`_Y]^aBBBB
@HWI-ST0866_0110:5:1101:1418:2201#GATCAG/1
TCTTTATTGGCATCAGGCATCACCACACCATGGTTCTTGGCTCCCATGTTGGCCTGGACTCTCTTGCCATTCCGGGATCCTCTCTCATAGATGTACTCGC
+HWI-ST0866_0110:5:1101:1418:2201#GATCAG/1
__P`cccceegge]eghhhhdfhhhhhhhhhfhhefghffffhffhhfheg^eeffgfegf`fghhhffhhggadcX[`bbbbbbbbbcbbbcbR]aabaa
@HWI-ST0866_0110:5:1101:1561:2232#GATCAG/1
CCGAAACCCCGAAAGCACCCCAAAATCCCTGTGGGGAACCCCGAAAATCCCGAAATTACCCCAAAATACCTGTGGGATACCCTGAAAACCCGAAAGCACC
+HWI-ST0866_0110:5:1101:1561:2232#GATCAG/1
__[VJ\``\e[eefdgbafagfffagfd`Rc[cac`a_ef[a_N`]aced]\X]Z^RGYYYXa^_``bb_YYYbBBBBBBBBBBBBBBBBBBBBBBBBBB
@HWI-ST0866_0110:5:1101:1675:2246#GATCAG/1
GCTCAAGTCCCGGAGGAGGTCAGAGCTGGCATCTCTTCCCCAGCTGCTGCTCAGGAGTGTAAGCACCTGCAAACAGCTGCCAGCCAGGGAGCTGTGACTT
+HWI-ST0866_0110:5:1101:1675:2246#GATCAG/1
__J\`accccec[eagag`gggedbffhhffgfhhhhheaaefaghhhfdghhhdfd`ddgbd]_^abbbb__^ababbGXY_[aa^`aOOT[`bbGYYS
@HWI-ST0866_0110:5:1101:1752:2075#GATCAG/1
CAGCNGCTCTGGGCACCCTGTGCCAGGGCNTGNCCACCCTCCCAGCCAAGAATTCCTTCCCNATATCTAACCCAAATTTCTTCCCNGTAGGAGCAGGATG
+HWI-ST0866_0110:5:1101:1752:2075#GATCAG/1
Z_[aBQQQ`ccace_d_Y`a_Xd^ecc]fBPYBOYacedeZeVRbWVW_\bcS\bdde`VBKKT^accab]GT\Z_YY``_]YGBKKWWO]`_W^[W_R
@HWI-ST0866_0110:5:1101:1888:2141#GATCAG/1
CAGATGAGGACTTTTGCTCCAAATGGGAAAGGAGAAAACCTCAGTCCGTAGAGATGCTCCCAGAGATAGTCCTAAAGATGAAGATGATGAAGACCCTTTG
+HWI-ST0866_0110:5:1101:1888:2141#GATCAG/1
abbeeeeeggggiihiiiiiiiiiiiiiiiiiiiiiiiiiiiiiihiiiihighiihiihichifggdgeebaceebbddbbdcecdcccbccbccc
@HWI-ST0866_0110:5:1101:1930:2172#GATCAG/1
ATCCAAGTTAAAACAGAGGCCTGTGACAGACTCTTGGCCCATCGTGTTGATACTAAAATGAAAGGAAACAAAGTGAATGAAGTACTGAATAGATTACACT
+HWI-ST0866_0110:5:1101:1930:2172#GATCAG/1
_^_a^cccegcgghhgZc`ghhc^egggd^_[d]defcdfd^Z^OXWaQ^adghWWaff_H_cbdbbd\dbddV^_ZRMHHZGUZ_b_YRTGTT]`__b[
@HWI-ST0866_0110:5:1101:1945:2183#GATCAG/1
CTCACGATGGTCCCCAGGCTGTCCACAGTTGCCACACACTGATAATATCCTTCATCAGGTTTATTATGCTTGGAATGCACCACACTGTTAATTAATAAAG
+HWI-ST0866_0110:5:1101:1945:2183#GATCAG/1
^__ccc\ccY`^ae`Z_`bR`b]fa]dec^ceeeffc^fcdceXc]cehehaebefd`eW\\b]bebeeedde`R_aa_c]b\baaZ`accdc[`]a`a
@HWI-ST0866_0110:5:1101:1920:2205#GATCAG/1
GCCAGTACAGCTGTAGTAGTCTGTCCTTCCCATCCGTGCCCATGTGACACAGCAGGTTCACAGCATGGTGACCAGTTTGAAGCTTCCTACCTCTGTGGTA
+HWI-ST0866_0110:5:1101:1920:2205#GATCAG/1
bbbeeeeegfggghfhfhhihiiiiihiiiiiiiiighiiiiiiiiiiihiihghiiiiiihhhbdghgggggeeceeebddcddccccbccccc
@HWI-ST0866_0110:5:1101:2095:2167#GATCAG/1
GTTCAGACAAGTTCGATCTCTTGTGCATCGACTGTGCTGGATGATAGTTTTTCAGTGAGTATTATGGTTAGTAGATATAGTACCAGGCTGCAAATAGCTA
+HWI-ST0866_0110:5:1101:2095:2167#GATCAG/1
a_P\ccegggggiighihiiiighhiiiihhiihiiiiihihiiiehhiiifhhibfaedfhiiifghih____dgeeddgeeeeeddc_bbccccbb
@HWI-ST0866_0110:5:1101:2494:2131#GATCAG/1
CTCGAAATCCAGGGCAACGTAGCACAGCTTCTCCTTGATGTCACGCACAATTTCTCTCTCAGCTGTGGTGGTGAAGCTGTAGCCTCTCTCTGTCAGGATC
+HWI-ST0866_0110:5:1101:2494:2131#GATCAG/1
_aaeceeeggggdfgfihghffhhhiiihffgiiiiihihhfiigghdgdhffhiiifdhihd`^bV^aabb__bdc_]bZ`bccccccccb]`bcc
@HWI-ST0866_0110:5:1101:2424:2217#GATCAG/1
TAACAGTCCCCCTGGTATGAAATGGCACCTTGGTTACACTGAGGGAGGGGTGAGGTTACAGGGAGTAATTTTCATGTGTAACTGGGGTTAAAAAAAAAAA
+HWI-ST0866_0110:5:1101:2424:2217#GATCAG/1
_bPaceeegfggcghfgfhhiiiiiiffh_fghghhfhhhgfghfcghhiT_bddddeeeeaac`bbcccb`cb`cbbc`ccbcccaacbbbbcaaccc
@HWI-ST0866_0110:5:1101:2485:2220#GATCAG/1
CCTGGATGGTGGGCTGATCAACTTTGAGAAGAGAAGGAAGGAGTTCGAAGTCATCGCGCAGATCAAGCTGCTCCAGTCGGCCTGCAACAACTACAGCTTC
+HWI-ST0866_0110:5:1101:2485:2220#GATCAG/1
_bbeeeeegeggiiiiiiiiiiiihidghhihiiiiiiiiigghhhhihiiiiiihigeeddddddccccccccbccccccaccccccccccccbbccc
@HWI-ST0866_0110:5:1101:2476:2244#GATCAG/1
CAGTACTCTTTGTACCGCTCATCTGCATCTCCAAACACTTTGTACCTGCTGCCTTTTATTTTGTATGTTTACCTGTGTCAGAGAGTCGCCAAGTTTGTTC
+HWI-ST0866_0110:5:1101:2476:2244#GATCAG/1
abbeeeeefgggghfiihiiihigaghefhhdgheghhhhf^afgffhfhiiihidfhhihihhiiggbg`gfdg`bdddeade`]`_Z^Y[]bcbccb
@HWI-ST0866_0110:5:1101:2502:2189#GATCAG/1
AACAAAACGGGCTGTTTTAGGACCCTTGGTCCCAAGGGGTAATGGCCCTCAGCACCCACTATCCCTGCTCTCCAGGGCTCTCTAGGGATTTAGTGCTGAT
+HWI-ST0866_0110:5:1101:2502:2189#GATCAG/1
_b_ceeeeggggiiiiihiiiiiiiiiiiiiiiiiiefhiihiiiihihihhhgggeeee_dddcddccccbcaccccccccbcbcccccccccdccc
@HWI-ST0866_0110:5:1101:2517:2226#GATCAG/1
TATCAATTTGCGCTTGATTACTAGTGCTACCTTCCCATACATTGCAGAGAGCGCGTGTCCATAGTGTATGGTACAGTACAACCAGCACCACAGCTTAGAG
+HWI-ST0866_0110:5:1101:2517:2226#GATCAG/1
__P`^ceeggff`fhiiiiighfhgghihghhhghhdfdffh__agffhghcddgggggedeeea^^acddZ_bRU]`]bcZZ[`ab[`^c_aaccccbb
@HWI-ST0866_0110:5:1101:2659:2245#GATCAG/1
TGCCTATGGAATTACGTTAATTTACACAAACAAATTCCATATTAGCTTTAAAAAATAAACCTACTTCTAACTAGAAGTGAAAGAAGTTTAAAAGTGCTGC
+HWI-ST0866_0110:5:1101:2659:2245#GATCAG/1
^YYccacc]`ae^affhhhbc^dgeghfeeafg`dfhhgf]fhhSYcgfd]cae_fghhRW\eb]ffhZV\`d]dR\V^Z_dced`bGZ`b`bbZ]]_Y`
:

Paired-End



BILS
Bioinformatics
Infrastructure
for Life Sciences

Insert size


Insert size
Read 1
DNA-fragment
Read 2
Adapter+primer
Inner mate distance

Paired-end gives you two files

FASTQ format (old):

```
@61DFRAAXX100204:1:100:10494:3070/1
AAACAACAGGGCACATTGTCACTCTTGTATTTGAAAAACACTTTCCGGCCAT
+
ACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCBC?CCCCCCCCC@@CACCCCCA
```

```
@61DFRAAXX100204:1:100:10494:3070/2
ATCCAAGTTAAAACAGAGGCCTGTGACAGACTCTTGGCCCATCGTGTTGATA
+
_^_a^cccegcgghhgZc`ghhc^egggd^_[d]defcdfd^Z^OXWaQ^ad
```

New: @<instrument>:<run number>:<flowcell ID>:<lane>:<tile>:<x-pos>:<y-pos>
<read>:<is filtered>:<control number>:<sample number>

Example:
@SIM:1:FCX:1:15:6329:1045 1:N:0:2
TCGCACTCAACGCCCTGCATATGACAAGACAGAATC
+
<>;##=><9=AAAAAAAAAA9#:<#<;<<<????#=

Transcript Reconstruction from RNA-Seq Reads

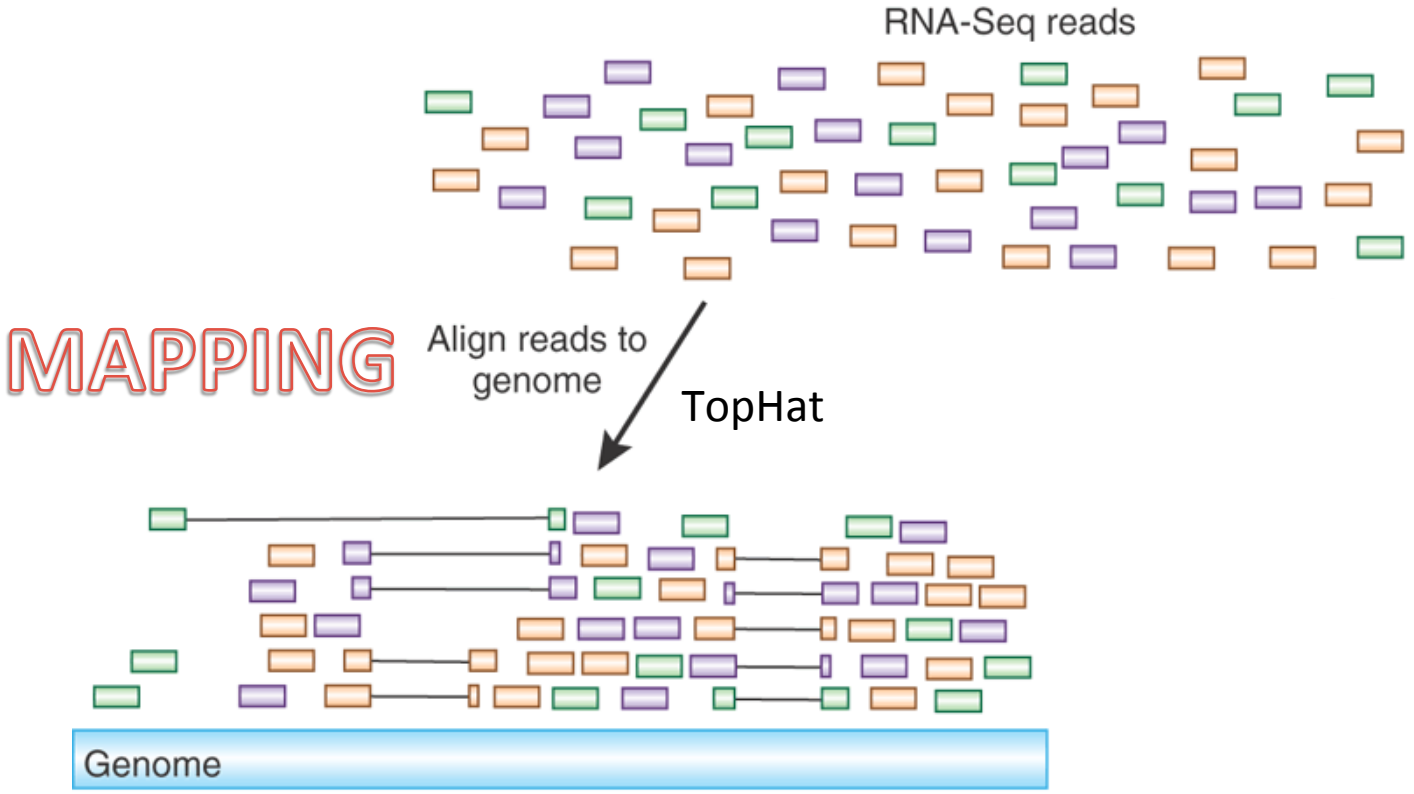
Advancing RNA-Seq analysis

Brian J Haas & Michael C Zody

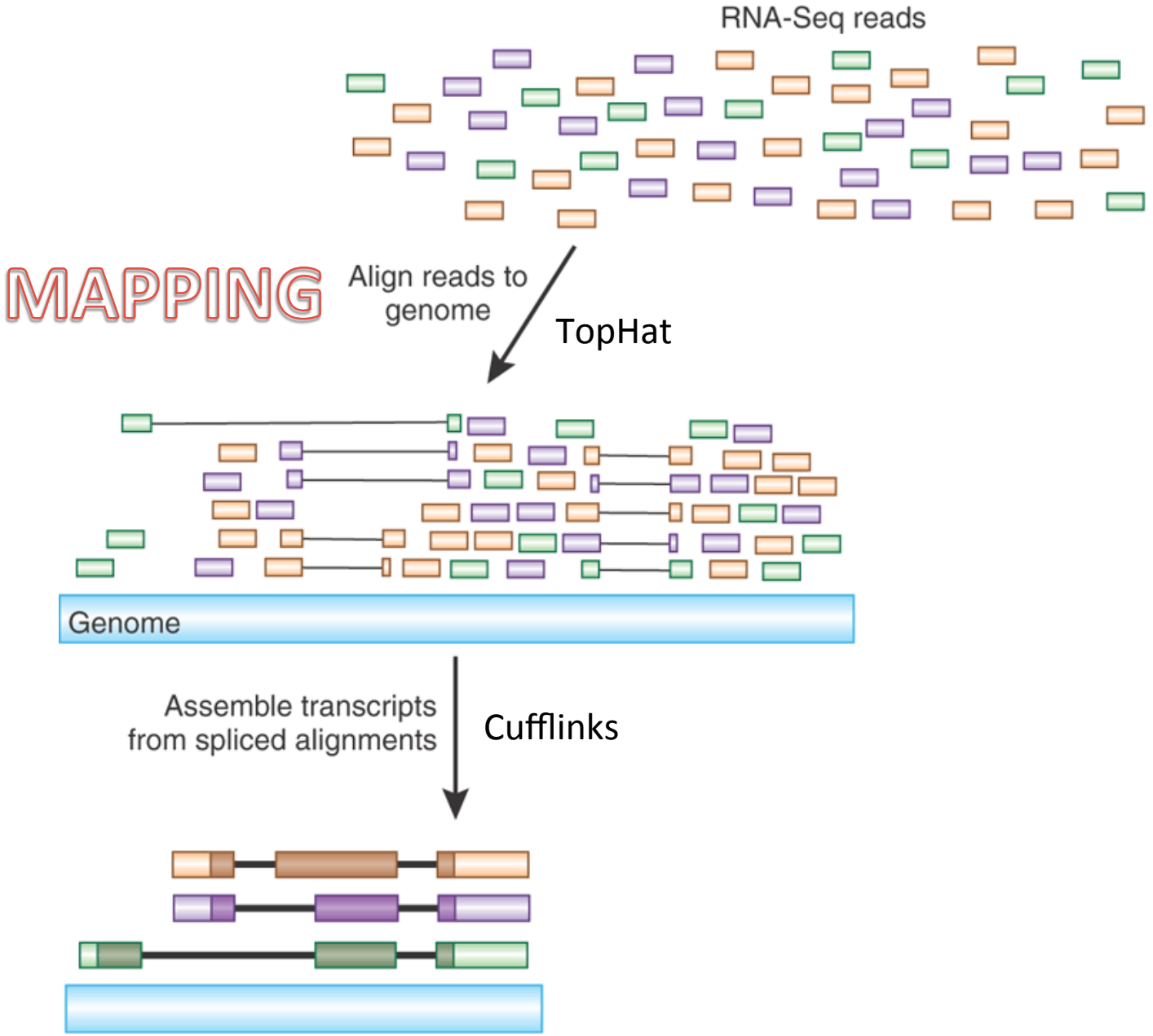
Nature Biotech, 2010

New methods for analyzing RNA-Seq data enable *de novo* reconstruction of the transcriptome.

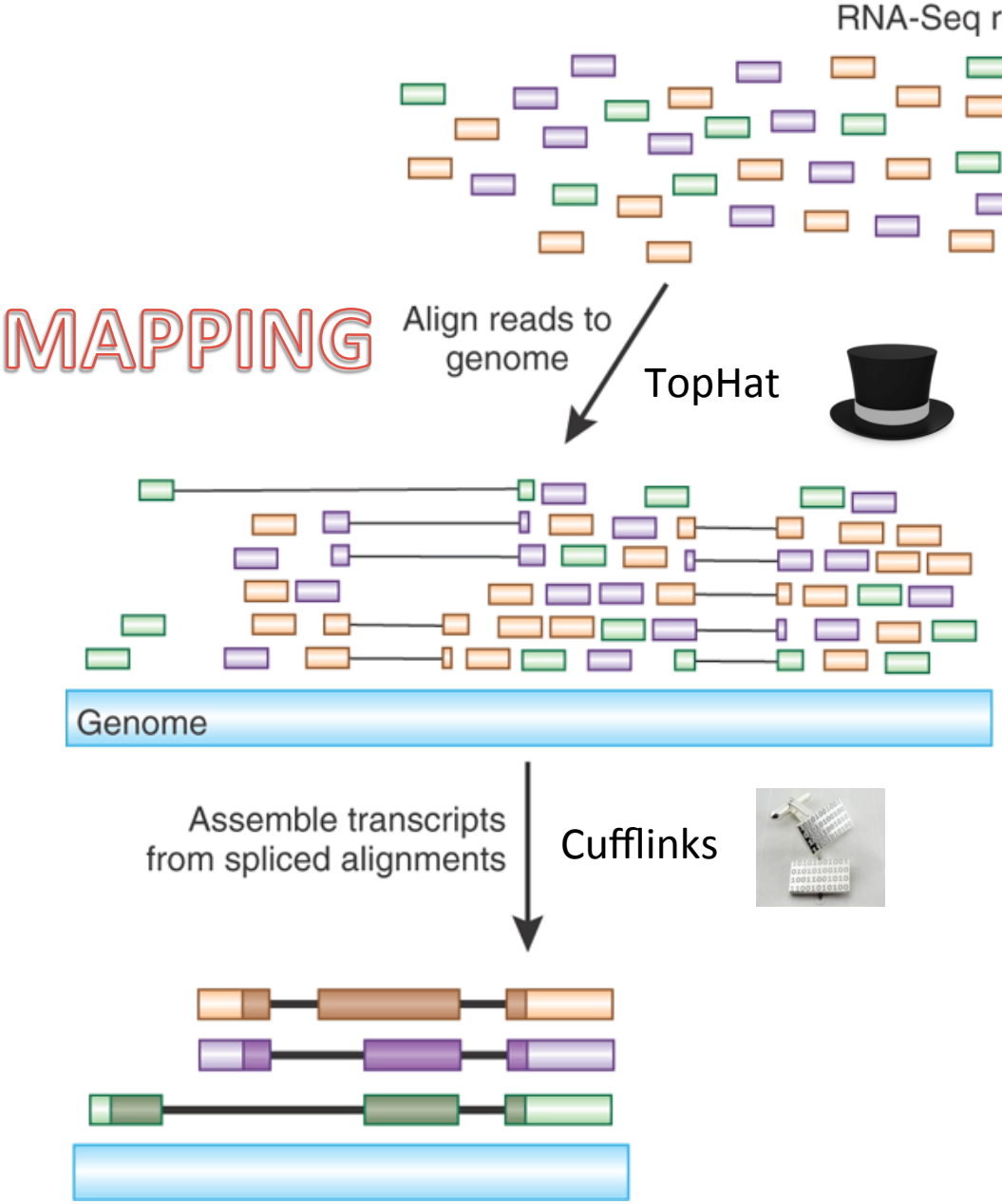
Transcript Reconstruction from RNA-Seq Reads



Transcript Reconstruction from RNA-Seq Reads



Transcript Reconstruction from RNA-Seq Reads



NATURE PROTOCOLS | PROTOCOL

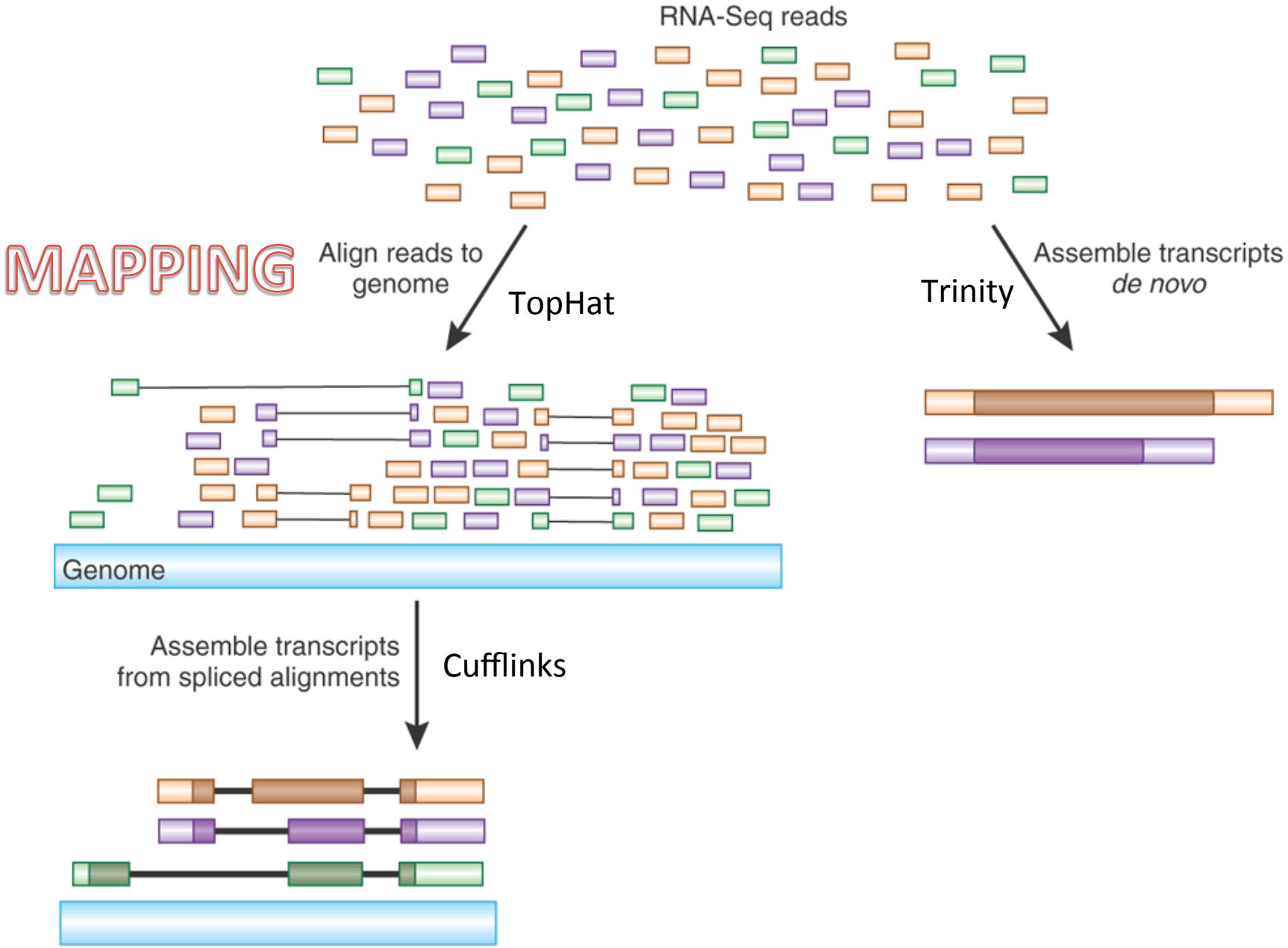
Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks

Cole Trapnell, Adam Roberts, Loyal Goff, Geo Pertea, Daehwan Kim, David R Kelley, Harold Pimentel, Steven L Salzberg, John L Rinn & Lior Pachter

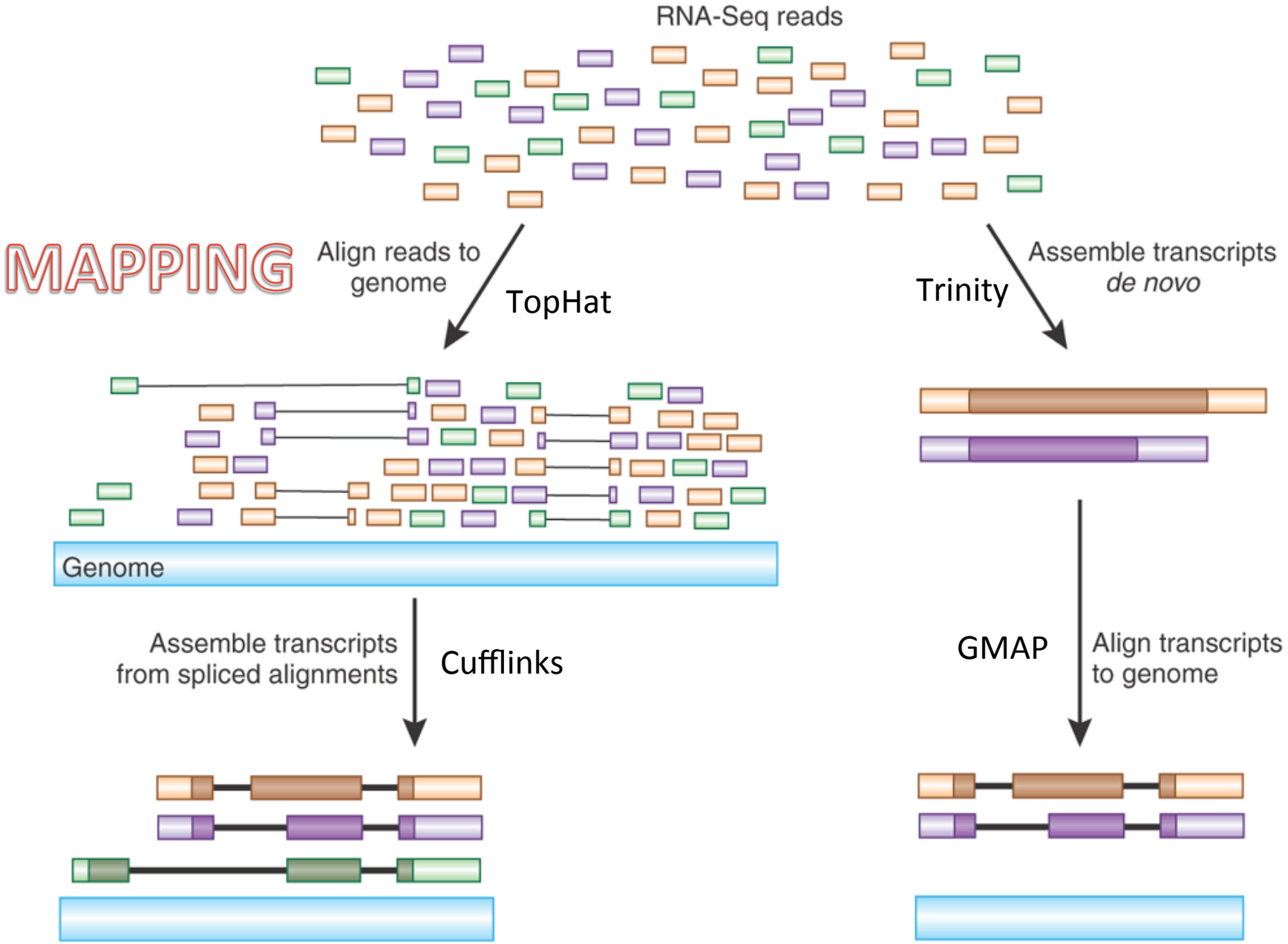
Affiliations | Contributions | Corresponding author

Nature Protocols **7**, 562–578 (2012) | doi:10.1038/nprot.2012.016
Published online 01 March 2012

Transcript Reconstruction from RNA-Seq Reads



Transcript Reconstruction from RNA-Seq Reads



Transcript Reconstruction from RNA-Seq Reads

End-to-end **Transcriptome**-based
RNA-Seq Analysis
Software Package

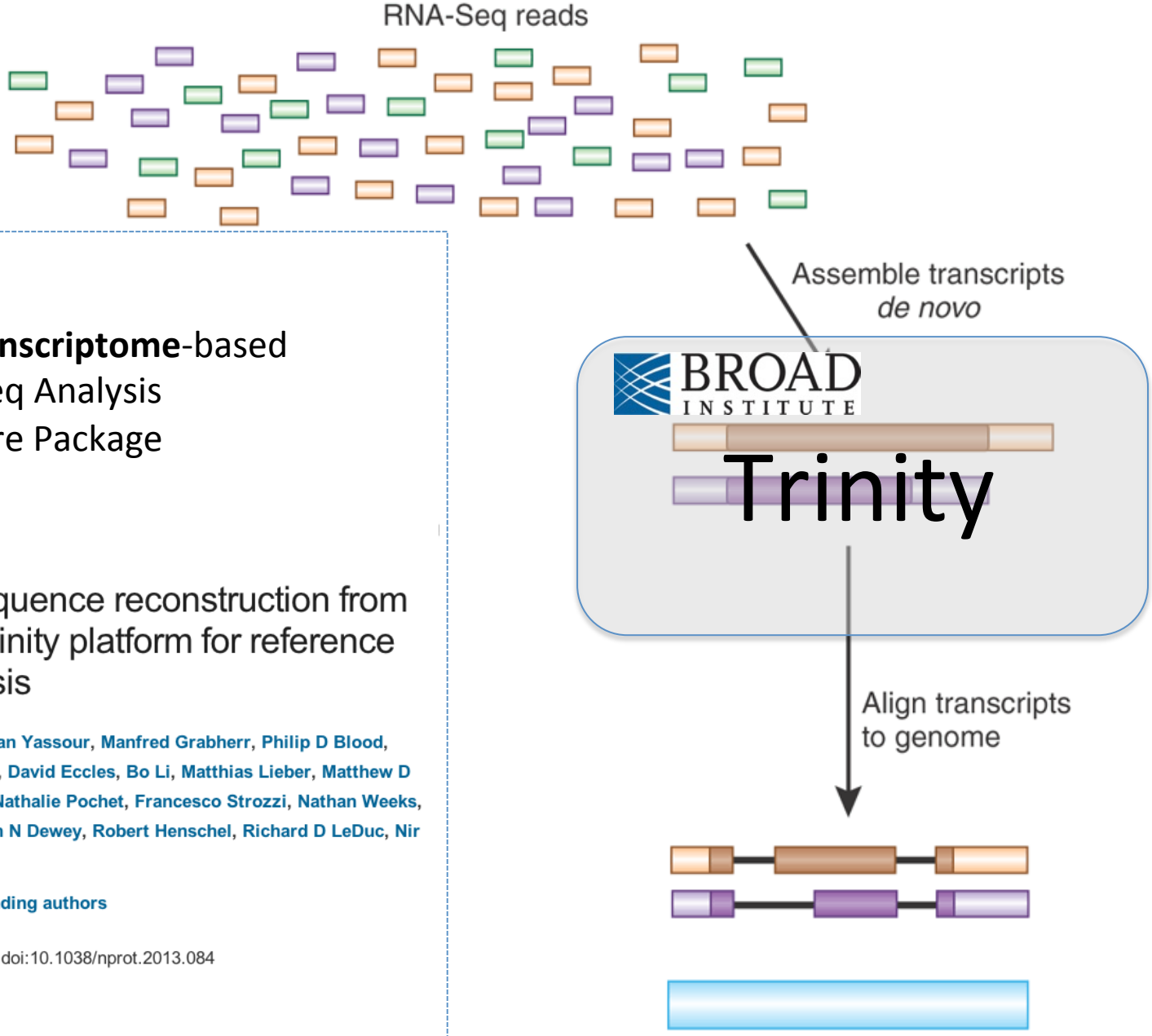
NATURE PROTOCOLS | PROTOCOL

De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis

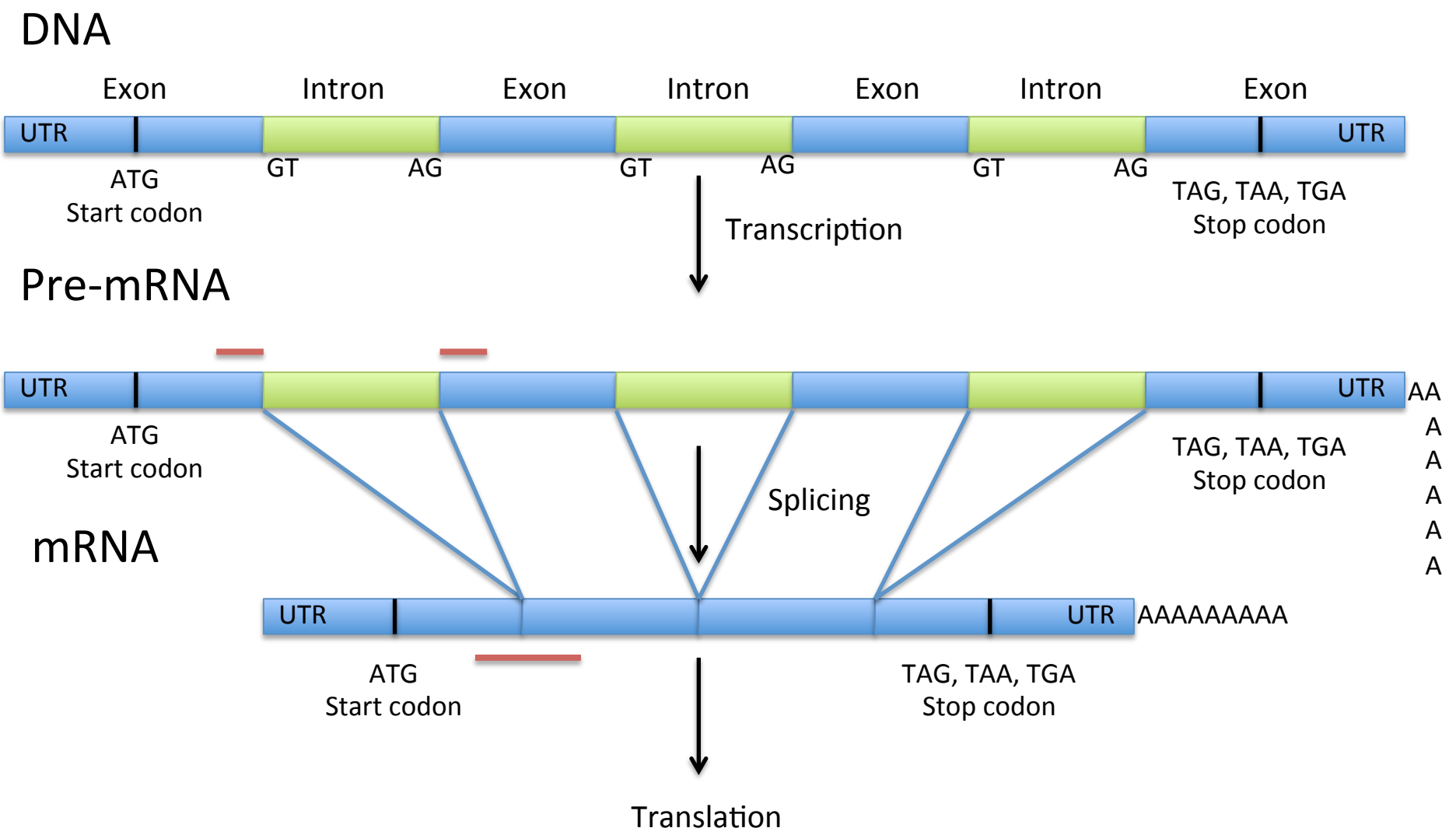
Brian J Haas, Alexie Papanicolaou, Moran Yassour, Manfred Grabherr, Philip D Blood, Joshua Bowden, Matthew Brian Couger, David Eccles, Bo Li, Matthias Lieber, Matthew D MacManes, Michael Ott, Joshua Orvis, Nathalie Pochet, Francesco Strozzi, Nathan Weeks, Rick Westerman, Thomas William, Colin N Dewey, Robert Henschel, Richard D LeDuc, Nir Friedman & Aviv Regev

Affiliations | Contributions | Corresponding authors

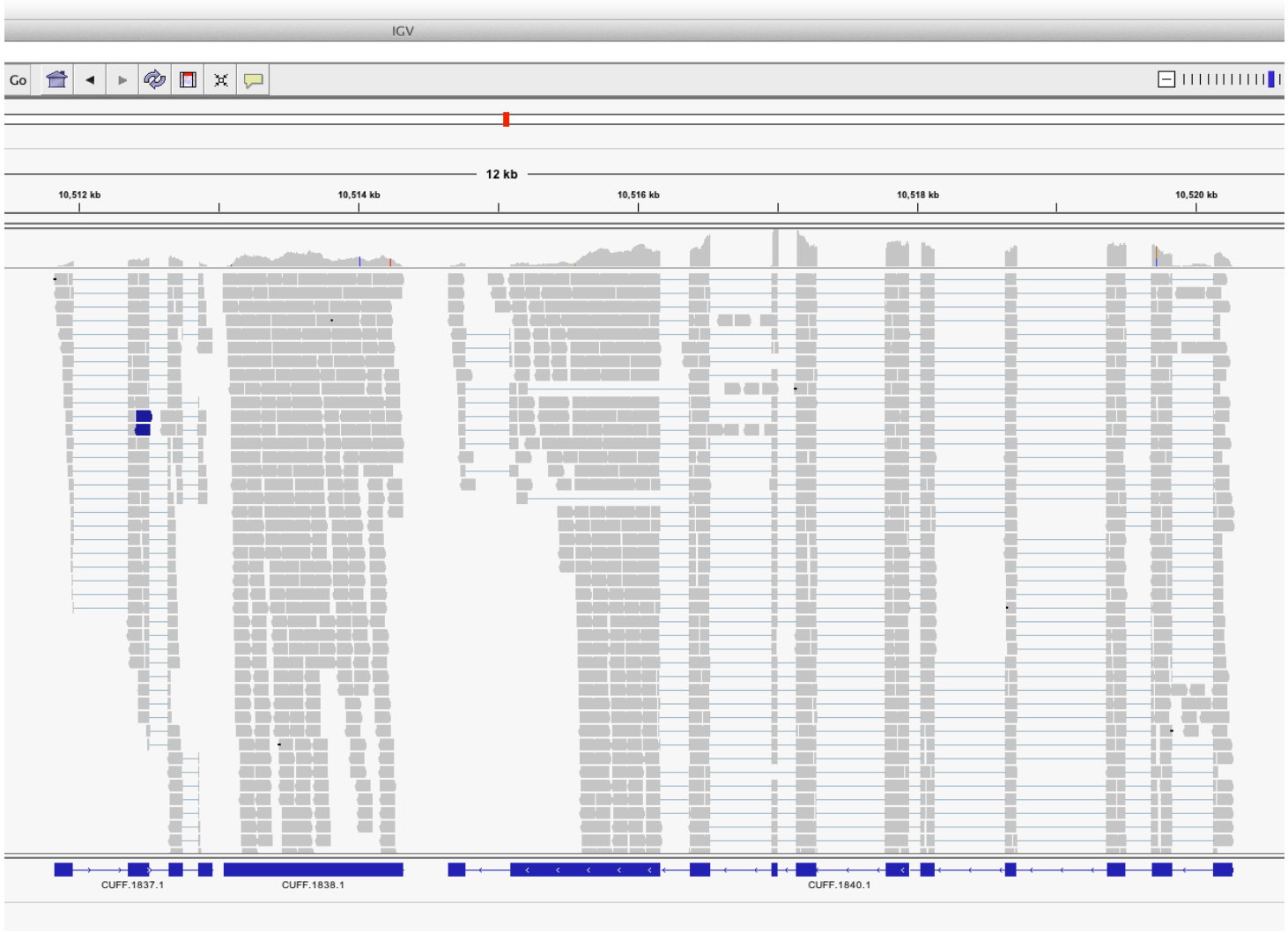
Nature Protocols **8**, 1494–1512 (2013) | doi:10.1038/nprot.2013.084
Published online 11 July 2013



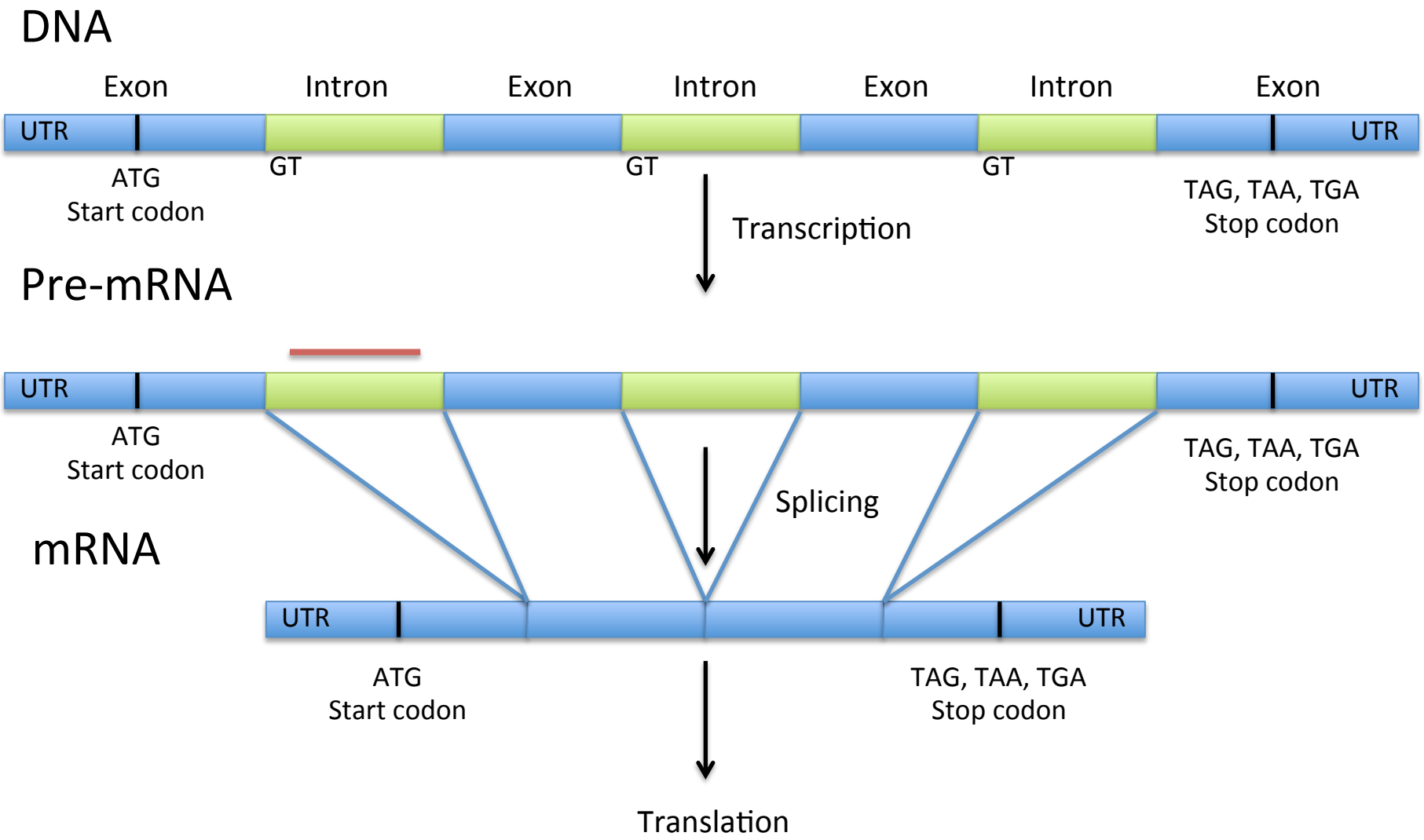
Basic concepts of mapping-based RNA-seq - Spliced reads



RNA-seq - Spliced reads

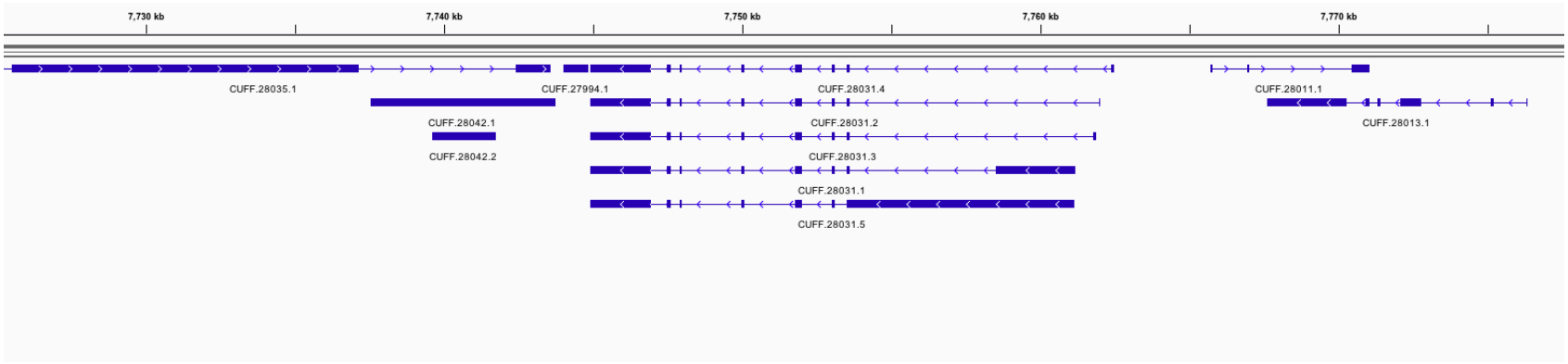


Pre-mRNA

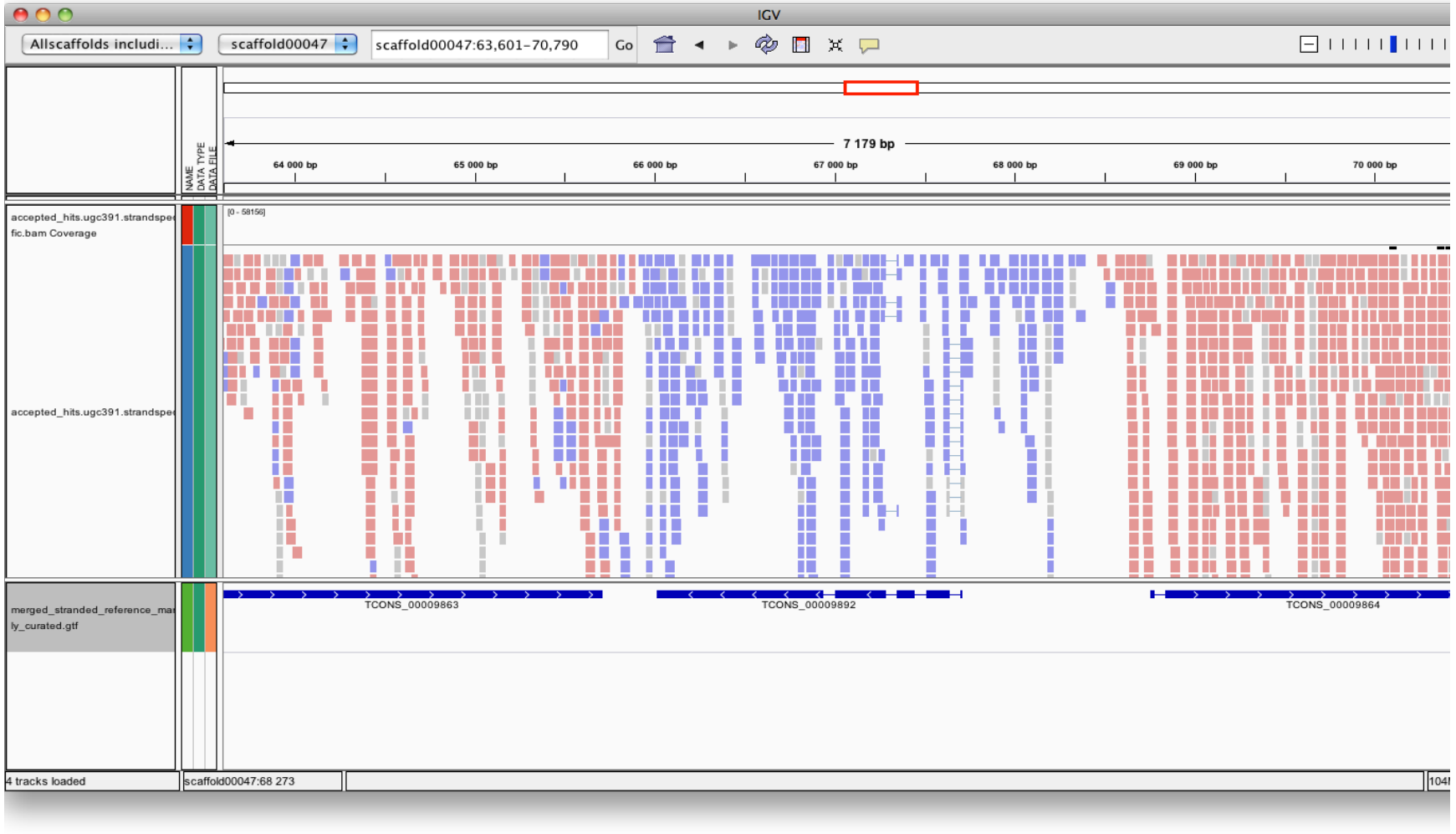

DNA
Exon
Intron
Exon
Intron
Exon
Intron
Exon
UTR
UTR
ATG
Start codon
GT
GT
GT
TAG, TAA, TGA
Stop codon
Transcription
Pre-mRNA
UTR
UTR
ATG
Start codon
TAG, TAA, TGA
Stop codon
Splicing
mRNA
UTR
UTR
ATG
Start codon
TAG, TAA, TGA
Stop codon
Translation

Pre-mRNA

Pre-mRNA


7,730 kb
7,740 kb
7,750 kb
7,760 kb
7,770 kb
CUFF.28035.1
CUFF.27994.1
CUFF.28031.4
CUFF.28011.1
CUFF.28042.1
CUFF.28031.2
CUFF.28013.1
CUFF.28042.2
CUFF.28031.3
CUFF.28031.1
CUFF.28031.5

Stranded rna-seq



Overview of RNASeq

Bowtie (fast short-read alignment)

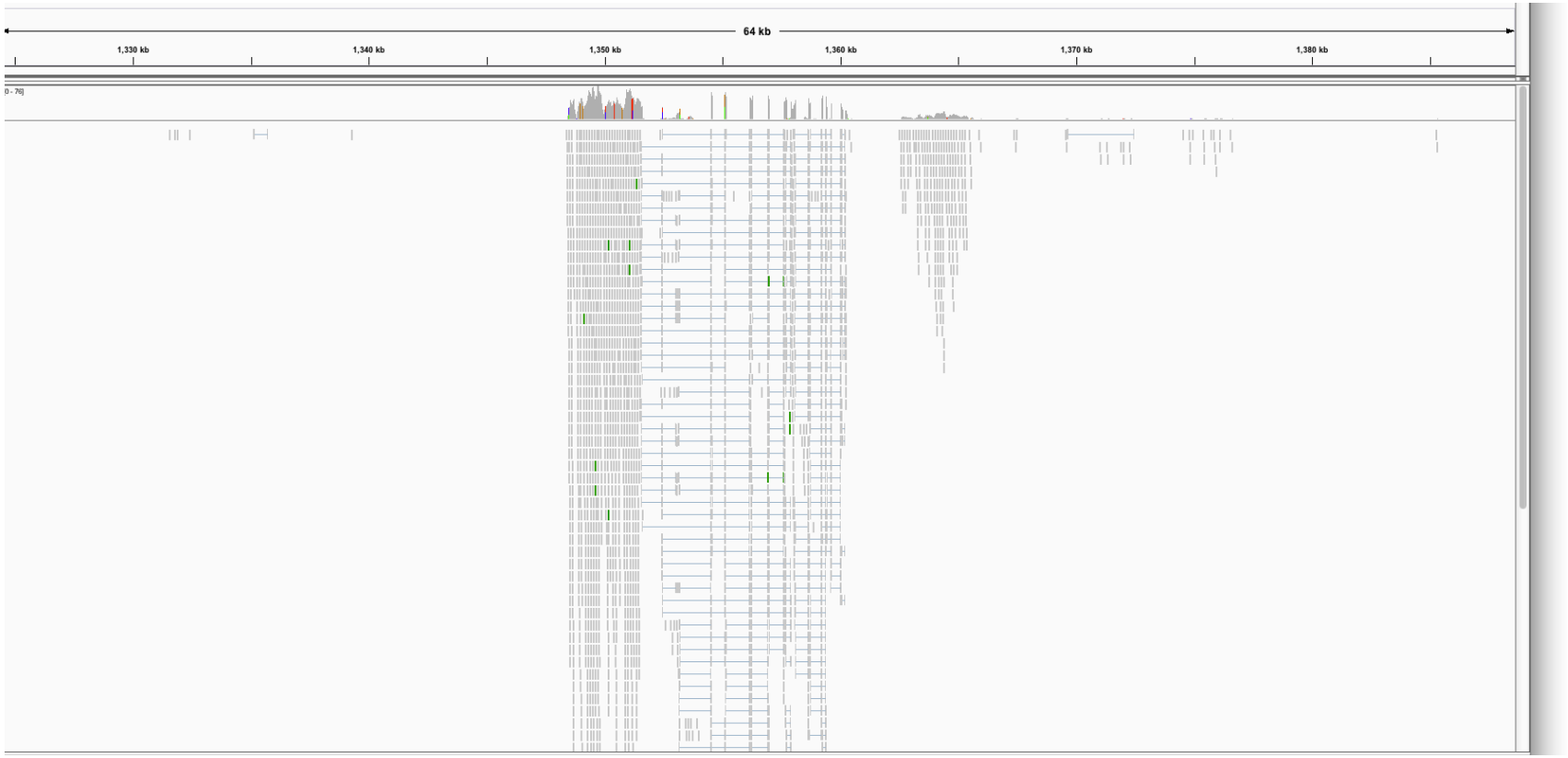
TopHat (spliced short-read alignment)

Cufflinks (transcript reconstruction from alignments)

Cuffdiff (differential expression analysis)

CummeRbund (visualization & analysis)

Tophat-mapped reads



Alignments are reported in a compact representation: SAM format

```
0       61G9EAAXX100520:5:100:10095:16477
1       83
2       chr1
3       51986
4       38
5       46M
6       =
7       51789
8       -264
9       CCCAAACAAGCCGAACTAGCTGATTTGGCTCGTAAAGACCCGGAAA
10      ###CB?=ADDBCBCDEEFFDEFFFDEFFGDBEFGEDGCFGFGGGGG
11      MD:Z:67
12      NH:i:1
13      HI:i:1
14      NM:i:0
15      SM:i:38
16      XQ:i:40
17      X2:i:0
```

SAM format specification: http://samtools.sourceforge.net/SAM1.pdf

Alignments are reported in a compact representation: SAM format

```
0     61G9EAAXX100520:5:100:10095:16477
1     83
2     chr1
3     51986
4     38
5     46M
6     =
7     51789
8     -264
9     CCCAAACAAGCCGAACTAGCTGATTTGGCTCGTAAAGACCCGGAAA
10    ###CB?=ADDBCBCDEEFFDEFFFDEFFGDBEFGEDGCFGFGGGGG
11    MD:Z:67
12    NH:i:1
13    HI:i:1
14    NM:i:0
15    SM:i:38
16    XQ:i:40
17    X2:i:0
```

(read name)

(FLAGS stored as bit fields; 83 = 00001010011)

(alignment target)

(position alignment starts)

(Compact description of the alignment in CIGAR format)

(read sequence, oriented according to the forward alignment)

(base quality values)

(Metadata)

SAM format specification: http://samtools.sourceforge.net/SAM1.pdf

Alignments are reported in a compact representation: SAM format

```
0       61G9EAAXX100520:5:100:10095:16477  (read name)
1       83   (FLAGS stored as bit fields; 83 = 00001010011 )
2       chr1   (alignment target)
```

Still not compact enough…
Millions to billions of reads takes up a lot of space!!

Convert SAM to binary – BAM format.

```
15      SM:i:38
16      XQ:i:40
17      X2:i:0
```

SAM format specification: http://samtools.sourceforge.net/SAM1.pdf

Samtools

- Tools for
  - converting SAM <-> BAM
  - Viewing BAM files (eg. samtools view file.bam | less)
  - Sorting BAM files, and lots more:

```
Program: samtools (Tools for alignments in the SAM format)
Version: 0.1.18 (r982:295)

Usage:   samtools <command> [options]

Command: view        SAM<->BAM conversion
         sort        sort alignment file
         mpileup     multi-way pileup
         depth       compute the depth
         faidx       index/extract FASTA
         tview       text alignment viewer
         index       index alignment
         idxstats    BAM index stats (r595 or later)
         fixmate     fix mate information
         flagstat    simple stats
         calmd       recalculate MD/NM tags and '=' bases
         merge       merge sorted alignments
         rmdup       remove PCR duplicates
         reheader    replace BAM header
         cat         concatenate BAMs
         targetcut   cut fosmid regions (for fosmid pool only)
         phase       phase heterozygotes
```

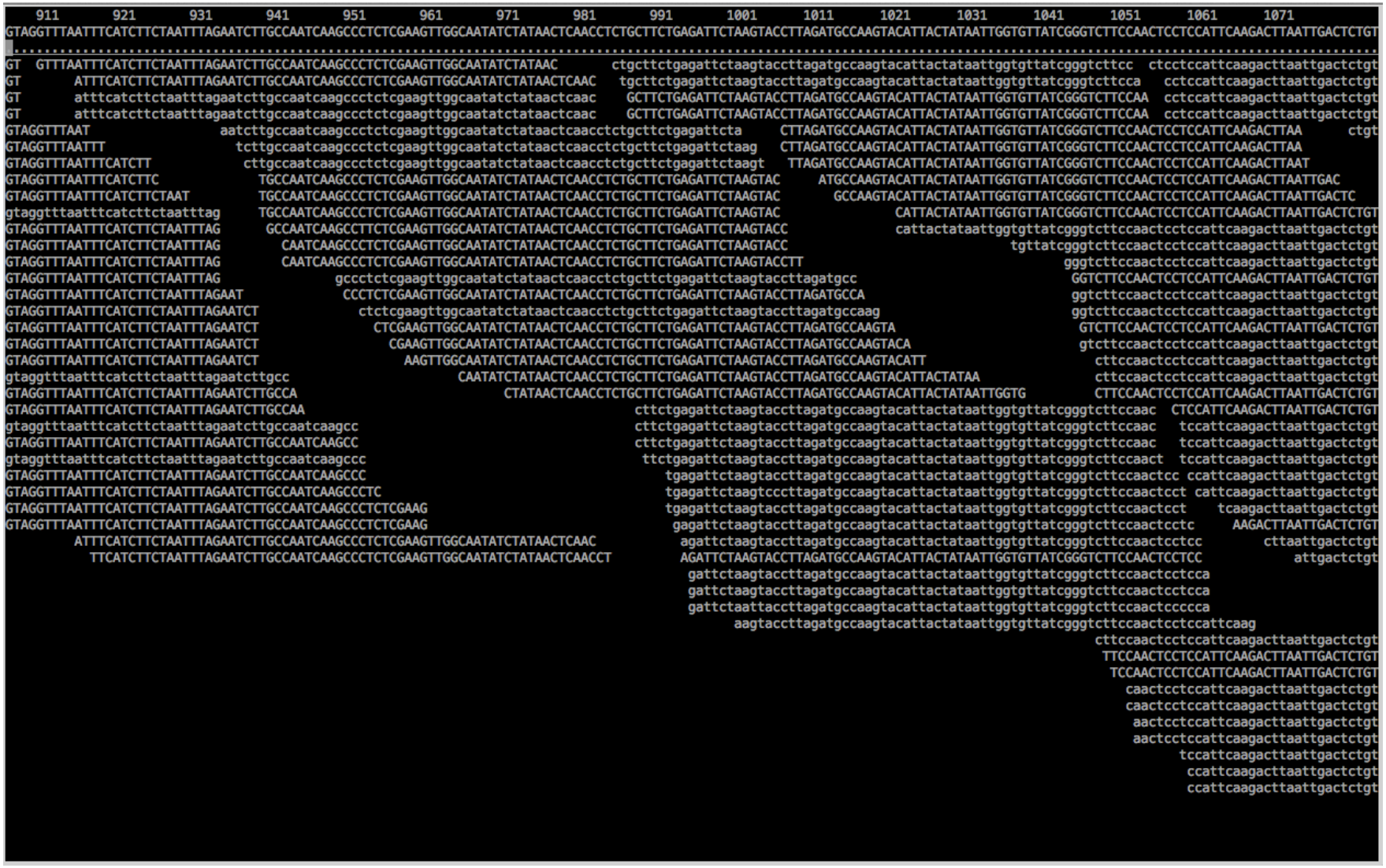
There is also CRAM…

CRAM compression rate File format	**File size (GB)**
SAM	7.4
BAM	1.9
CRAM lossless	1.4
CRAM 8 bins	0.8
CRAM no quality scores	0.26

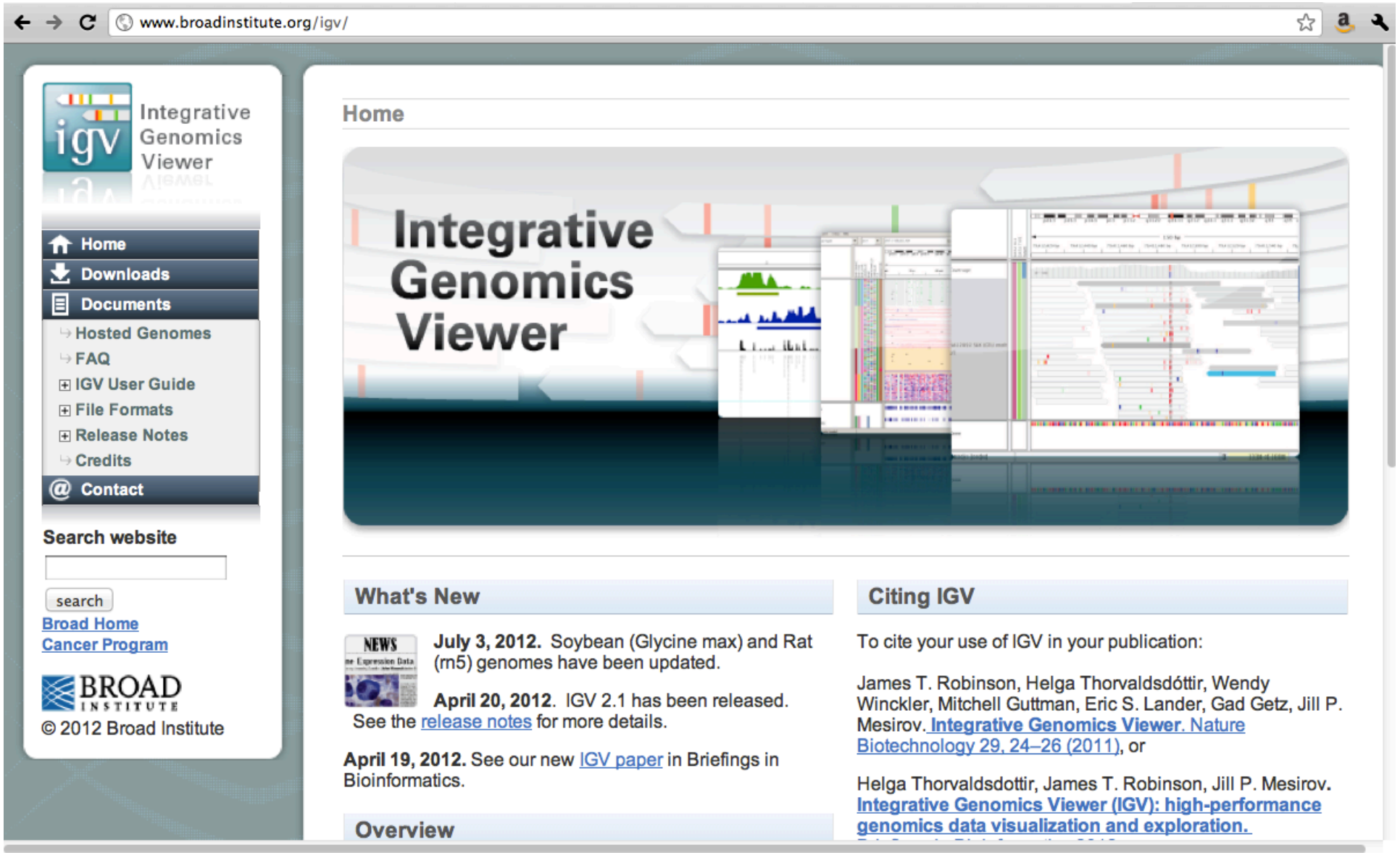
Visualizing Alignments of RNA-Seq reads

Text-based Alignment Viewer

% samtools tview alignments.bam target.fasta

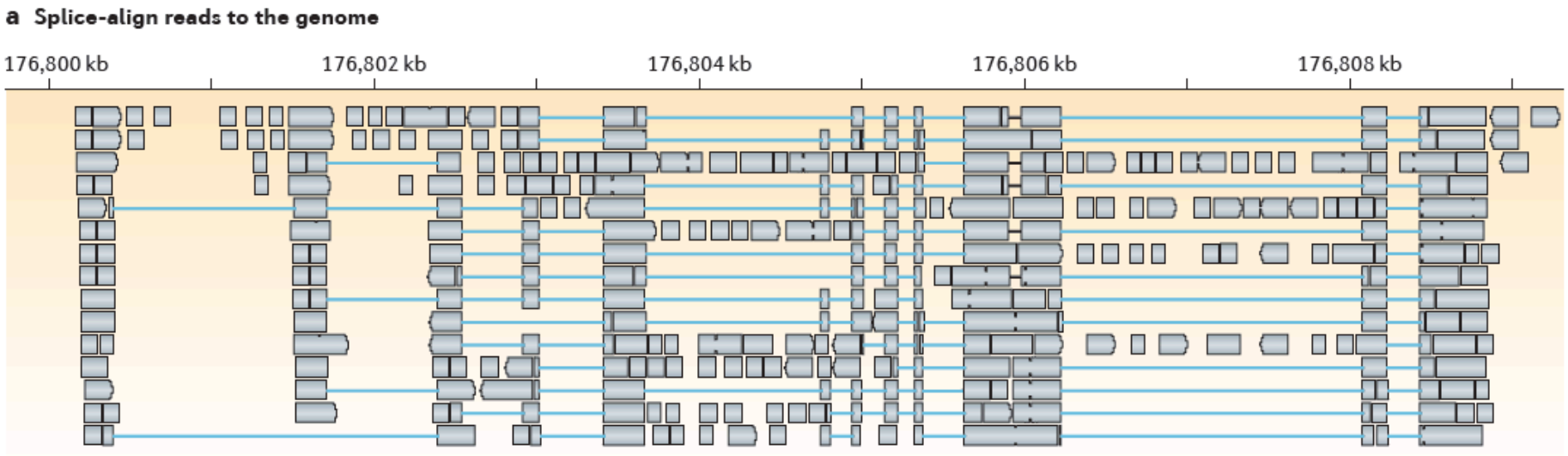


IGV


www.broadinstitute.org/igv/
Integrative Genomics Viewer
Home
Downloads
Documents
Hosted Genomes
FAQ
IGV User Guide
File Formats
Release Notes
Credits
Contact
Search website
search
Broad Home
Cancer Program
BROAD INSTITUTE
© 2012 Broad Institute
Home
Integrative Genomics Viewer
What's New
July 3, 2012. Soybean (Glycine max) and Rat (rn5) genomes have been updated.
April 20, 2012. IGV 2.1 has been released. See the release notes for more details.
April 19, 2012. See our new IGV paper in Briefings in Bioinformatics.
Overview
Citing IGV
To cite your use of IGV in your publication:
James T. Robinson, Helga Thorvaldsdóttir, Wendy Winckler, Mitchell Guttman, Eric S. Lander, Gad Getz, Jill P. Mesirov. Integrative Genomics Viewer. Nature Biotechnology 29, 24–26 (2011), or
Helga Thorvaldsdottir, James T. Robinson, Jill P. Mesirov. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration.

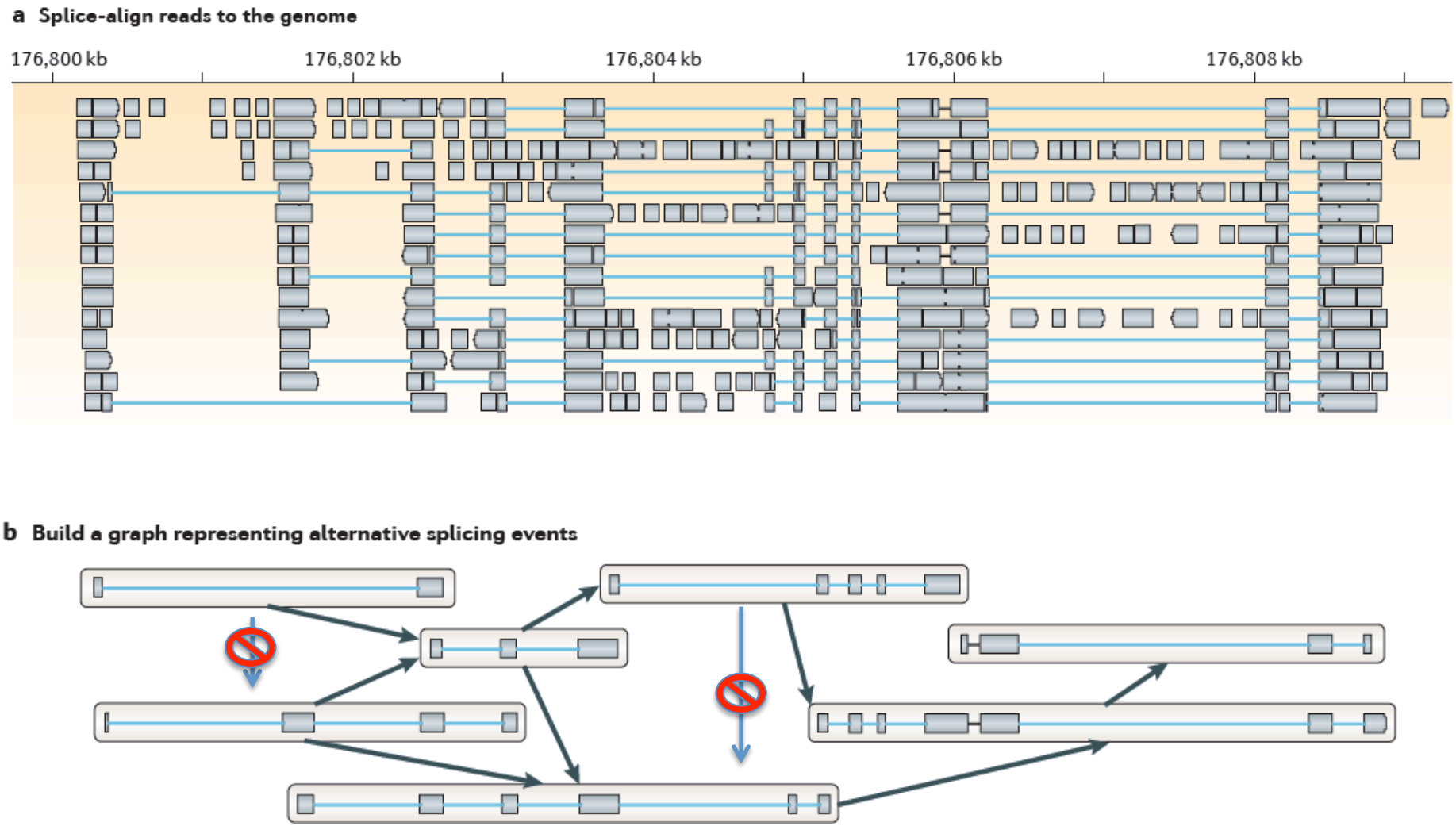
IGV: Viewing Tophat Alignments

Transcript Reconstruction Using Cufflinks



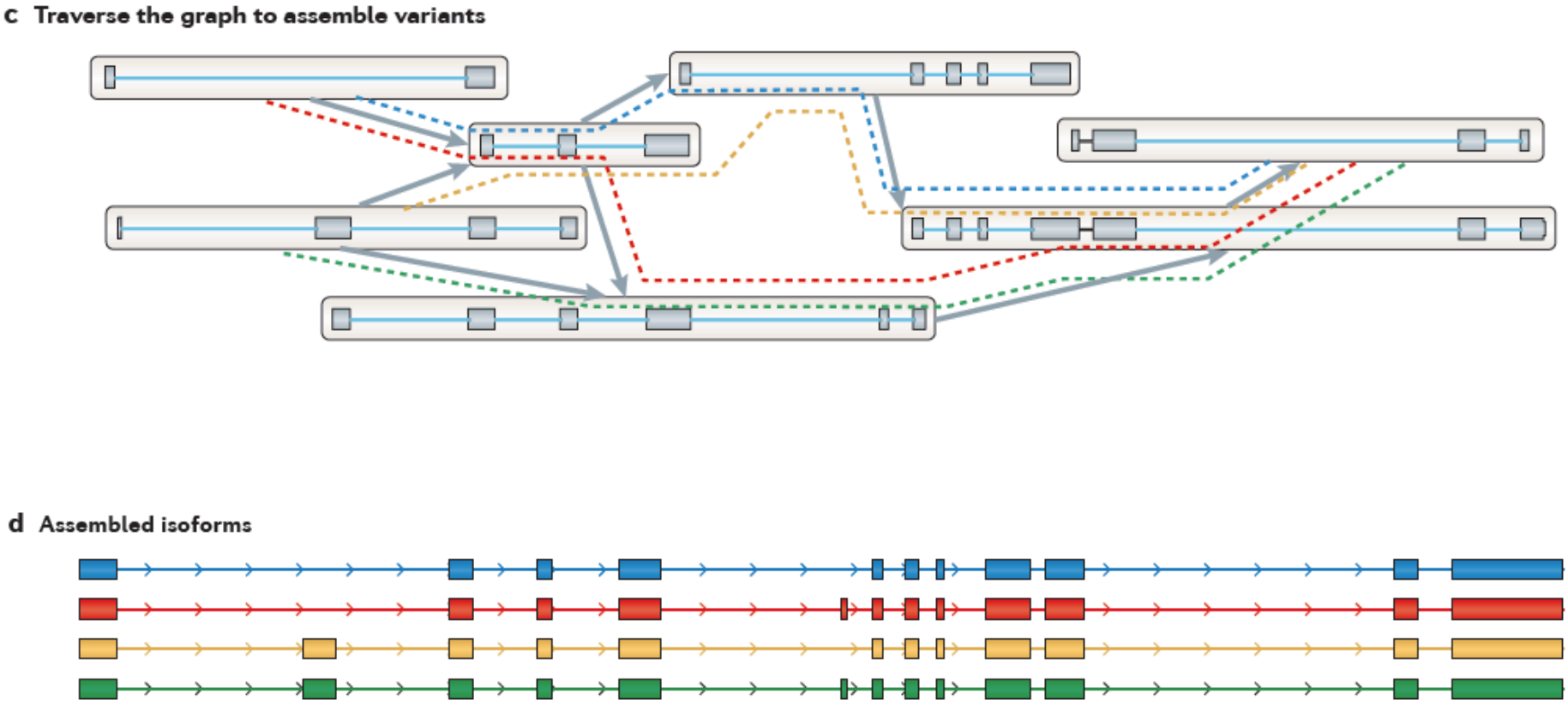
From Martin & Wang. Nature Reviews in Genetics. 2011

Transcript Reconstruction Using Cufflinks



From Martin & Wang. Nature Reviews in Genetics. 2011

Transcript Reconstruction Using Cufflinks



From Martin & Wang. Nature Reviews in Genetics. 2011

GFF file format

```
##gff-version 3
scaffold_7      maker   gene    133848  144662  .       -       .       ID=C55462A8A38E2878A71E2ABDAFB34661;Name=maker-scaffold_7-augustus-gene-0.11
scaffold_7      maker   mRNA    133848  144662  .       -       .       ID=A649E923246BADE2184E579FA9124ABD;Parent=C55462A8A38E2878A71E2ABDAFB34661;Name=1:cornix-all_reads.72406.1;_AED=1.00;_eAED=1.00;_QI=71|0|0|0|0|0|8|414|347
scaffold_7      maker   exon    138974  139077  .       -       .       ID=A649E923246BADE2184E579FA9124ABD:exon:7;Parent=A649E923246BADE2184E579FA9124ABD
scaffold_7      maker   exon    135098  135281  .       -       .       ID=A649E923246BADE2184E579FA9124ABD:exon:6;Parent=A649E923246BADE2184E579FA9124ABD
scaffold_7      maker   exon    139616  139836  .       -       .       ID=A649E923246BADE2184E579FA9124ABD:exon:5;Parent=A649E923246BADE2184E579FA9124ABD
scaffold_7      maker   exon    144511  144662  .       -       .       ID=A649E923246BADE2184E579FA9124ABD:exon:4;Parent=A649E923246BADE2184E579FA9124ABD
scaffold_7      maker   exon    136342  136437  .       -       .       ID=A649E923246BADE2184E579FA9124ABD:exon:3;Parent=A649E923246BADE2184E579FA9124ABD
scaffold_7      maker   exon    133848  134338  .       -       .       ID=A649E923246BADE2184E579FA9124ABD:exon:2;Parent=A649E923246BADE2184E579FA9124ABD
scaffold_7      maker   exon    141262  141383  .       -       .       ID=A649E923246BADE2184E579FA9124ABD:exon:1;Parent=A649E923246BADE2184E579FA9124ABD
scaffold_7      maker   exon    144138  144296  .       -       .       ID=A649E923246BADE2184E579FA9124ABD:exon:0;Parent=A649E923246BADE2184E579FA9124ABD
scaffold_7      maker   five_prime_UTR  144592  144662  .       -       .       ID=A649E923246BADE2184E579FA9124ABD:five_prime_utr;Parent=A649E923246BADE2184E579FA9124ABD
scaffold_7      maker   CDS     144511  144591  .       -       0       ID=A649E923246BADE2184E579FA9124ABD:cds;Parent=A649E923246BADE2184E579FA9124ABD
scaffold_7      maker   CDS     144138  144296  .       -       0       ID=A649E923246BADE2184E579FA9124ABD:cds;Parent=A649E923246BADE2184E579FA9124ABD
scaffold_7      maker   CDS     141262  141383  .       -       0       ID=A649E923246BADE2184E579FA9124ABD:cds;Parent=A649E923246BADE2184E579FA9124ABD
scaffold_7      maker   CDS     139616  139836  .       -       1       ID=A649E923246BADE2184E579FA9124ABD:cds;Parent=A649E923246BADE2184E579FA9124ABD
scaffold_7      maker   CDS     138974  139077  .       -       2       ID=A649E923246BADE2184E579FA9124ABD:cds;Parent=A649E923246BADE2184E579FA9124ABD
scaffold_7      maker   CDS     136342  136437  .       -       0       ID=A649E923246BADE2184E579FA9124ABD:cds;Parent=A649E923246BADE2184E579FA9124ABD
scaffold_7      maker   CDS     135098  135281  .       -       0       ID=A649E923246BADE2184E579FA9124ABD:cds;Parent=A649E923246BADE2184E579FA9124ABD
scaffold_7      maker   CDS     134262  134338  .       -       2       ID=A649E923246BADE2184E579FA9124ABD:cds;Parent=A649E923246BADE2184E579FA9124ABD
scaffold_7      maker   three_prime_UTR 133848  134261  .       -       .       ID=A649E923246BADE2184E579FA9124ABD:three_prime_utr;Parent=A649E923246BADE2184E579FA9124ABD
scaffold_7      maker   gene    83101   117593  .       +       .       ID=D3BD9A5F27797F56A844A2E890FF6B99;Name=maker-scaffold_7-augustus-gene-0.6
scaffold_7      maker   mRNA    83101   117593  .       +       .       ID=CFF5DDA190832937C45A0D2E674C9C26;Parent=D3BD9A5F27797F56A844A2E890FF6B99;Name=maker-scaffold_7-augustus-gene-0.6-mRNA-1;_AED=1.00;_eAED=1.00;_QI=0|0|0|0|0|0|21|4|706
scaffold_7      maker   exon    95748   95871   .       +       .       ID=CFF5DDA190832937C45A0D2E674C9C26:exon:8;Parent=CFF5DDA190832937C45A0D2E674C9C26
scaffold_7      maker   exon    99113   99137   .       +       .       ID=CFF5DDA190832937C45A0D2E674C9C26:exon:9;Parent=CFF5DDA190832937C45A0D2E674C9C26
scaffold_7      maker   exon    90664   90748   .       +       .       ID=CFF5DDA190832937C45A0D2E674C9C26:exon:10;Parent=CFF5DDA190832937C45A0D2E674C9C26
scaffold_7      maker   exon    110231  110356  .       +       .       ID=CFF5DDA190832937C45A0D2E674C9C26:exon:11;Parent=CFF5DDA190832937C45A0D2E674C9C26
scaffold_7      maker   exon    113609  113679  .       +       .       ID=CFF5DDA190832937C45A0D2E674C9C26:exon:12;Parent=CFF5DDA190832937C45A0D2E674C9C26
scaffold_7      maker   exon    94057   94117   .       +       .       ID=CFF5DDA190832937C45A0D2E674C9C26:exon:13;Parent=CFF5DDA190832937C45A0D2E674C9C26
scaffold_7      maker   exon    84578   84670   .       +       .       ID=CFF5DDA190832937C45A0D2E674C9C26:exon:14;Parent=CFF5DDA190832937C45A0D2E674C9C26
scaffold_7      maker   exon    115452  115536  .       +       .       ID=CFF5DDA190832937C45A0D2E674C9C26:exon:15;Parent=CFF5DDA190832937C45A0D2E674C9C26
scaffold_7      maker   exon    111579  111669  .       +       .       ID=CFF5DDA190832937C45A0D2E674C9C26:exon:16;Parent=CFF5DDA190832937C45A0D2E674C9C26
scaffold_7      maker   exon    102917  103016  .       +       .       ID=CFF5DDA190832937C45A0D2E674C9C26:exon:17;Parent=CFF5DDA190832937C45A0D2E674C9C26
scaffold_7      maker   exon    96766   96849   .       +       .       ID=CFF5DDA190832937C45A0D2E674C9C26:exon:18;Parent=CFF5DDA190832937C45A0D2E674C9C26
scaffold_7      maker   exon    86666   86750   .       +       .       ID=CFF5DDA190832937C45A0D2E674C9C26:exon:19;Parent=CFF5DDA190832937C45A0D2E674C9C26
scaffold_7      maker   exon    99944   100109  .       +       .       ID=CFF5DDA190832937C45A0D2E674C9C26:exon:20;Parent=CFF5DDA190832937C45A0D2E674C9C26
scaffold_7      maker   exon    109766  109860  .       +       .       ID=CFF5DDA190832937C45A0D2E674C9C26:exon:21;Parent=CFF5DDA190832937C45A0D2E674C9C26
scaffold_7      maker   exon    93154   93282   .       +       .       ID=CFF5DDA190832937C45A0D2E674C9C26:exon:22;Parent=CFF5DDA190832937C45A0D2E674C9C26
scaffold_7      maker   exon    114737  114825  .       +       .       ID=CFF5DDA190832937C45A0D2E674C9C26:exon:23;Parent=CFF5DDA190832937C45A0D2E674C9C26
scaffold_7      maker   exon    83101   83155   .       +       .       ID=CFF5DDA190832937C45A0D2E674C9C26:exon:24;Parent=CFF5DDA190832937C45A0D2E674C9C26
scaffold_7      maker   exon    108533  108795  .       +       .       ID=CFF5DDA190832937C45A0D2E674C9C26:exon:25;Parent=CFF5DDA190832937C45A0D2E674C9C26
scaffold_7      maker   exon    117477  117593  .       +       .       ID=CFF5DDA190832937C45A0D2E674C9C26:exon:26;Parent=CFF5DDA190832937C45A0D2E674C9C26
scaffold_7      maker   exon    106779  106866  .       +       .       ID=CFF5DDA190832937C45A0D2E674C9C26:exon:27;Parent=CFF5DDA190832937C45A0D2E674C9C26
scaffold_7      maker   exon    105743  105835  .       +       .       ID=CFF5DDA190832937C45A0D2E674C9C26:exon:28;Parent=CFF5DDA190832937C45A0D2E674C9C26
scaffold_7      maker   CDS     83101   83155   .       +       0       ID=CFF5DDA190832937C45A0D2E674C9C26:cds;Parent=CFF5DDA190832937C45A0D2E674C9C26
scaffold_7      maker   CDS     84578   84670   .       +       2       ID=CFF5DDA190832937C45A0D2E674C9C26:cds;Parent=CFF5DDA190832937C45A0D2E674C9C26
scaffold_7      maker   CDS     86666   86750   .       +       2       ID=CFF5DDA190832937C45A0D2E674C9C26:cds;Parent=CFF5DDA190832937C45A0D2E674C9C26
scaffold_7      maker   CDS     90664   90748   .       +       1       ID=CFF5DDA190832937C45A0D2E674C9C26:cds;Parent=CFF5DDA190832937C45A0D2E674C9C26
scaffold_7      maker   CDS     93154   93282   .       +       0       ID=CFF5DDA190832937C45A0D2E674C9C26:cds;Parent=CFF5DDA190832937C45A0D2E674C9C26
scaffold_7      maker   CDS     94057   94117   .       +       0       ID=CFF5DDA190832937C45A0D2E674C9C26:cds;Parent=CFF5DDA190832937C45A0D2E674C9C26
scaffold_7      maker   CDS     95748   95871   .       +       2       ID=CFF5DDA190832937C45A0D2E674C9C26:cds;Parent=CFF5DDA190832937C45A0D2E674C9C26
scaffold_7      maker   CDS     96766   96849   .       +       1       ID=CFF5DDA190832937C45A0D2E674C9C26:cds;Parent=CFF5DDA190832937C45A0D2E674C9C26
scaffold_7      maker   CDS     99113   99137   .       +       1       ID=CFF5DDA190832937C45A0D2E674C9C26:cds;Parent=CFF5DDA190832937C45A0D2E674C9C26
scaffold_7      maker   CDS     99944   100109  .       +       0       ID=CFF5DDA190832937C45A0D2E674C9C26:cds;Parent=CFF5DDA190832937C45A0D2E674C9C26
scaffold_7      maker   CDS     102917  103016  .       +       2       ID=CFF5DDA190832937C45A0D2E674C9C26:cds;Parent=CFF5DDA190832937C45A0D2E674C9C26
scaffold_7      maker   CDS     105743  105835  .       +       1       ID=CFF5DDA190832937C45A0D2E674C9C26:cds;Parent=CFF5DDA190832937C45A0D2E674C9C26
scaffold_7      maker   CDS     106779  106866  .       +       1       ID=CFF5DDA190832937C45A0D2E674C9C26:cds;Parent=CFF5DDA190832937C45A0D2E674C9C26
scaffold_7      maker   CDS     108533  108795  .       +       0       ID=CFF5DDA190832937C45A0D2E674C9C26:cds;Parent=CFF5DDA190832937C45A0D2E674C9C26
scaffold_7      maker   CDS     109766  109860  .       +       1       ID=CFF5DDA190832937C45A0D2E674C9C26:cds;Parent=CFF5DDA190832937C45A0D2E674C9C26
scaffold_7      maker   CDS     110231  110356  .       +       2       ID=CFF5DDA190832937C45A0D2E674C9C26:cds;Parent=CFF5DDA190832937C45A0D2E674C9C26
scaffold_7      maker   CDS     111579  111669  .       +       2       ID=CFF5DDA190832937C45A0D2E674C9C26:cds;Parent=CFF5DDA190832937C45A0D2E674C9C26
scaffold_7      maker   CDS     113609  113679  .       +       1       ID=CFF5DDA190832937C45A0D2E674C9C26:cds;Parent=CFF5DDA190832937C45A0D2E674C9C26
scaffold_7      maker   CDS     114737  114825  .       +       2       ID=CFF5DDA190832937C45A0D2E674C9C26:cds;Parent=CFF5DDA190832937C45A0D2E674C9C26
scaffold_7      maker   CDS     115452  115536  .       +       0       ID=CFF5DDA190832937C45A0D2E674C9C26:cds;Parent=CFF5DDA190832937C45A0D2E674C9C26
scaffold_7      maker   CDS     117477  117589  .       +       2       ID=CFF5DDA190832937C45A0D2E674C9C26:cds;Parent=CFF5DDA190832937C45A0D2E674C9C26
scaffold_7      maker   three_prime_UTR 117590  117593  .       +       .       ID=CFF5DDA190832937C45A0D2E674C9C26:three_prime_utr;Parent=CFF5DDA190832937C45A0D2E674C9C26
scaffold_7      maker   gene    22451   37514   .       -       .       ID=E1E94A56071E5D7940C93BE70FD79E56;Name=E1E94A56071E5D7940C93BE70FD79E56
scaffold_7      maker   mRNA    22451   37514   .       -       .       ID=2BB3B792BBFCD8B1C7A4070E76500CC5;Parent=E1E94A56071E5D7940C93BE70FD79E56;Name=scaffold_7:22450-37514;_AED=1.00;_eAED=1.00;_QI=3008|0|0|0|0|0|11|762|414
scaffold_7      maker   mRNA    26810   37514   .       -       .       ID=1D629B10725FB97175585BA23623AFA6;Parent=E1E94A56071E5D7940C93BE70FD79E56;Name=maker-scaffold_7-augustus-gene-0.10-mRNA-1;_AED=1.00;_eAED=1.00;_QI=3008|0|0|0|0|0|10|0|400
scaffold_7      maker   exon    26810   27088   .       -       .       ID=1D629B10725FB97175585BA23623AFA6:exon:40;Parent=1D629B10725FB97175585BA23623AFA6
scaffold_7      maker   exon    27950   28090   .       -       .       ID=2BB3B792BBFCD8B1C7A4070E76500CC5:exon:39;Parent=2BB3B792BBFCD8B1C7A4070E76500CC5,1D629B10725FB97175585BA23623AFA6
scaffold_7      maker   exon    33513   33610   .       -       .       ID=2BB3B792BBFCD8B1C7A4070E76500CC5:exon:38;Parent=2BB3B792BBFCD8B1C7A4070E76500CC5,1D629B10725FB97175585BA23623AFA6
maker.gff
```

GFF3 file format

Seqid	source	type	start	end	score	strand	phase	attributes
Chr1	Snap	gene	234	3657	.	+	.	ID=gene1; Name=Snap1;
Chr1	Snap	mRNA	234	3657	.	+	.	ID=gene1.m1; Parent=gene1;
Chr1	Snap	exon	234	1543	.	+	.	ID=gene1.m1.exon1; Parent=gene1.m1;
Chr1	Snap	CDS	577	1543	.	+	0	ID=gene1.m1.CDS1; Parent=gene1.m1;
Chr1	Snap	exon	1822	2674	.	+	.	ID=gene1.m1.exon2; Parent=gene1.m1;
Chr1	Snap	CDS	1822	2674	.	+	2	ID=gene1.m1.CDS2; Parent=gene1.m1;
		start_ codon						Alias, note, ontology_term …
		stop_ codon						

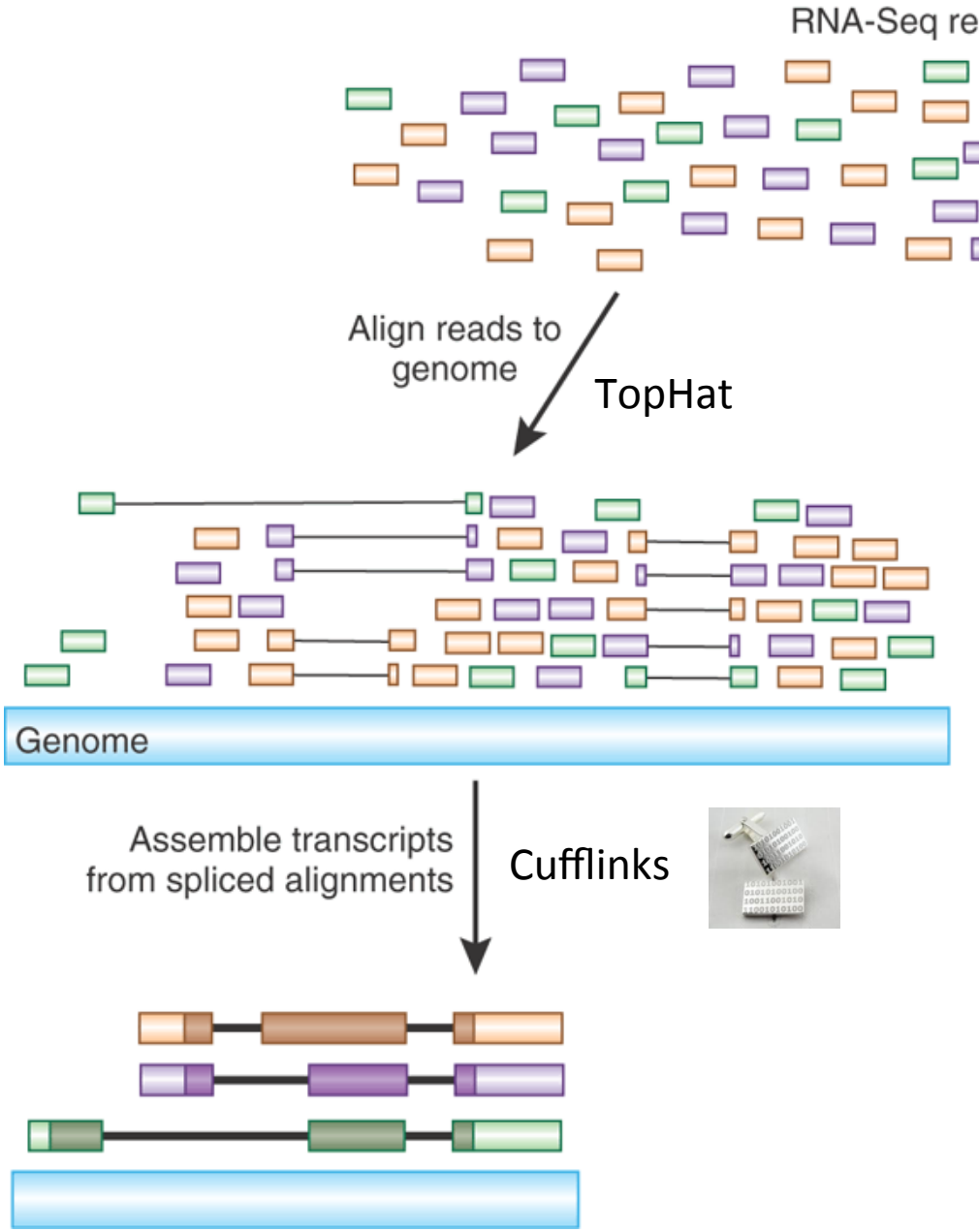
GTF file format

```
Sb_20131119_contig_1    Cufflinks       transcript      1522    2095    1000    .       .       gene_id "CUFF.1"; transcript_id "CUFF.1.1"; FPKM "2.6064385494"; frac "1.000000"; conf_lo "0.948975"; conf_hi "3.440036"; cov "4.817376";
Sb_20131119_contig_1    Cufflinks       exon    1522    2095    1000    .       .       gene_id "CUFF.1"; transcript_id "CUFF.1.1"; exon_number "1"; FPKM "2.6064385494"; frac "1.000000"; conf_lo "0.948975"; conf_hi "3.440036"; cov "4.817376";
Sb_20131119_contig_1    Cufflinks       transcript      3626    4118    1000    .       .       gene_id "CUFF.3"; transcript_id "CUFF.3.1"; FPKM "3.1548106029"; frac "1.000000"; conf_lo "0.828669"; conf_hi "3.729011"; cov "8.517668";
Sb_20131119_contig_1    Cufflinks       exon    3626    4118    1000    .       .       gene_id "CUFF.3"; transcript_id "CUFF.3.1"; exon_number "1"; FPKM "3.1548106029"; frac "1.000000"; conf_lo "0.828669"; conf_hi "3.729011"; cov "8.517668";
Sb_20131119_contig_1    Cufflinks       transcript      4855    5340    1000    .       .       gene_id "CUFF.5"; transcript_id "CUFF.5.1"; FPKM "7.0235237898"; frac "1.000000"; conf_lo "2.521814"; conf_hi "6.164435"; cov "16.171080";
Sb_20131119_contig_1    Cufflinks       exon    4855    5340    1000    .       .       gene_id "CUFF.5"; transcript_id "CUFF.5.1"; exon_number "1"; FPKM "7.0235237898"; frac "1.000000"; conf_lo "2.521814"; conf_hi "6.164435"; cov "16.171080";
Sb_20131119_contig_1    Cufflinks       transcript      5398    5975    1000    .       .       gene_id "CUFF.4"; transcript_id "CUFF.4.1"; FPKM "3.1706609980"; frac "1.000000"; conf_lo "1.178010"; conf_hi "3.651831"; cov "4.764328";
Sb_20131119_contig_1    Cufflinks       exon    5398    5975    1000    .       .       gene_id "CUFF.4"; transcript_id "CUFF.4.1"; exon_number "1"; FPKM "3.1706609980"; frac "1.000000"; conf_lo "1.178010"; conf_hi "3.651831"; cov "4.764328";
Sb_20131119_contig_10   Cufflinks       transcript      954     2795    1000    .       .       gene_id "CUFF.6"; transcript_id "CUFF.6.1"; FPKM "6.8195889357"; frac "1.000000"; conf_lo "5.175059"; conf_hi "7.577765"; cov "16.605534";
Sb_20131119_contig_10   Cufflinks       exon    954     2795    1000    .       .       gene_id "CUFF.6"; transcript_id "CUFF.6.1"; exon_number "1"; FPKM "6.8195889357"; frac "1.000000"; conf_lo "5.175059"; conf_hi "7.577765"; cov "16.605534";
Sb_20131119_contig_1    Cufflinks       transcript      4502    4718    1000    .       .       gene_id "CUFF.2"; transcript_id "CUFF.2.1"; FPKM "37.5296486924"; frac "1.000000"; conf_lo "2.510193"; conf_hi "9.099450"; cov "60.418389";
Sb_20131119_contig_1    Cufflinks       exon    4502    4718    1000    .       .       gene_id "CUFF.2"; transcript_id "CUFF.2.1"; exon_number "1"; FPKM "37.5296486924"; frac "1.000000"; conf_lo "2.510193"; conf_hi "9.099450"; cov "60.418389";
Sb_20131119_contig_1    Cufflinks       transcript      10522   13208   1000    .       .       gene_id "CUFF.23"; transcript_id "CUFF.23.1"; FPKM "55.6377793473"; frac "1.000000"; conf_lo "48.931832"; conf_hi "55.241530"; cov "121.429110";
Sb_20131119_contig_1    Cufflinks       exon    10522   13208   1000    .       .       gene_id "CUFF.23"; transcript_id "CUFF.23.1"; exon_number "1"; FPKM "55.6377793473"; frac "1.000000"; conf_lo "48.931832"; conf_hi "55.241530"; cov "121.429110";
Sb_20131119_contig_1    Cufflinks       transcript      13270   14623   1000    .       .       gene_id "CUFF.7"; transcript_id "CUFF.7.1"; FPKM "41.2374406123"; frac "1.000000"; conf_lo "31.982715"; conf_hi "39.274371"; cov "89.788421";
Sb_20131119_contig_1    Cufflinks       exon    13270   14623   1000    .       .       gene_id "CUFF.7"; transcript_id "CUFF.7.1"; exon_number "1"; FPKM "41.2374406123"; frac "1.000000"; conf_lo "31.982715"; conf_hi "39.274371"; cov "89.788421";
Sb_20131119_contig_100022       Cufflinks       transcript      3991    4547    1000    .       .       gene_id "CUFF.54"; transcript_id "CUFF.54.1"; FPKM "52.7578565123"; frac "1.000000"; conf_lo "27.382285"; conf_hi "37.895127"; cov "66.397320";
Sb_20131119_contig_100022       Cufflinks       exon    3991    4547    1000    .       .       gene_id "CUFF.54"; transcript_id "CUFF.54.1"; exon_number "1"; FPKM "52.7578565123"; frac "1.000000"; conf_lo "27.382285"; conf_hi "37.895127"; cov "66.397320";
Sb_20131119_contig_100023       Cufflinks       transcript      1097    2009    1000    .       .       gene_id "CUFF.9"; transcript_id "CUFF.9.1"; FPKM "7.4263254644"; frac "1.000000"; conf_lo "4.474632"; conf_hi "7.830606"; cov "16.282075";
Sb_20131119_contig_100023       Cufflinks       exon    1097    2009    1000    .       .       gene_id "CUFF.9"; transcript_id "CUFF.9.1"; exon_number "1"; FPKM "7.4263254644"; frac "1.000000"; conf_lo "4.474632"; conf_hi "7.830606"; cov "16.282075";
Sb_20131119_contig_100023       Cufflinks       transcript      1       123     1000    .       .       gene_id "CUFF.8"; transcript_id "CUFF.8.1"; FPKM "6279.663115693"; frac "1.000000"; conf_lo "42.624813"; conf_hi "69.749694"; cov "8277.399810";
Sb_20131119_contig_100023       Cufflinks       exon    1       123     1000    .       .       gene_id "CUFF.8"; transcript_id "CUFF.8.1"; exon_number "1"; FPKM "6279.663115693"; frac "1.000000"; conf_lo "42.624813"; conf_hi "69.749694"; cov "8277.399810";
Sb_20131119_contig_100040       Cufflinks       transcript      1       221     1000    .       .       gene_id "CUFF.10"; transcript_id "CUFF.10.1"; FPKM "233.0138642752"; frac "1.000000"; conf_lo "22.182837"; conf_hi "38.203775"; cov "306.188212";
Sb_20131119_contig_100040       Cufflinks       exon    1       221     1000    .       .       gene_id "CUFF.10"; transcript_id "CUFF.10.1"; exon_number "1"; FPKM "233.0138642752"; frac "1.000000"; conf_lo "22.182837"; conf_hi "38.203775"; cov "306.188212";
Sb_20131119_contig_100107       Cufflinks       transcript      2       255     1000    .       .       gene_id "CUFF.11"; transcript_id "CUFF.11.1"; FPKM "29.5909083872"; frac "1.000000"; conf_lo "3.216803"; conf_hi "9.918474"; cov "41.309085";
Sb_20131119_contig_100107       Cufflinks       exon    2       255     1000    .       .       gene_id "CUFF.11"; transcript_id "CUFF.11.1"; exon_number "1"; FPKM "29.5909083872"; frac "1.000000"; conf_lo "3.216803"; conf_hi "9.918474"; cov "41.309085";
Sb_20131119_contig_100107       Cufflinks       transcript      2041    2331    1000    .       .       gene_id "CUFF.12"; transcript_id "CUFF.12.1"; FPKM "18.4360530685"; frac "1.000000"; conf_lo "2.807793"; conf_hi "8.657363"; cov "24.668034";
Sb_20131119_contig_100107       Cufflinks       exon    2041    2331    1000    .       .       gene_id "CUFF.12"; transcript_id "CUFF.12.1"; exon_number "1"; FPKM "18.4360530685"; frac "1.000000"; conf_lo "2.807793"; conf_hi "8.657363"; cov "24.668034";
Sb_20131119_contig_100111       Cufflinks       transcript      21      129     1000    .       .       gene_id "CUFF.14"; transcript_id "CUFF.14.1"; FPKM "3960.0565774823"; frac "1.000000"; conf_lo "5.622026"; conf_hi "18.740088"; cov "4272.822539";
Sb_20131119_contig_100111       Cufflinks       exon    21      129     1000    .       .       gene_id "CUFF.14"; transcript_id "CUFF.14.1"; exon_number "1"; FPKM "3960.0565774823"; frac "1.000000"; conf_lo "5.622026"; conf_hi "18.740088"; cov "4272.822539";
Sb_20131119_contig_100121       Cufflinks       transcript      1756    2236    1000    .       .       gene_id "CUFF.17"; transcript_id "CUFF.17.1"; FPKM "52.3538569152"; frac "1.000000"; conf_lo "23.640043"; conf_hi "34.398386"; cov "75.482038";
Sb_20131119_contig_100121       Cufflinks       exon    1756    2236    1000    .       .       gene_id "CUFF.17"; transcript_id "CUFF.17.1"; exon_number "1"; FPKM "52.3538569152"; frac "1.000000"; conf_lo "23.640043"; conf_hi "34.398386"; cov "75.482038";
Sb_20131119_contig_100192       Cufflinks       transcript      1840    2212    1000    .       .       gene_id "CUFF.20"; transcript_id "CUFF.20.1"; FPKM "24.9098132799"; frac "1.000000"; conf_lo "7.484312"; conf_hi "14.786080"; cov "49.916331";
Sb_20131119_contig_100192       Cufflinks       exon    1840    2212    1000    .       .       gene_id "CUFF.20"; transcript_id "CUFF.20.1"; exon_number "1"; FPKM "24.9098132799"; frac "1.000000"; conf_lo "7.484312"; conf_hi "14.786080"; cov "49.916331";
Sb_20131119_contig_100107       Cufflinks       transcript      430     902     1000    .       .       gene_id "CUFF.13"; transcript_id "CUFF.13.1"; FPKM "25.0500552879"; frac "1.000000"; conf_lo "10.508448"; conf_hi "18.137870"; cov "45.523657";
Sb_20131119_contig_100107       Cufflinks       exon    430     902     1000    .       .       gene_id "CUFF.13"; transcript_id "CUFF.13.1"; exon_number "1"; FPKM "25.0500552879"; frac "1.000000"; conf_lo "10.508448"; conf_hi "18.137870"; cov "45.523657";
Sb_20131119_contig_100192       Cufflinks       transcript      1       616     1000    .       .       gene_id "CUFF.19"; transcript_id "CUFF.19.1"; FPKM "42.2951435419"; frac "1.000000"; conf_lo "23.433223"; conf_hi "33.823425"; cov "88.806988";
Sb_20131119_contig_100192       Cufflinks       exon    1       616     1000    .       .       gene_id "CUFF.19"; transcript_id "CUFF.19.1"; exon_number "1"; FPKM "42.2951435419"; frac "1.000000"; conf_lo "23.433223"; conf_hi "33.823425"; cov "88.806988";
Sb_20131119_contig_10011        Cufflinks       transcript      219     353     1000    .       .       gene_id "CUFF.15"; transcript_id "CUFF.15.1"; FPKM "4118.4602721559"; frac "1.000000"; conf_lo "53.966826"; conf_hi "84.228598"; cov "5420.068745";
Sb_20131119_contig_10011        Cufflinks       exon    219     353     1000    .       .       gene_id "CUFF.15"; transcript_id "CUFF.15.1"; exon_number "1"; FPKM "4118.4602721559"; frac "1.000000"; conf_lo "53.966826"; conf_hi "84.228598"; cov "5420.068745";
Sb_20131119_contig_100040       Cufflinks       transcript      945     2276    1000    -       .       gene_id "CUFF.38"; transcript_id "CUFF.38.1"; FPKM "69.4702369875"; frac "1.000000"; conf_lo "60.275496"; conf_hi "70.712087"; cov "143.103653";
Sb_20131119_contig_100040       Cufflinks       exon    945     1966    1000    -       .       gene_id "CUFF.38"; transcript_id "CUFF.38.1"; exon_number "1"; FPKM "69.4702369875"; frac "1.000000"; conf_lo "60.275496"; conf_hi "70.712087"; cov "143.103653";
Sb_20131119_contig_100040       Cufflinks       exon    2079    2276    1000    -       .       gene_id "CUFF.38"; transcript_id "CUFF.38.1"; exon_number "2"; FPKM "69.4702369875"; frac "1.000000"; conf_lo "60.275496"; conf_hi "70.712087"; cov "143.103653";
Sb_20131119_contig_100121       Cufflinks       transcript      1       150     1000    .       .       gene_id "CUFF.16"; transcript_id "CUFF.16.1"; FPKM "1080.5400540118"; frac "1.000000"; conf_lo "23.150255"; conf_hi "44.484805"; cov "1391.311243";
Sb_20131119_contig_100121       Cufflinks       exon    1       150     1000    .       .       gene_id "CUFF.16"; transcript_id "CUFF.16.1"; exon_number "1"; FPKM "1080.5400540118"; frac "1.000000"; conf_lo "23.150255"; conf_hi "44.484805"; cov "1391.311243";
Sb_20131119_contig_10022        Cufflinks       transcript      1510    2616    1000    .       .       gene_id "CUFF.22"; transcript_id "CUFF.22.1"; FPKM "10.9257515912"; frac "1.000000"; conf_lo "7.442428"; conf_hi "11.194395"; cov "24.833916";
Sb_20131119_contig_10022        Cufflinks       exon    1510    2616    1000    .       .       gene_id "CUFF.22"; transcript_id "CUFF.22.1"; exon_number "1"; FPKM "10.9257515912"; frac "1.000000"; conf_lo "7.442428"; conf_hi "11.194395"; cov "24.833916";
Sb_20131119_contig_100112       Cufflinks       transcript      933     3862    1000    .       .       gene_id "CUFF.21"; transcript_id "CUFF.21.1"; FPKM "13.6659124939"; frac "1.000000"; conf_lo "11.688997"; conf_hi "14.524101"; cov "32.588772";
Sb_20131119_contig_100112       Cufflinks       exon    933     3862    1000    .       .       gene_id "CUFF.21"; transcript_id "CUFF.21.1"; exon_number "1"; FPKM "13.6659124939"; frac "1.000000"; conf_lo "11.688997"; conf_hi "14.524101"; cov "32.588772";
Sb_20131119_contig_10022        Cufflinks       transcript      2857    6365    1000    .       .       gene_id "CUFF.24"; transcript_id "CUFF.24.1"; FPKM "13.7863265713"; frac "1.000000"; conf_lo "12.049946"; conf_hi "14.805328"; cov "32.698360";
Sb_20131119_contig_10022        Cufflinks       exon    2857    6365    1000    .       .       gene_id "CUFF.24"; transcript_id "CUFF.24.1"; exon_number "1"; FPKM "13.7863265713"; frac "1.000000"; conf_lo "12.049946"; conf_hi "14.805328"; cov "32.698360";
Sb_20131119_contig_100080       Cufflinks       transcript      39      1611    1000    -       .       gene_id "CUFF.18"; transcript_id "CUFF.18.1"; FPKM "44.1667134200"; frac "1.000000"; conf_lo "34.717689"; conf_hi "42.267104"; cov "77.065569";
Sb_20131119_contig_100080       Cufflinks       exon    39      135     1000    -       .       gene_id "CUFF.18"; transcript_id "CUFF.18.1"; exon_number "1"; FPKM "44.1667134200"; frac "1.000000"; conf_lo "34.717689"; conf_hi "42.267104"; cov "77.065569";
Sb_20131119_contig_100080       Cufflinks       exon    207     1041    1000    -       .       gene_id "CUFF.18"; transcript_id "CUFF.18.1"; exon_number "2"; FPKM "44.1667134200"; frac "1.000000"; conf_lo "34.717689"; conf_hi "42.267104"; cov "77.065569";
Sb_20131119_contig_100080       Cufflinks       exon    1128    1611    1000    -       .       gene_id "CUFF.18"; transcript_id "CUFF.18.1"; exon_number "3"; FPKM "44.1667134200"; frac "1.000000"; conf_lo "34.717689"; conf_hi "42.267104"; cov "77.065569";
Sb_20131119_contig_10026        Cufflinks       transcript      1       1002    1000    .       .       gene_id "CUFF.49"; transcript_id "CUFF.49.1"; FPKM "49.5933867311"; frac "1.000000"; conf_lo "35.335602"; conf_hi "43.965643"; cov "84.354408";
Sb_20131119_contig_10026        Cufflinks       exon    1       1002    1000    .       .       gene_id "CUFF.49"; transcript_id "CUFF.49.1"; exon_number "1"; FPKM "49.5933867311"; frac "1.000000"; conf_lo "35.335602"; conf_hi "43.965643"; cov "84.354408";
Sb_20131119_contig_10031        Cufflinks       transcript      26      377     1000    .       .       gene_id "CUFF.25"; transcript_id "CUFF.25.1"; FPKM "43.3120795956"; frac "1.000000"; conf_lo "12.379816"; conf_hi "22.244981"; cov "78.396470";
Sb_20131119_contig_10031        Cufflinks       exon    26      377     1000    .       .       gene_id "CUFF.25"; transcript_id "CUFF.25.1"; exon_number "1"; FPKM "43.3120795956"; frac "1.000000"; conf_lo "12.379816"; conf_hi "22.244981"; cov "78.396470";
Sb_20131119_contig_10032        Cufflinks       transcript      5174    5386    1000    .       .       gene_id "CUFF.29"; transcript_id "CUFF.29.1"; FPKM "386.7080614787"; frac "1.000000"; conf_lo "33.884660"; conf_hi "52.744990"; cov "584.846045";
Sb_20131119_contig_10032        Cufflinks       exon    5174    5386    1000    .       .       gene_id "CUFF.29"; transcript_id "CUFF.29.1"; exon_number "1"; FPKM "386.7080614787"; frac "1.000000"; conf_lo "33.884660"; conf_hi "52.744990"; cov "584.846045";
Sb_20131119_contig_10032        Cufflinks       transcript      160     3305    1000    .       .       gene_id "CUFF.35"; transcript_id "CUFF.35.1"; FPKM "5.9167626598"; frac "1.000000"; conf_lo "4.912969"; conf_hi "6.795913"; cov "12.795227";
Sb_20131119_contig_10032        Cufflinks       exon    160     3305    1000    .       .       gene_id "CUFF.35"; transcript_id "CUFF.35.1"; exon_number "1"; FPKM "5.9167626598"; frac "1.000000"; conf_lo "4.912969"; conf_hi "6.795913"; cov "12.795227";
Sb_20131119_contig_100403       Cufflinks       transcript      822     1053    1000    .       .       gene_id "CUFF.30"; transcript_id "CUFF.30.1"; FPKM "90.3398746415"; frac "1.000000"; conf_lo "9.098097"; conf_hi "19.663630"; cov "118.213868";
Sb_20131119_contig_100403       Cufflinks       exon    822     1053    1000    .       .       gene_id "CUFF.30"; transcript_id "CUFF.30.1"; exon_number "1"; FPKM "90.3398746415"; frac "1.000000"; conf_lo "9.098097"; conf_hi "19.663630"; cov "118.213868";
Sb_20131119_contig_100412       Cufflinks       transcript      226     2367    1000    .       .       gene_id "CUFF.32"; transcript_id "CUFF.32.1"; FPKM "5.4971611120"; frac "1.000000"; conf_lo "4.259535"; conf_hi "6.262152"; cov "11.356891";
Sb_20131119_contig_100412       Cufflinks       exon    226     2367    1000    .       .       gene_id "CUFF.32"; transcript_id "CUFF.32.1"; exon_number "1"; FPKM "5.4971611120"; frac "1.000000"; conf_lo "4.259535"; conf_hi "6.262152"; cov "11.356891";
Sb_20131119_contig_10031        Cufflinks       transcript      1415    3072    1000    .       .       gene_id "CUFF.36"; transcript_id "CUFF.36.1"; FPKM "55.9644866986"; frac "1.000000"; conf_lo "45.789638"; conf_hi "53.715558"; cov "119.323264";
Sb_20131119_contig_10031        Cufflinks       exon    1415    3072    1000    .       .       gene_id "CUFF.36"; transcript_id "CUFF.36.1"; exon_number "1"; FPKM "55.9644866986"; frac "1.000000"; conf_lo "45.789638"; conf_hi "53.715558"; cov "119.323264";
Sb_20131119_contig_100403       Cufflinks       transcript      2827    3748    1000    .       .       gene_id "CUFF.31"; transcript_id "CUFF.31.1"; FPKM "5.2492236693"; frac "1.000000"; conf_lo "3.027818"; conf_hi "5.981787"; cov "12.544434";
Sb_20131119_contig_100403       Cufflinks       exon    2827    3748    1000    .       .       gene_id "CUFF.31"; transcript_id "CUFF.31.1"; exon_number "1"; FPKM "5.2492236693"; frac "1.000000"; conf_lo "3.027818"; conf_hi "5.981787"; cov "12.544434";
Sb_20131119_contig_100328       Cufflinks       transcript      1505    3535    1000    .       .       gene_id "CUFF.33"; transcript_id "CUFF.33.1"; FPKM "4.5481695975"; frac "1.000000"; conf_lo "3.453060"; conf_hi "5.296928"; cov "10.333563";
[redacted]
```

GTF file format

Seqid	source	type	start	end	score	strand	phase	attributes
Chr1	Snap	exon	234	1543	.	+	.	gene_id “gene1”; transcript_id “transcript1”;
Chr1	Snap	CDS	577	1543	.	+	0	gene_id “gene1”; transcript_id “transcript1”;
Chr1	Snap	exon	1822	2674	.	+	.	gene_id “gene1”; transcript_id “transcript1”;
Chr1	Snap	CDS	1822	2674	.	+	2	gene_id “gene1”; transcript_id “transcript1”;
		start_ codon						
		stop_ codon						

Transcript Reconstruction from RNA-Seq Reads



NATURE PROTOCOLS | PROTOCOL

Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks

Cole Trapnell, Adam Roberts, Loyal Goff, Geo Pertea, Daehwan Kim, David R Kelley, Harold Pimentel, Steven L Salzberg, John L Rinn & Lior Pachter

Affiliations | **Contributions** | **Corresponding author**

Nature Protocols **7**, 562–578 (2012) | doi:10.1038/nprot.2012.016
Published online 01 March 2012

Transcript Reconstruction from RNA-Seq Reads

End-to-end **Transcriptome**-based
RNA-Seq Analysis
Software Package

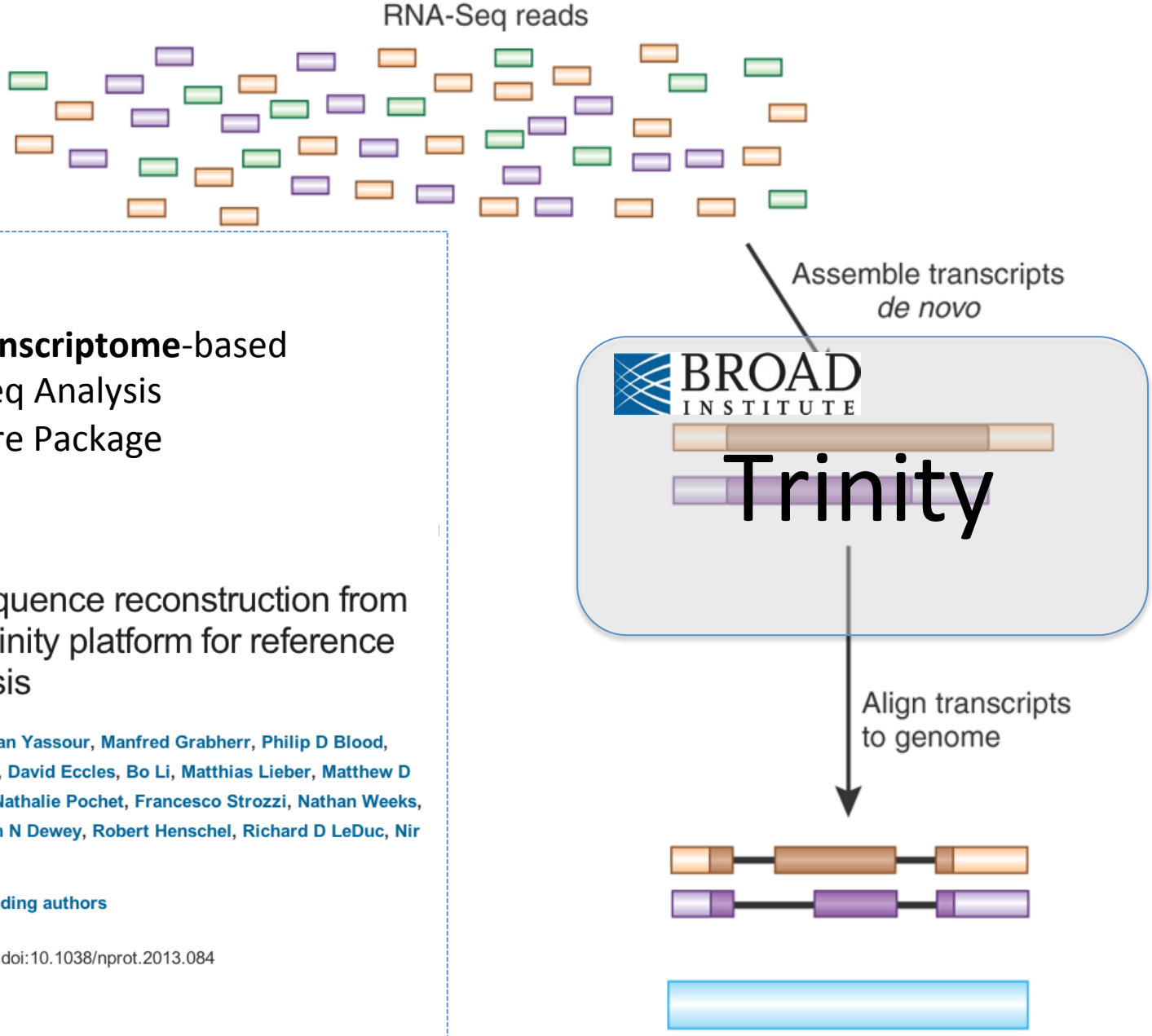
NATURE PROTOCOLS | PROTOCOL

De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis

Brian J Haas, Alexie Papanicolaou, Moran Yassour, Manfred Grabherr, Philip D Blood, Joshua Bowden, Matthew Brian Couger, David Eccles, Bo Li, Matthias Lieber, Matthew D MacManes, Michael Ott, Joshua Orvis, Nathalie Pochet, Francesco Strozzi, Nathan Weeks, Rick Westerman, Thomas William, Colin N Dewey, Robert Henschel, Richard D LeDuc, Nir Friedman & Aviv Regev

Affiliations | Contributions | Corresponding authors

Nature Protocols **8**, 1494–1512 (2013) | doi:10.1038/nprot.2013.084
Published online 11 July 2013



De novo transcriptome assembly

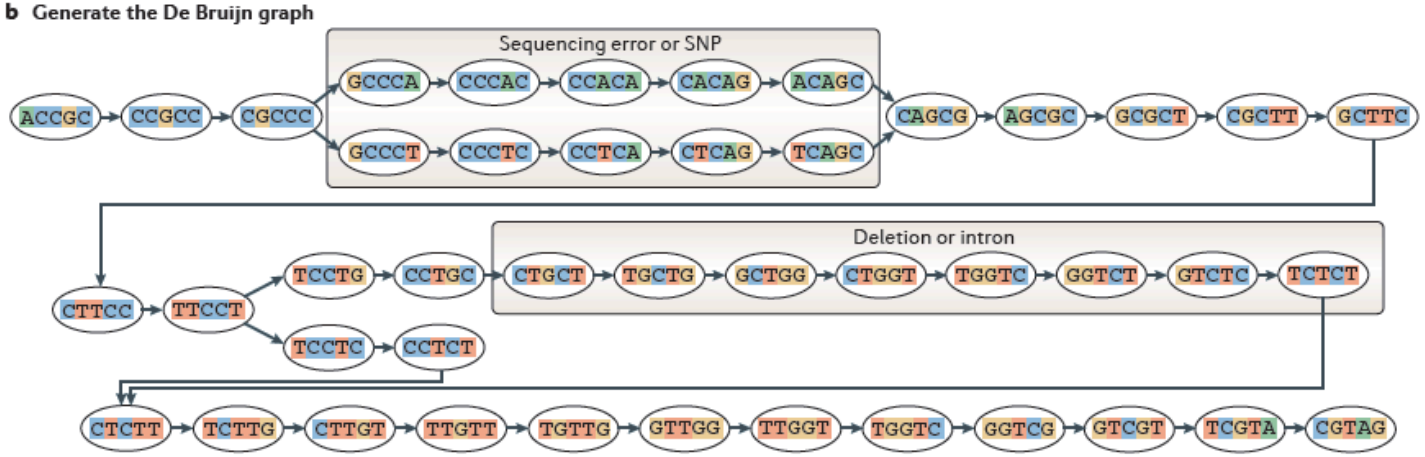
No genome required

Empower studies of non-model organisms

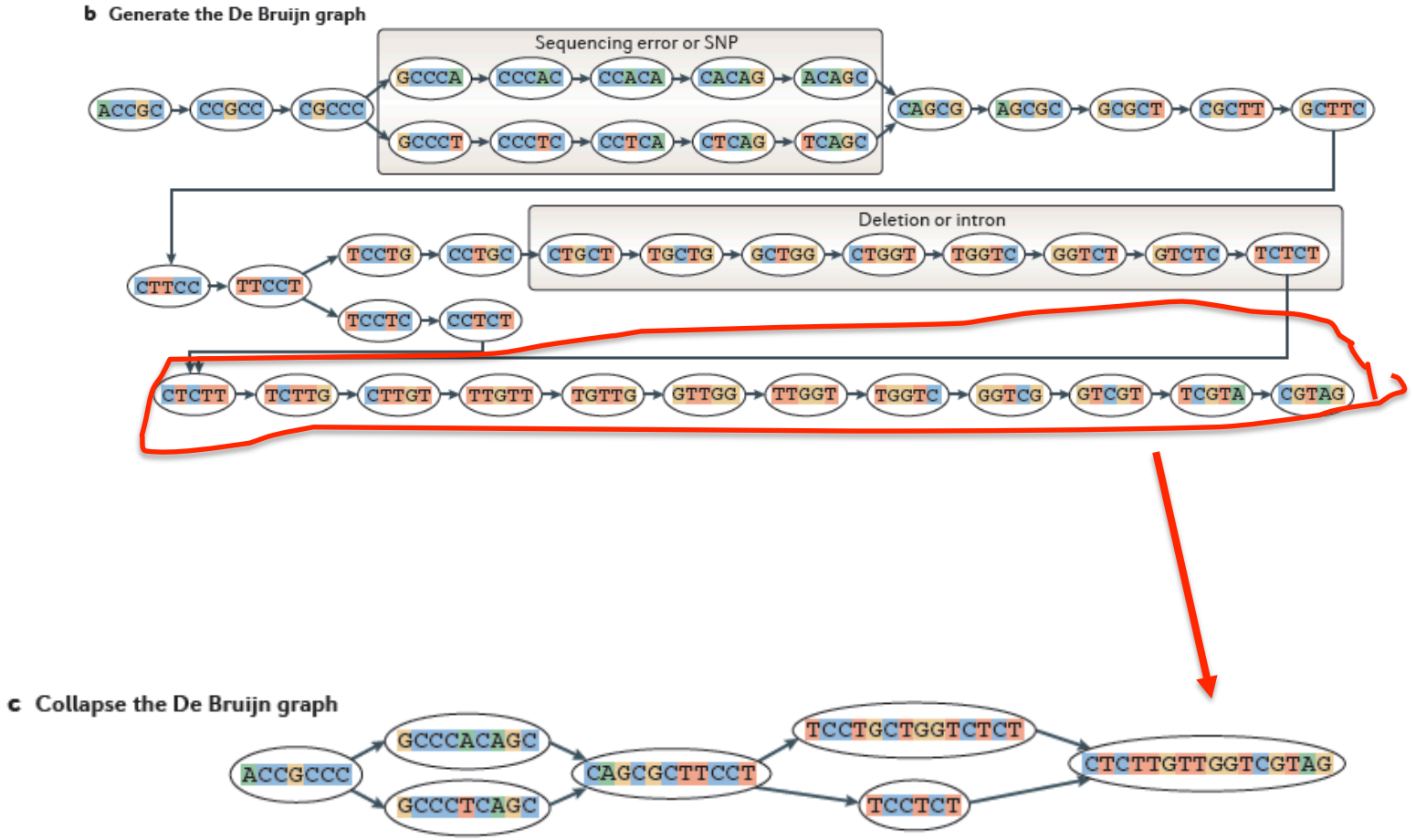
- expressed gene content
- transcript abundance
- differential expression

The General Approach to *De novo* RNA-Seq Assembly Using De Bruijn Graphs

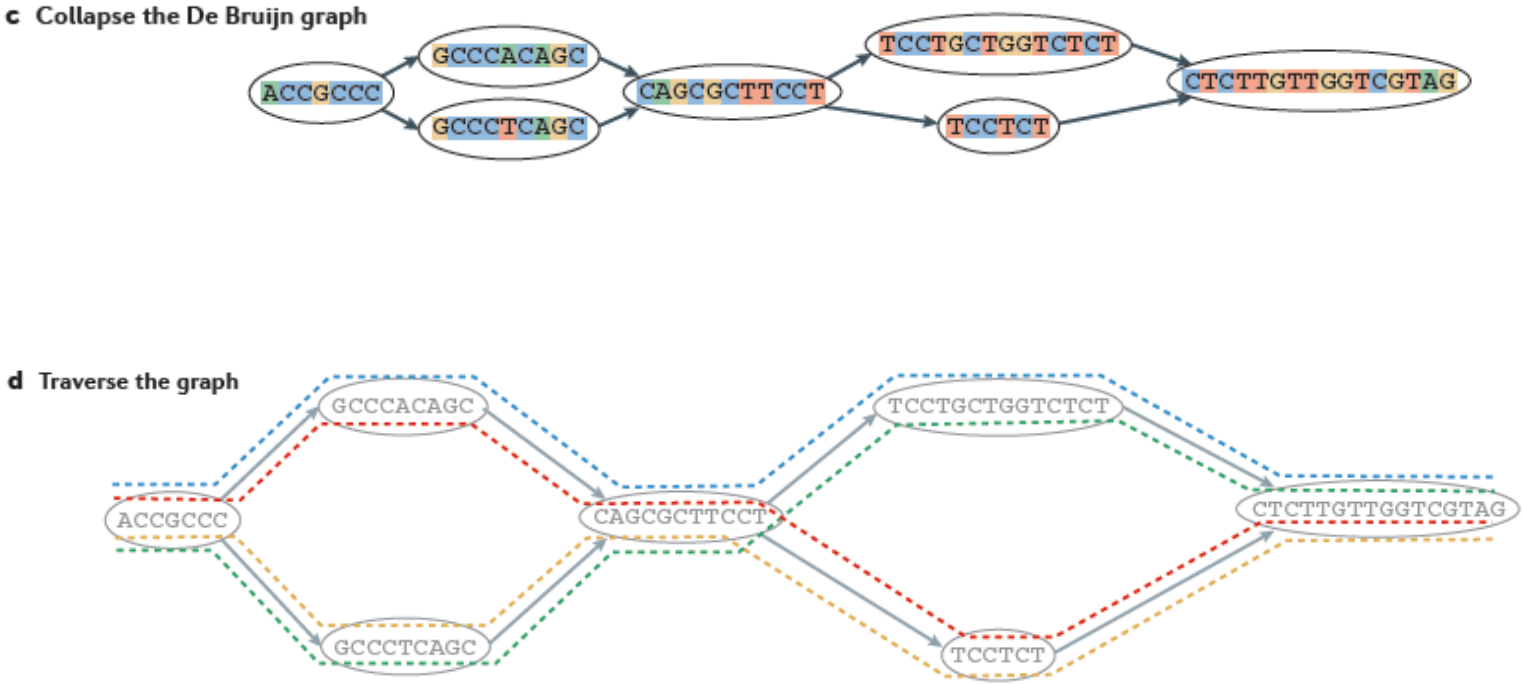
Sequence Assembly via De Bruijn Graphs



From Martin & Wang, Nat. Rev. Genet. 2011



From Martin & Wang, Nat. Rev. Genet. 2011



From Martin & Wang, Nat. Rev. Genet. 2011

Contrasting Genome and Transcriptome Assembly

Genome Assembly

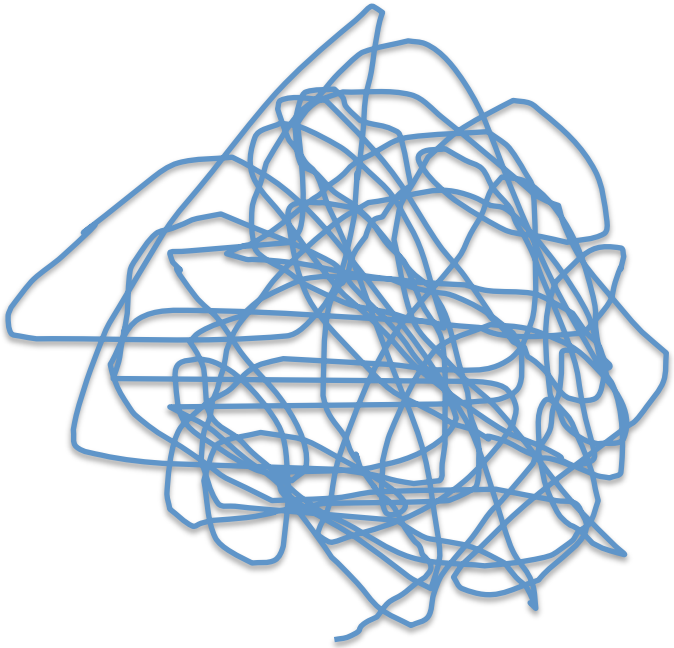
- Uniform coverage
- Single contig per locus
- Double-stranded

Transcriptome Assembly

- Exponentially distributed coverage levels
- Multiple contigs per locus (alt splicing)
- Strand-specific

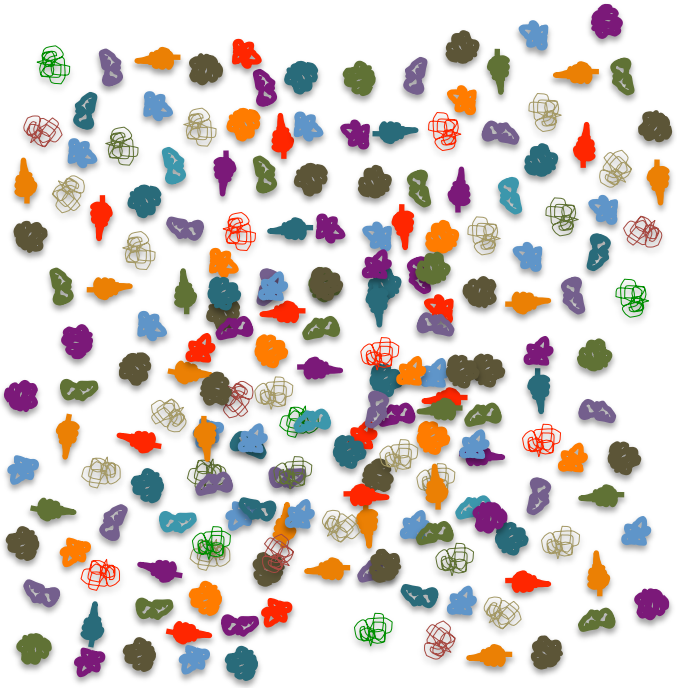
Trinity Aggregates Isolated Transcript Graphs

Genome Assembly
Single Massive Graph



Entire chromosomes represented.

Trinity Transcriptome Assembly
Many Thousands of Small Graphs



Ideally, one graph per expressed gene.

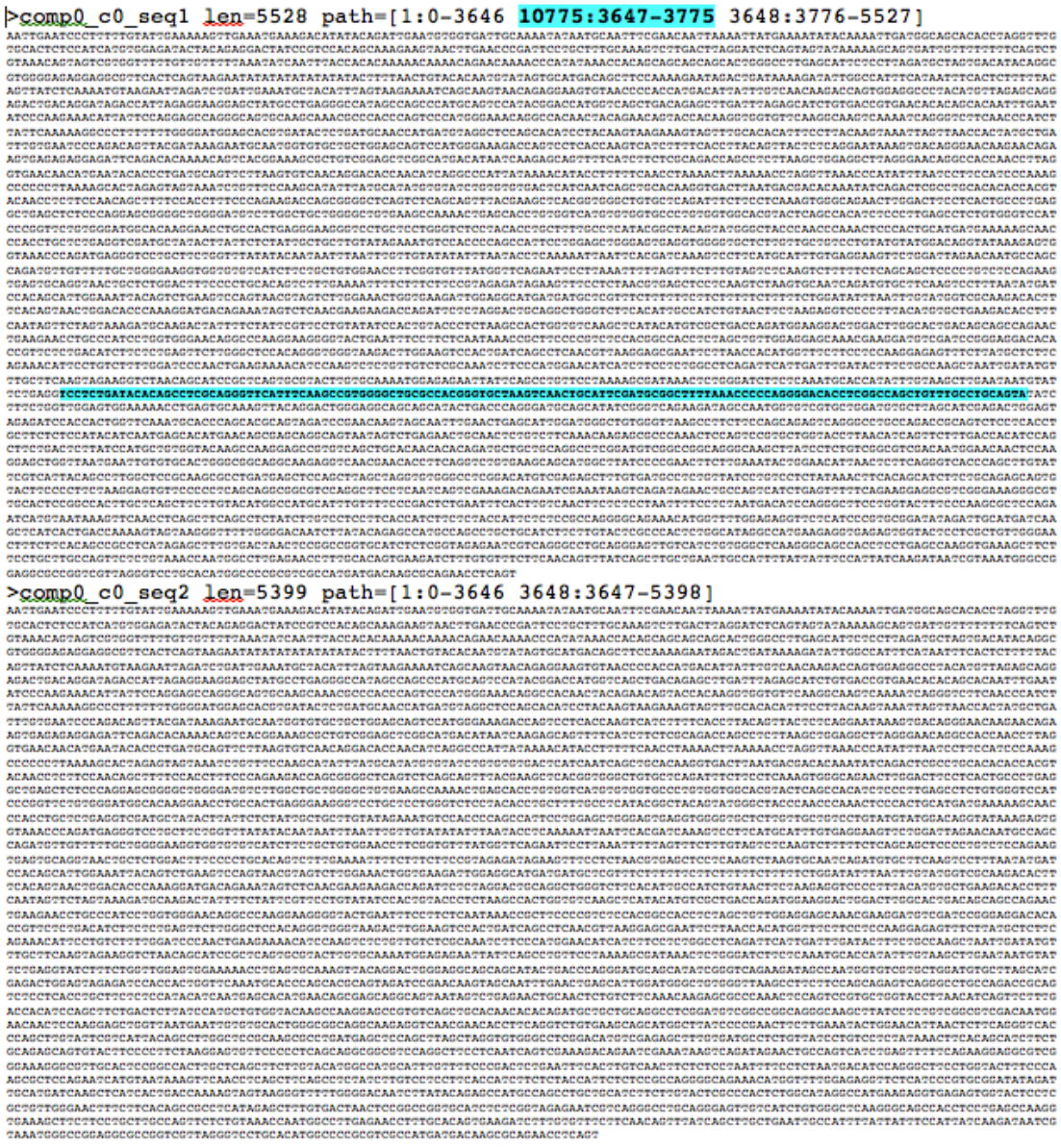
Trinity – How it works:

RNA-Seq reads → **Linear contigs** → **de-Bruijn graphs** → **Transcripts + Isoforms**

...CTTCGCAA...TGATCGGAT...
...ATTCGCAA...TCATCGGAT...

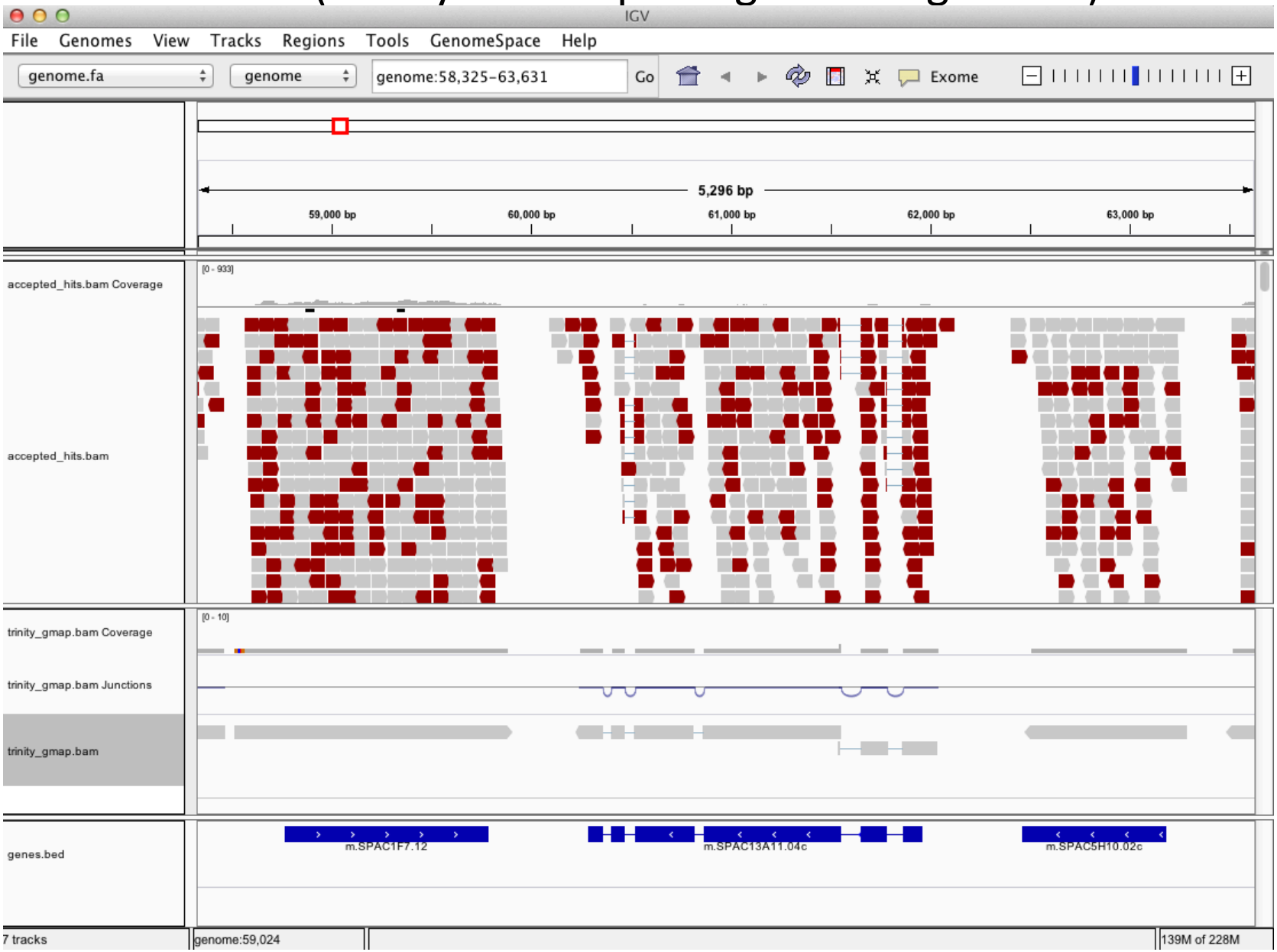
Thousands of disjoint graphs

Trinity output: A multi-fasta file



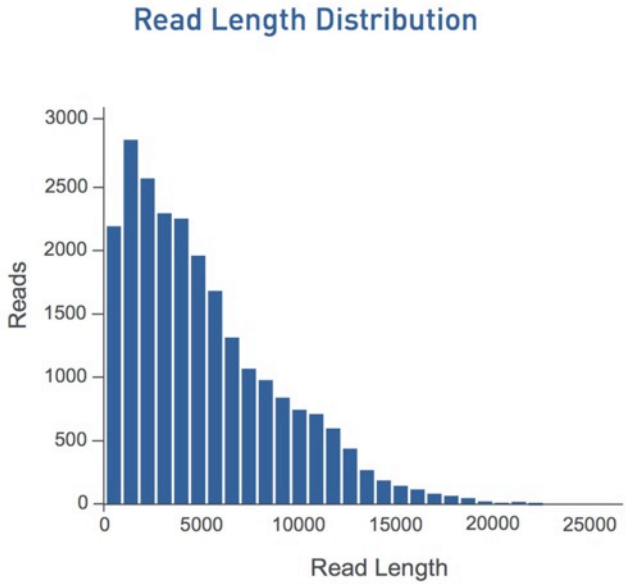
Can align Trinity transcripts to genome scaffolds to examine intron/exon structures

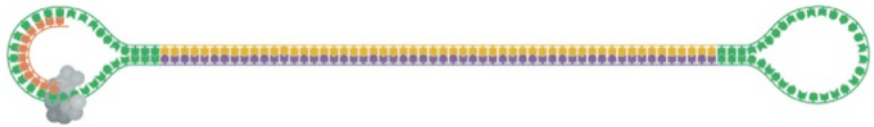
(Trinity transcripts aligned using GMAP)



An alternative: Pacific Biosciences (PacBio)

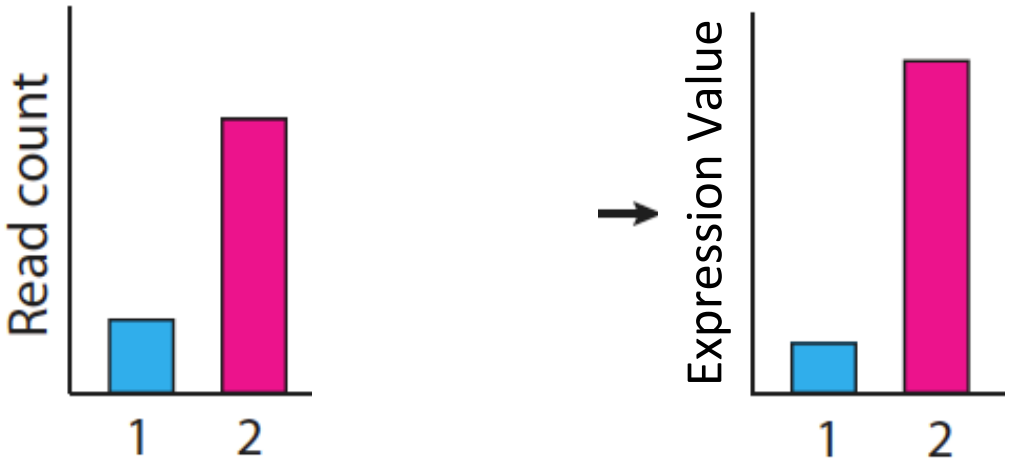
- Pros: Long reads (average 4.5 kbp), can give you full length transcripts in one read
- Cons: High error rate on longer fragments (15%), expensive





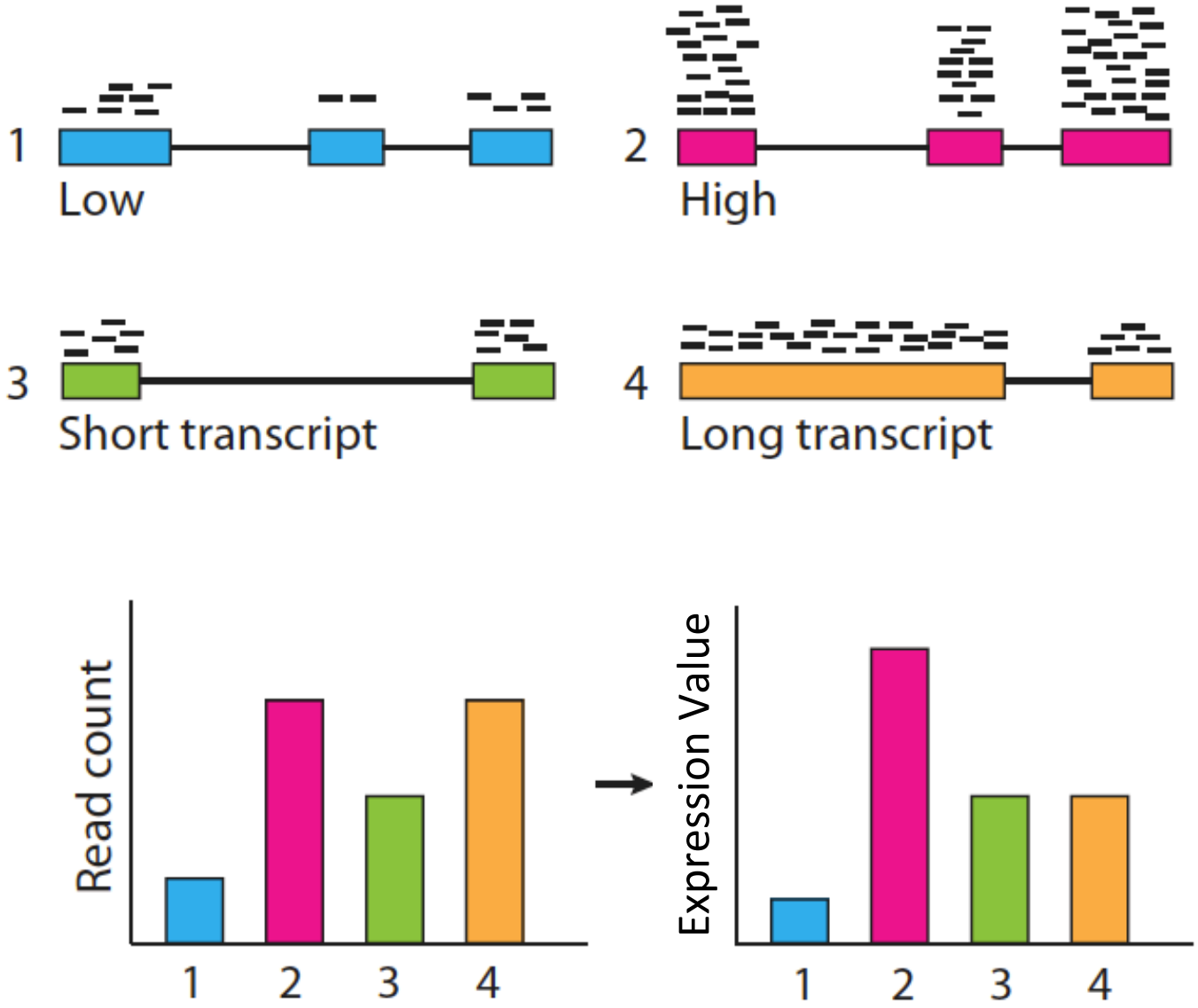
Abundance Estimation
(Computing Expression Values)

Calculating expression of genes and transcripts



Slide courtesy of Cole Trapnell

Calculating expression of genes and transcripts



Slide courtesy of Cole Trapnell

Normalized Expression Values

Gene expression for RNAseq analysis is based in how many reads map to an specific gene. For comparison purposes the counts needs to be normalized. There are different methodologies.

- **RPKM** (Mortazavi et al. 2008): Reads per Kilobase of Exon perMillion of Mapped reads.
- **Upper-quartile** (Bullard et al. 2010): Counts are divided per upper quartile of counts with at least one read.
- **TMM** (Robinson and Oshlack, 2010): Trimmed Means of M values (EdgeR).
- **FPKM** (Trapnell et al. 2010): Fragment per Kilobase of exon per Million of Mapped fragments (Cufflinks).

Normalized Expression Values

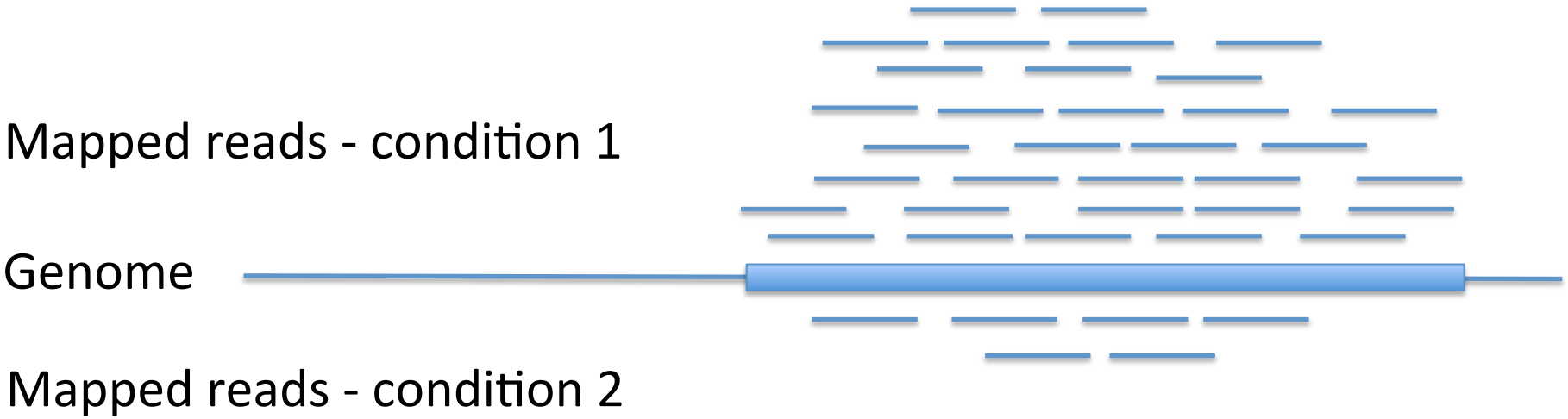
- Transcript-mapped read counts are normalized for both length of the transcript and total depth of sequencing.

- Reported as: Number of RNA-Seq **F**ragments
  Per **K**ilobase of transcript
  per total **M**illion fragments mapped

FPKM

Differential Expression Analysis Using RNA-Seq

Differential expression



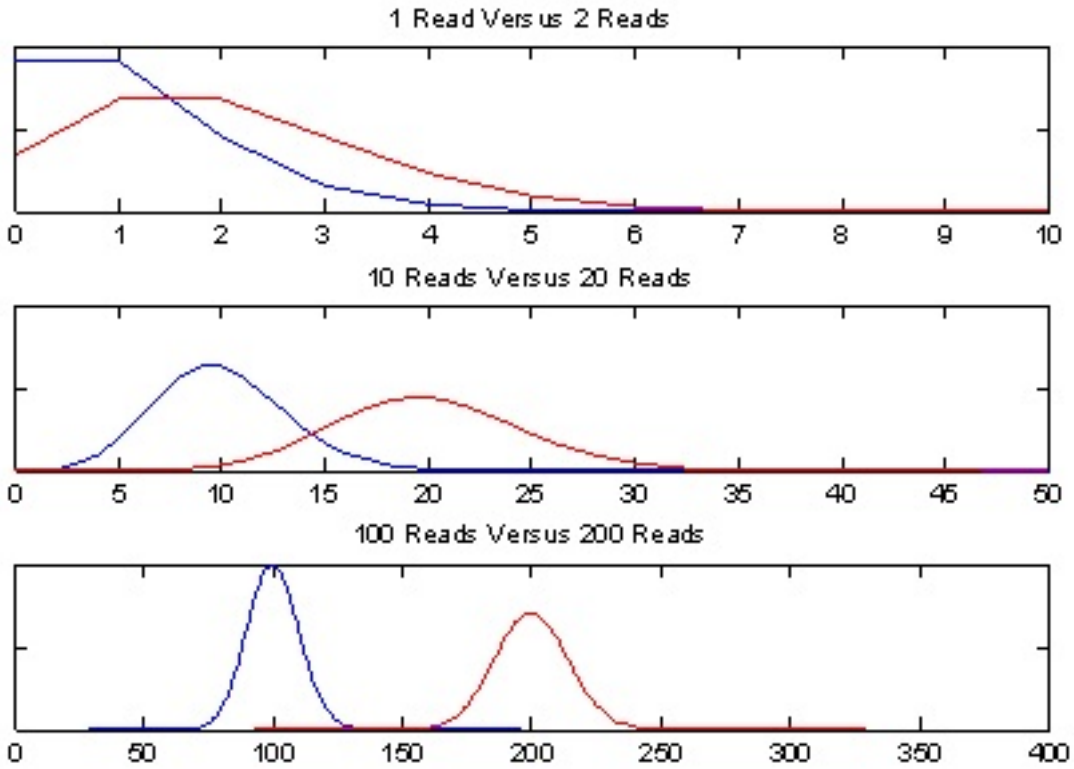
Diff. Expression Analysis Involves

- Counting reads
- Statistical significance testing

	Sample_A	Sample_B	Fold_Change	Significant?
Gene A	1	2	2-fold	No
Gene B	100	200	2-fold	Yes

Beware of concluding fold change from small numbers of counts

Poisson distributions for counts based on **2-fold** expression differences



No confidence in 2-fold difference. Likely observed by chance.

High confidence in 2-fold difference. Unlikely observed by chance.

From: http://gkno2.tumblr.com/post/24629975632/thinking-about-rna-seq-experimental-design-for

More Counts = More Statistical Power

Example: 5000 total reads per sample.

Observed 2-fold differences in read counts.

	SampleA	**Sample B**	**Fisher's Exact Test (P-value)**
geneA	1	2	1.00
geneB	10	20	0.098
geneC	100	200	< 0.001

Tools for DE analysis with RNA-Seq

ShrinkSeq
NoiSeq
baySeq
Vsf
Voom
SAMseq
TSPM
DESeq
EBSeq
NBPSeq
edgeR

+ other (not-R)
including CuffDiff

See: http://www.biomedcentral.com/1471-2105/14/91

Tools for DE analysis with RNA-Seq

Software	Normalization	Notes	URL
ERANGE	RPKM	Python	http://woldlab.caltech.edu/wiki/RNASeq
Scripture	RPKM	Java	http://www.broadinstitute.org/software/scripture
BitSeq*	RPKM	R/Bioconductor, Calculate DE	http://www.bioconductor.org/packages/2.12/bioc/html/BitSeq.html
EdgeR	TMM	R/Bioconductor, Calculate DE	http://www.bioconductor.org/packages/2.11/bioc/html/edgeR.html
Cufflinks*	FPKM	Isoforms, Calculate DE	http://cufflinks.cbcb.umd.edu/
MMSEQ*	FPKM	Isoforms, Haplotypes	http://bgx.org.uk/software/mmseq.html
RSEM*	FPKM	Calculate DE (EBSeq)	http://deweylab.biostat.wisc.edu/rsem/README.html

Glaus P. *et al* (2012) *Bioinformatics* 28:1721-1728 doi:10.1093/bioinformatics/bts260

Differential Gene Expression

Statistical test to evaluate if one gene has an differential expression between two or more conditions. These test can be based in different methodologies.

- **Negative binomial distribution** (DESeq, CuffLinks).
- **Bayesian methods for the negative binomial distribution** (EdgeR, BaySeq, BitSeq).
- **Non-parametric:** models the noise distribution of count changes by contrasting fold-change differences (M) and absolute expression differences (D) (NOISeq).

Tools for DE analysis with RNA-Seq

Software	Normalization	Need Replicas	Input	URL
EdgeR	Library Size / TMM	Yes	Raw Counts	http://www.bioconductor.org/packages/2.11/bioc/html/edgeR.html
DESeq	Library Size	No	Raw Counts	http://bioconductor.org/packages/release/bioc/html/DESeq.html
baySeq	Library Size	Yes	Raw Counts	http://www.bioconductor.org/packages/2.11/bioc/html/baySeq.html
NOISeq	Library Size / RPKM / UpperQ	No	Raw or Normalized Counts	http://bioinfo.cipf.es/noiseq/doku.php?id=start

Differential expression in RNA-seq: a matter of depth. *Genome Res.*21:2213-23

Explorative Data Mining Methods

For **gene expression** there are some common tasks and associated methods for the **data mining**:

- Clustering of the expression values and principal component analysis to reduce the variables.
- Classification using Gene Ontology terms and metabolic annotations
- Summarization visualizing the expression data through heat maps.

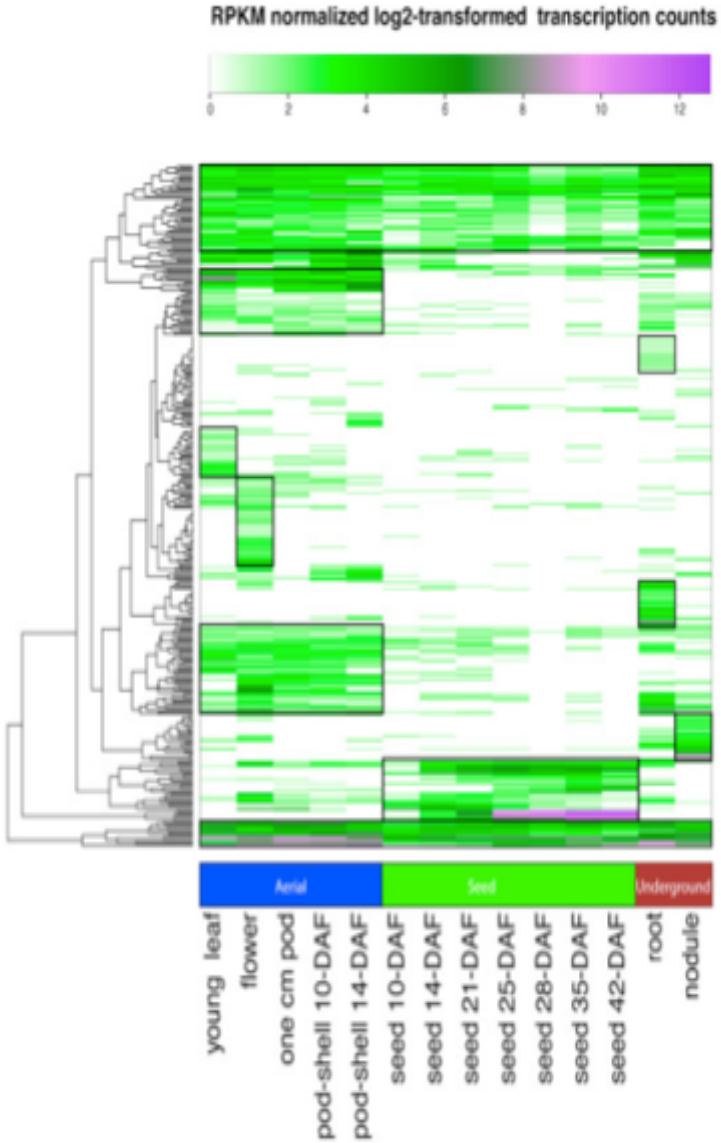
Cluster Analysis and Visualization

Cluster analysis or clustering is the task of **assigning a set of objects into groups** (called clusters) so that the objects in the same cluster are more similar (in some sense or another) to each other than to those in other clusters. Clustering is a main task of **explorative data mining**. The most used **clustering algorithm** for gene expression are:

- **Hierarchical clustering (HCL)**, where the distance between elements is used to build the clusters.
- **K-means clustering (KMC)**, where clusters are represented by a vector. The number of clusters is fixed and the elements are assigned based in its distance to the vector.

Cluster Analysis and Visualization

Severin AJ et al., 2010 BMC Plant Biology, 10:160



One of the most common classification data mining method is the use of gene annotations such as GO terms or metabolic annotations. These methodologies compare two groups between them to find if there are term more represented in one group than in other. Some examples are:

- **Gene Set Enrichment Analysis (GSEA)**, computational method that determines whether an a priori defined set of genes shows statistically significant.
- **Profile comparisons**, each group defines a profile based in the annotation groups (generally GO terms). Profiles are compared to find if they are significantly different.

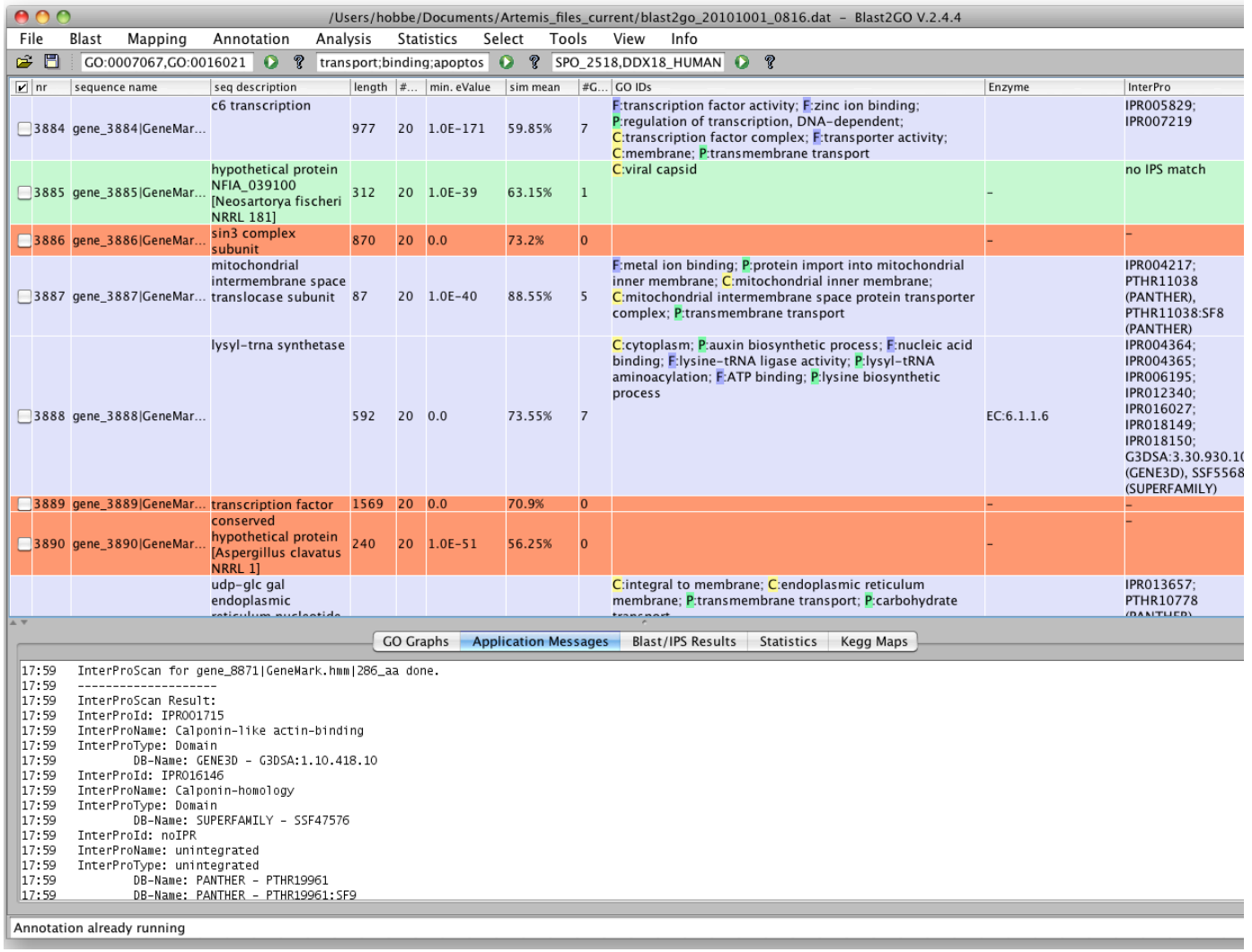
Use of transcripts

- Transcripts can be assembled de novo or from mapped reads and then used in gene expression/differential expression studies
- Can be functionally anntoated

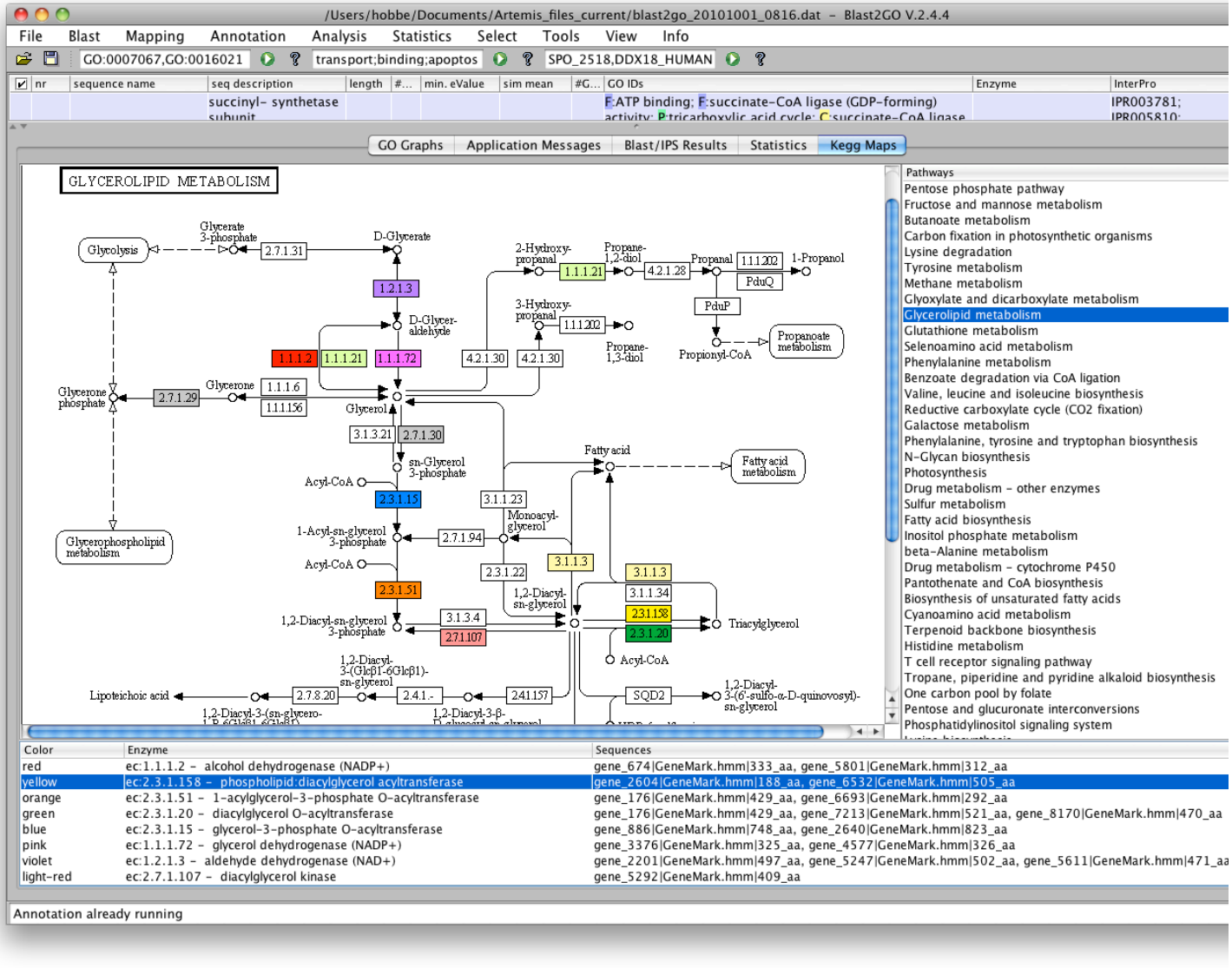
Functional annotation

- Take transcripts from Cufflinks or Trinity
- Annotate the sequences functionally in Blast2GO

Blast2GO


/Users/hobbe/Documents/Artemis_files_current/blast2go_20101001_0816.dat - Blast2GO V.2.4.4
File Blast Mapping Annotation Analysis Statistics Select Tools View Info
GO:0007067,GO:0016021
transport;binding;apoptos
SPO_2518,DDX18_HUMAN
nr | sequence name | seq description | length | #... | min. eValue | sim mean | #G... | GO IDs | Enzyme | InterPro
3884 | gene_3884|GeneMar... | c6 transcription | 977 | 20 | 1.0E-171 | 59.85% | 7 | F:transcription factor activity; F:zinc ion binding; P:regulation of transcription, DNA-dependent; C:transcription factor complex; F:transporter activity; C:membrane; P:transmembrane transport | | IPR005829; IPR007219
3885 | gene_3885|GeneMar... | hypothetical protein NFIA_039100 [Neosartorya fischeri NRRL 181] | 312 | 20 | 1.0E-39 | 63.15% | 1 | C:viral capsid | - | no IPS match
3886 | gene_3886|GeneMar... | sin3 complex subunit | 870 | 20 | 0.0 | 73.2% | 0 | | - | -
3887 | gene_3887|GeneMar... | mitochondrial intermembrane space translocase subunit | 87 | 20 | 1.0E-40 | 88.55% | 5 | F:metal ion binding; P:protein import into mitochondrial inner membrane; C:mitochondrial inner membrane; C:mitochondrial intermembrane space protein transporter complex; P:transmembrane transport | | IPR004217; PTHR11038 (PANTHER), PTHR11038:SF8 (PANTHER)
3888 | gene_3888|GeneMar... | lysyl-trna synthetase | 592 | 20 | 0.0 | 73.55% | 7 | C:cytoplasm; P:auxin biosynthetic process; F:nucleic acid binding; F:lysine-tRNA ligase activity; P:lysyl-tRNA aminoacylation; F:ATP binding; P:lysine biosynthetic process | EC:6.1.1.6 | IPR004364; IPR004365; IPR006195; IPR012340; IPR016027; IPR018149; IPR018150; G3DSA:3.30.930.1((GENE3D), SSF5568 (SUPERFAMILY)
3889 | gene_3889|GeneMar... | transcription factor | 1569 | 20 | 0.0 | 70.9% | 0 | | - | -
3890 | gene_3890|GeneMar... | conserved hypothetical protein [Aspergillus clavatus NRRL 1] | 240 | 20 | 1.0E-51 | 56.25% | 0 | | - | -
udp-glc gal endoplasmic | C:integral to membrane; C:endoplasmic reticulum membrane; P:transmembrane transport; P:carbohydrate | IPR013657; PTHR10778
GO Graphs Application Messages Blast/IPS Results Statistics Kegg Maps
17:59 InterProScan for gene_8871|GeneMark.hmm|286_aa done.
17:59 ----------
17:59 InterProScan Result:
17:59 InterProId: IPR001715
17:59 InterProName: Calponin-like actin-binding
17:59 InterProType: Domain
17:59 DB-Name: GENE3D - G3DSA:1.10.418.10
17:59 InterProId: IPR016146
17:59 InterProName: Calponin-homology
17:59 InterProType: Domain
17:59 DB-Name: SUPERFAMILY - SSF47576
17:59 InterProId: noIPR
17:59 InterProName: unintegrated
17:59 InterProType: unintegrated
17:59 DB-Name: PANTHER - PTHR19961
17:59 DB-Name: PANTHER - PTHR19961:SF9
Annotation already running

KEGG-mapping



Acknowledgement

Henrik, SciLifeLab