



# Canadian Bioinformatics Workshops

[www.bioinformatics.ca](http://www.bioinformatics.ca)

In collaboration with  
**Cold Spring Harbor Laboratory**  
&  
**New York Genome Center**



This page is available in the following languages:

Afrikaans বাংলাৰাখি Català Dansk Deutsch Ελληνικά English English (CA) English (GB) English (US) Esperanto  
Español Castellano (AR) Español (CL) Castellano (CO) Español (Ecuador) Castellano (MX) Castellano (PE)  
Euskara Suomi français français (CA) Galego עברית hrvatski Magyar Italiano 日本語 한국어 Macedonian Malayu  
Nederlands Norsk Sesotho sa Leboa polski Português română slovenski jezik српски srpski (latinica) Sotho svenska  
中文 華語 (台灣) isiZulu



## Attribution-Share Alike 2.5 Canada

### You are free:



**to Share** — to copy, distribute and transmit the work



**to Remix** — to adapt the work



### Under the following conditions:



**Attribution.** You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work).



**Share Alike.** If you alter, transform, or build upon this work, you may distribute the resulting work only under the same or similar licence to this one.

- For any reuse or distribution, you must make clear to others the licence terms of this work.
- Any of the above conditions can be waived if you get permission from the copyright holder.
- The author's moral rights are retained in this licence.

[Disclaimer](#)

Your fair dealing and other rights are in no way affected by the above.

This is a human-readable summary of the Legal Code (the full licence) available in the following languages:  
[English](#) [French](#)

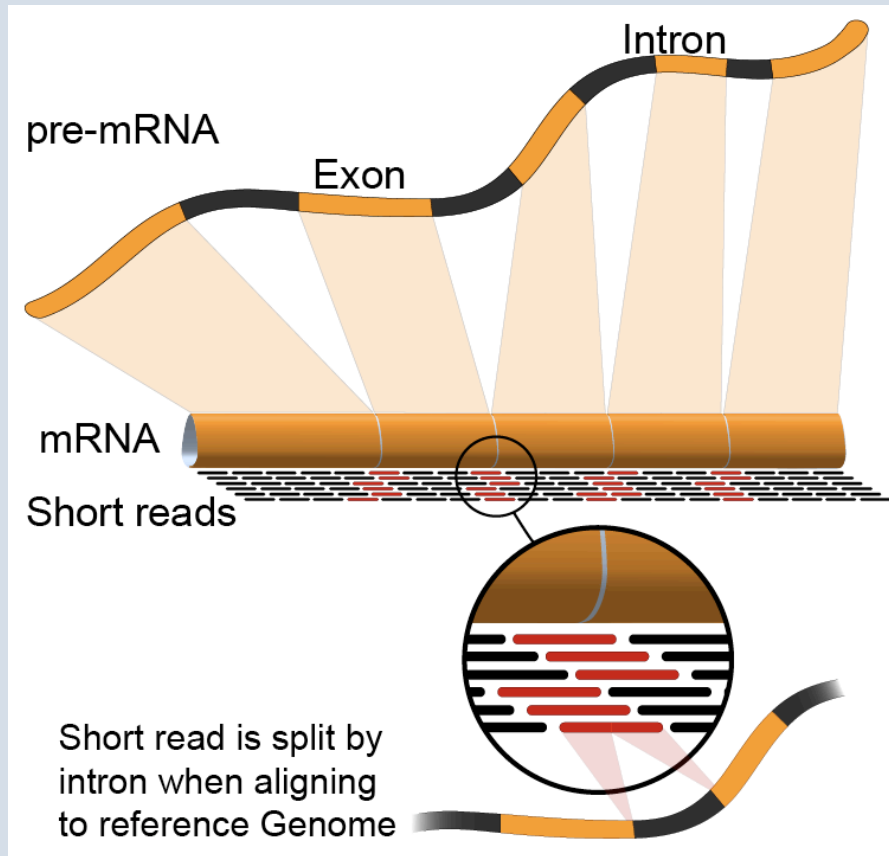
# Module 4

## Isoform discovery and alternative expression (lecture)

Malachi Griffith & Obi Griffith

High-throughput Biology: From Sequence to Networks

April 27-May 3, 2015



# Learning objectives of the course

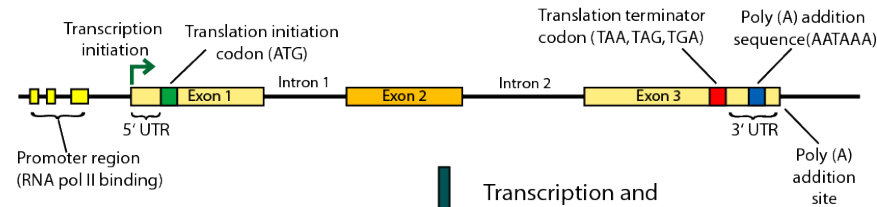
- Module 0: Introduction to cloud computing
- Module 1: Introduction to RNA sequencing
- Module 2: RNA-seq alignment and visualization
- Module 3: Expression and Differential Expression
- **Module 4: Isoform discovery and alternative expression**
  
- Tutorials
  - Provide a working example of an RNA-seq analysis pipeline
  - Run in a ‘reasonable’ amount of time with modest computer resources
  - Self contained, self explanatory, portable

# Learning Objectives of Module

- Explore use of Cufflinks in reference annotation based transcript (RABT) assembly mode and 'de novo' assembly mode.
  - Both modes require a reference genome sequence...

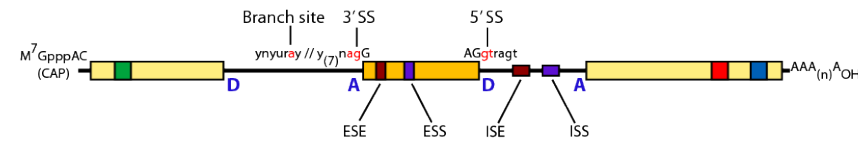
# Review of gene expression

## Double-stranded genomic DNA template



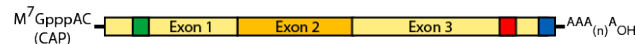
Transcription and polyadenylation

## Single-stranded pre-mRNA (nuclear RNA)



RNA processing

## Mature mRNA

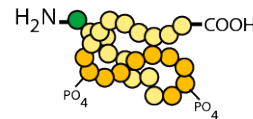


Export to cytoplasm and translation

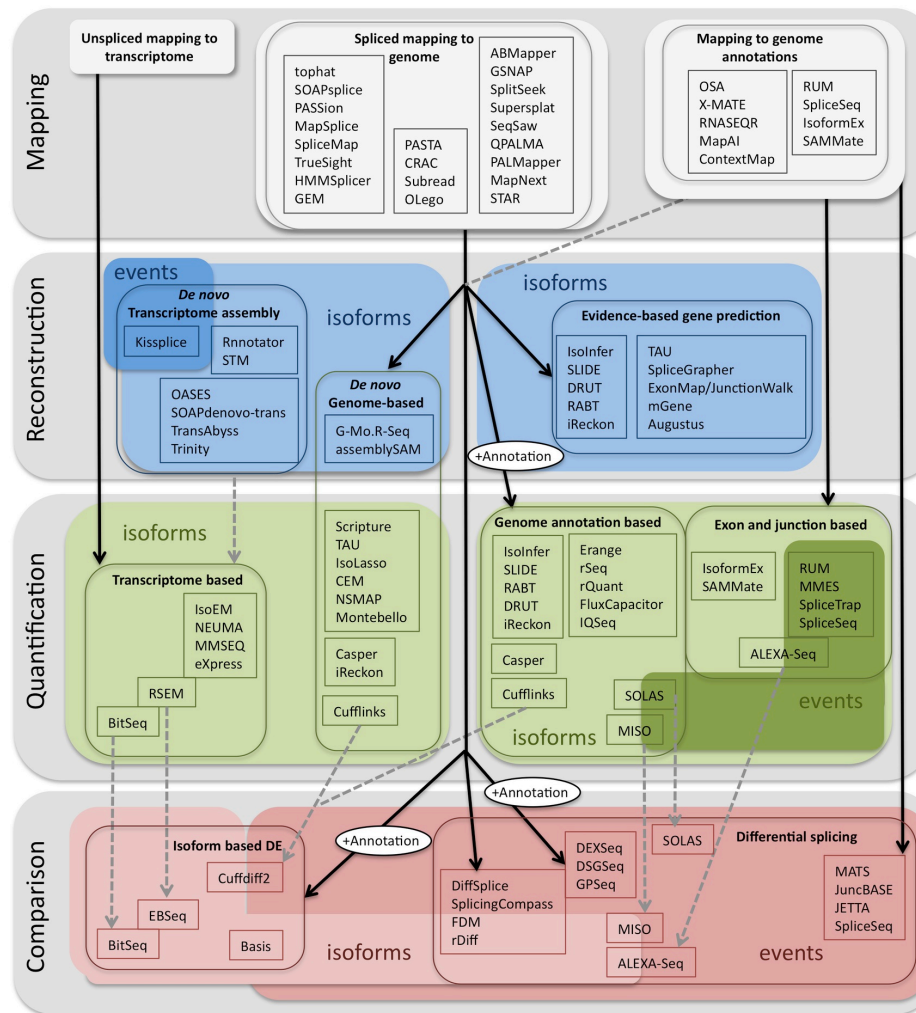
## Protein (amino acid sequence)



Folding, posttranslational modification, subcellular localization, etc.



# Methods to study splicing by RNA-seq



<http://www.rna-seqblog.com/data-analysis/splicing-junction/methods-to-study-splicing-from-rna-seq/>  
<http://arxiv.org/ftp/arxiv/papers/1304/1304.5952.pdf>

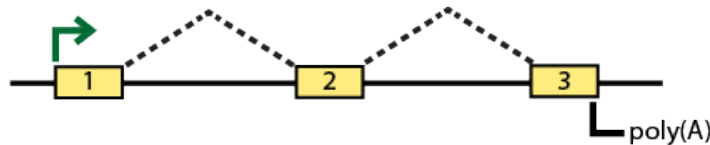


# Useful resources and discussion

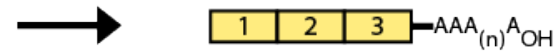
- Best approach to predict novel and alternative splicing events from RNA-seq data
  - <http://www.biostars.org/p/68966/>
  - <http://www.biostars.org/p/62728/>
- Alternative splicing detection
  - <http://www.biostars.org/p/65617/>
  - <http://www.biostars.org/p/11695/>
- Identifying genes that express different isoforms in cancer vs normal RNA-seq data
  - <http://www.biostars.org/p/50365/>
- Cufflinks / Cuffdiff Output - How are tests different?
  - <http://www.biostars.org/p/13525/>
- Visualisation of Alternative splicing events using RNA-seq data
  - <http://www.biostars.org/p/8979/>

# Types of alternative expression - part 1

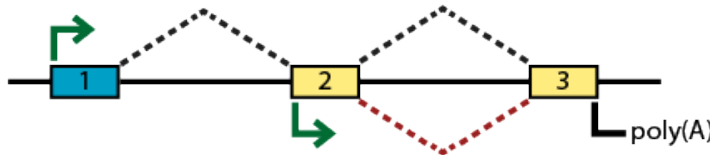
## Simple transcription



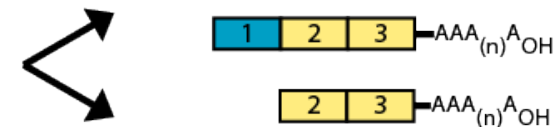
Canonical isoform



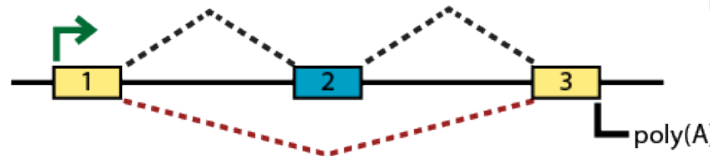
## Alternative transcript initiation



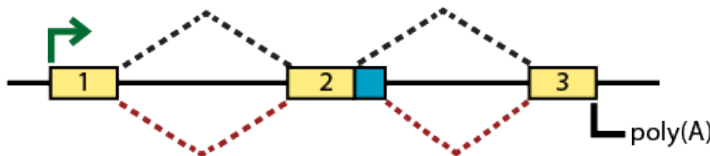
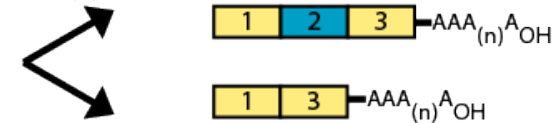
Alternate start sites



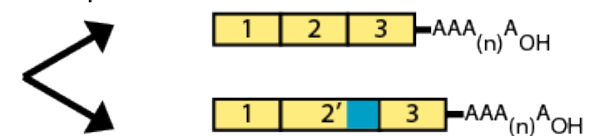
## Alternative splicing



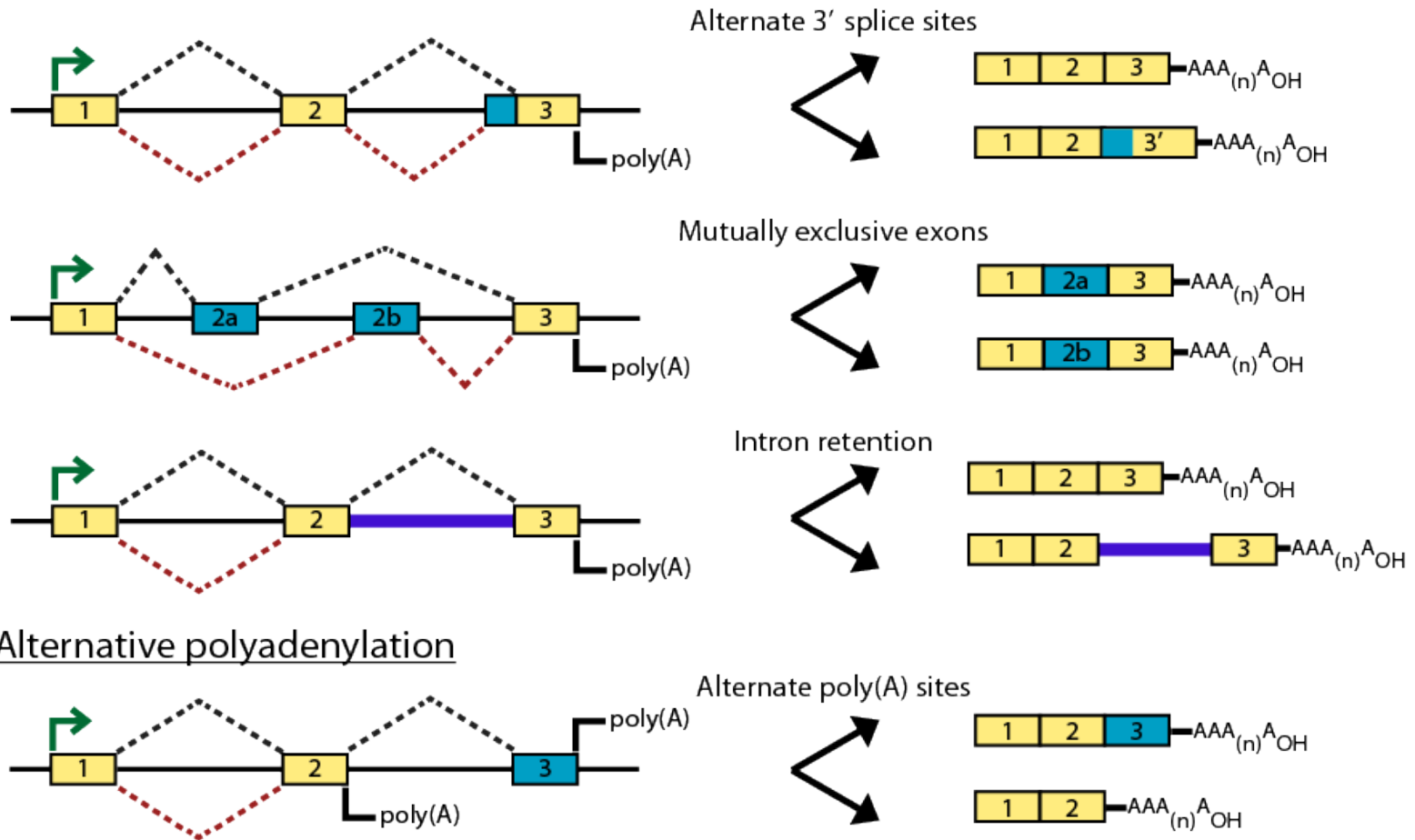
Cassette (skipped) exon



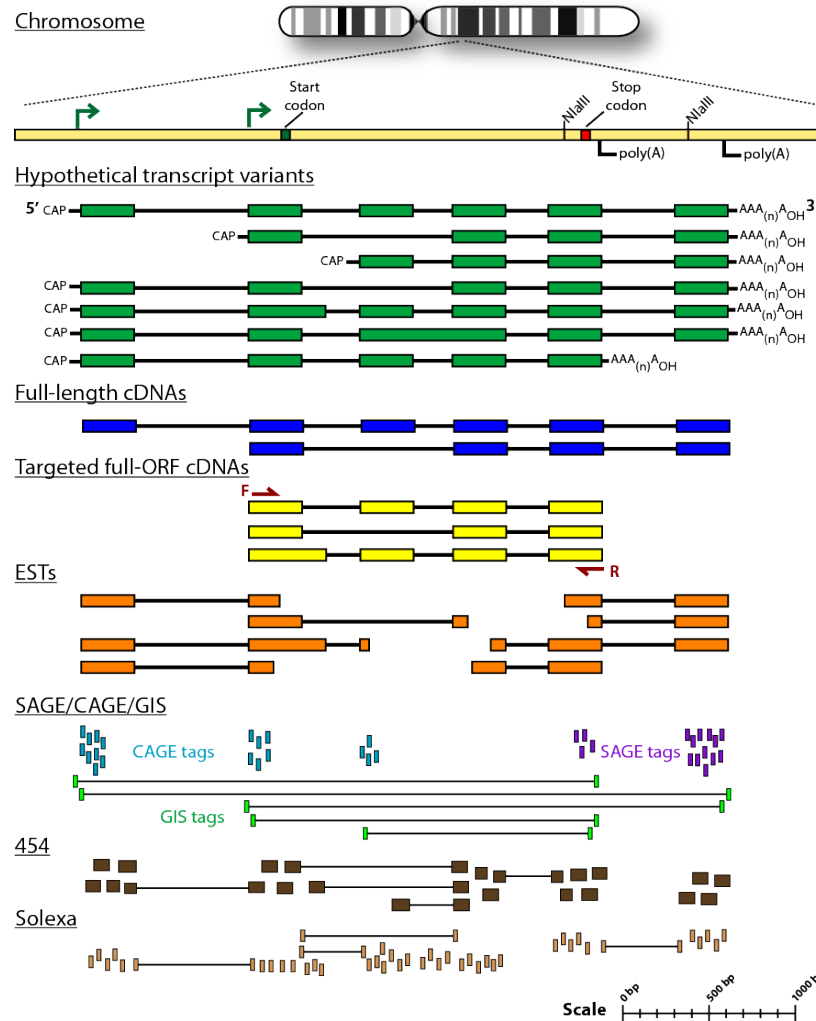
Alternate 5' splice sites



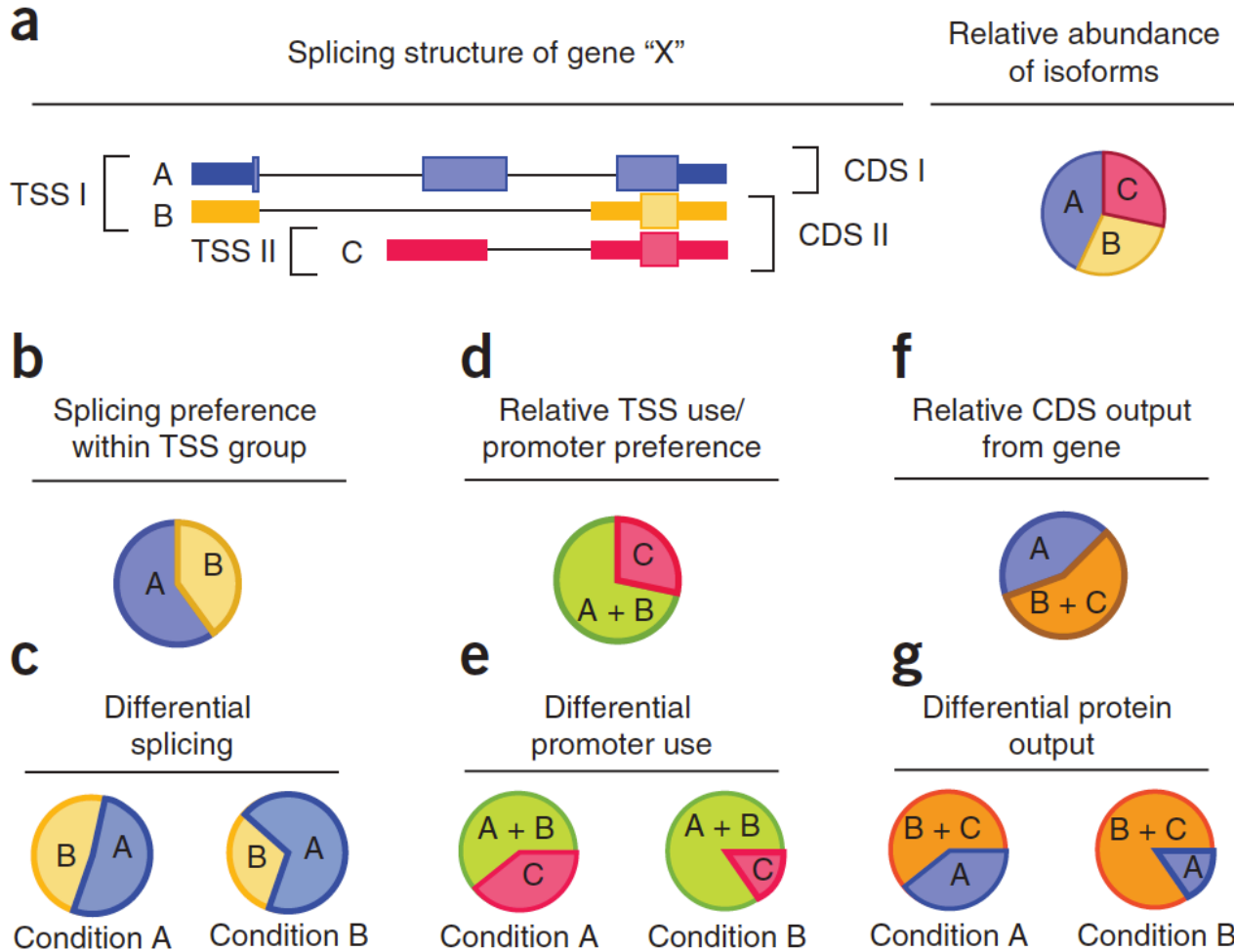
# Types of alternative expression – part 2



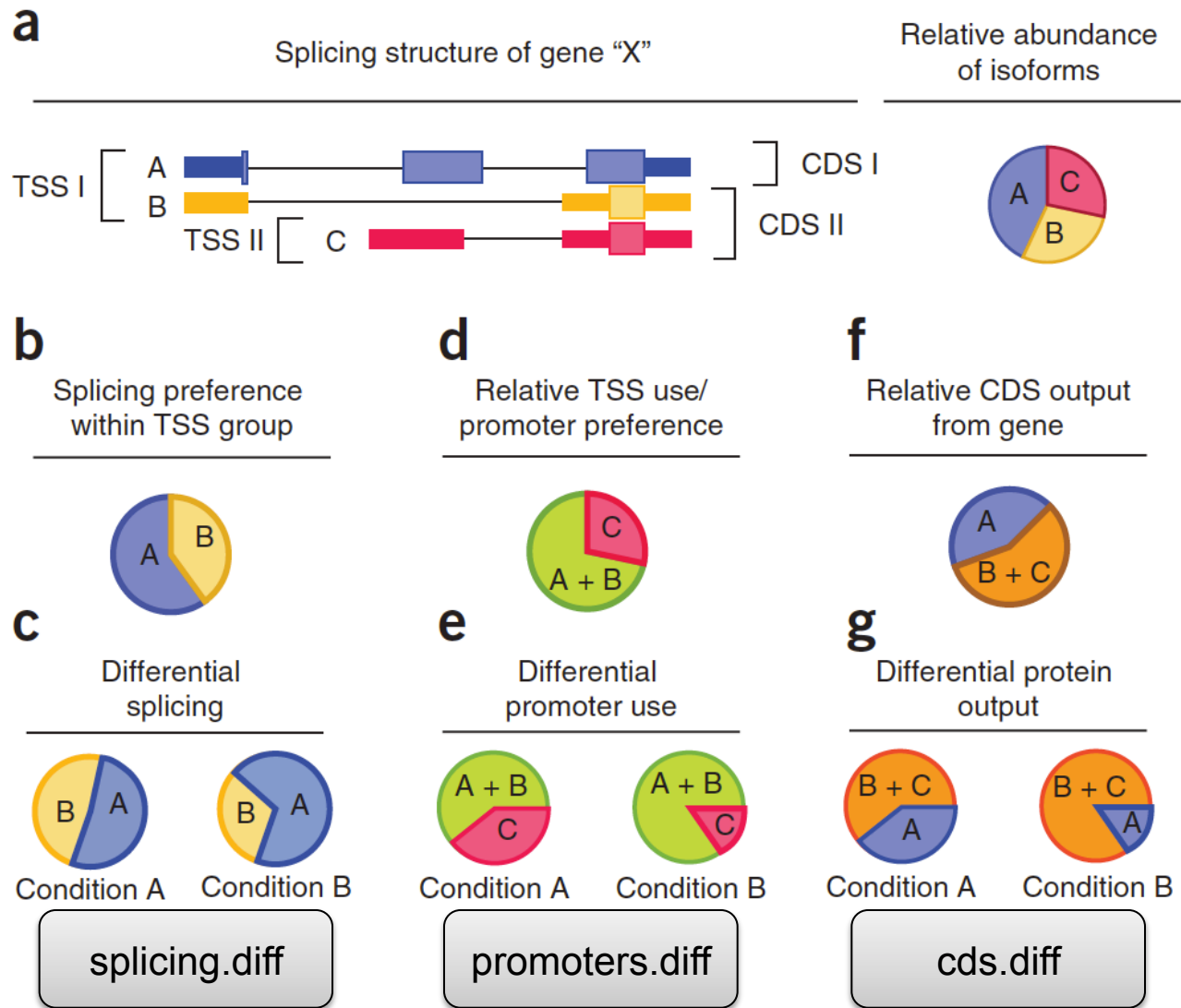
# Sequencing methods for studying alternative isoforms



# Cufflinks alternative splicing tests

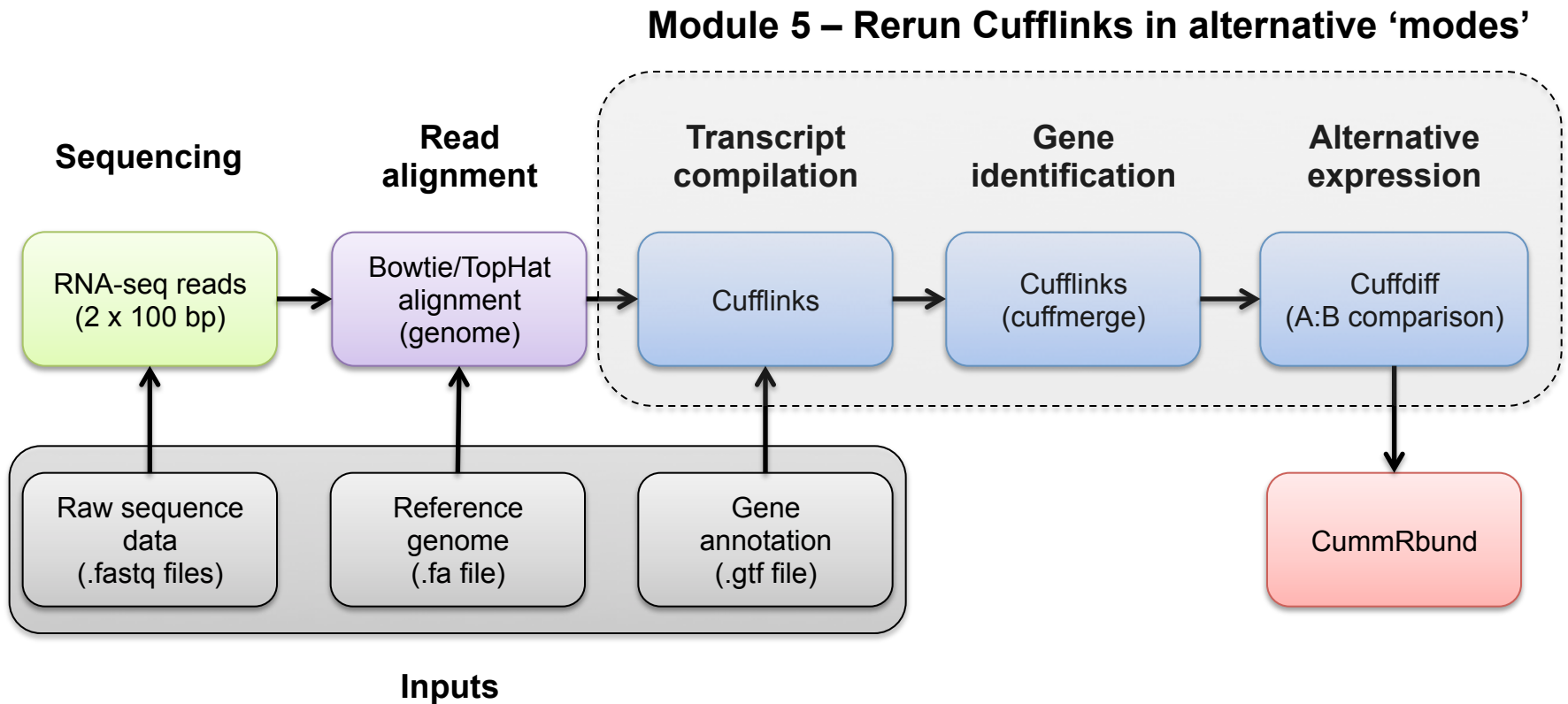


# Cufflinks alternative splicing tests



# **Introduction to tutorial (Module 5)**

# Bowtie/TopHat/Cufflinks/Cuffdiff RNA-seq Pipeline

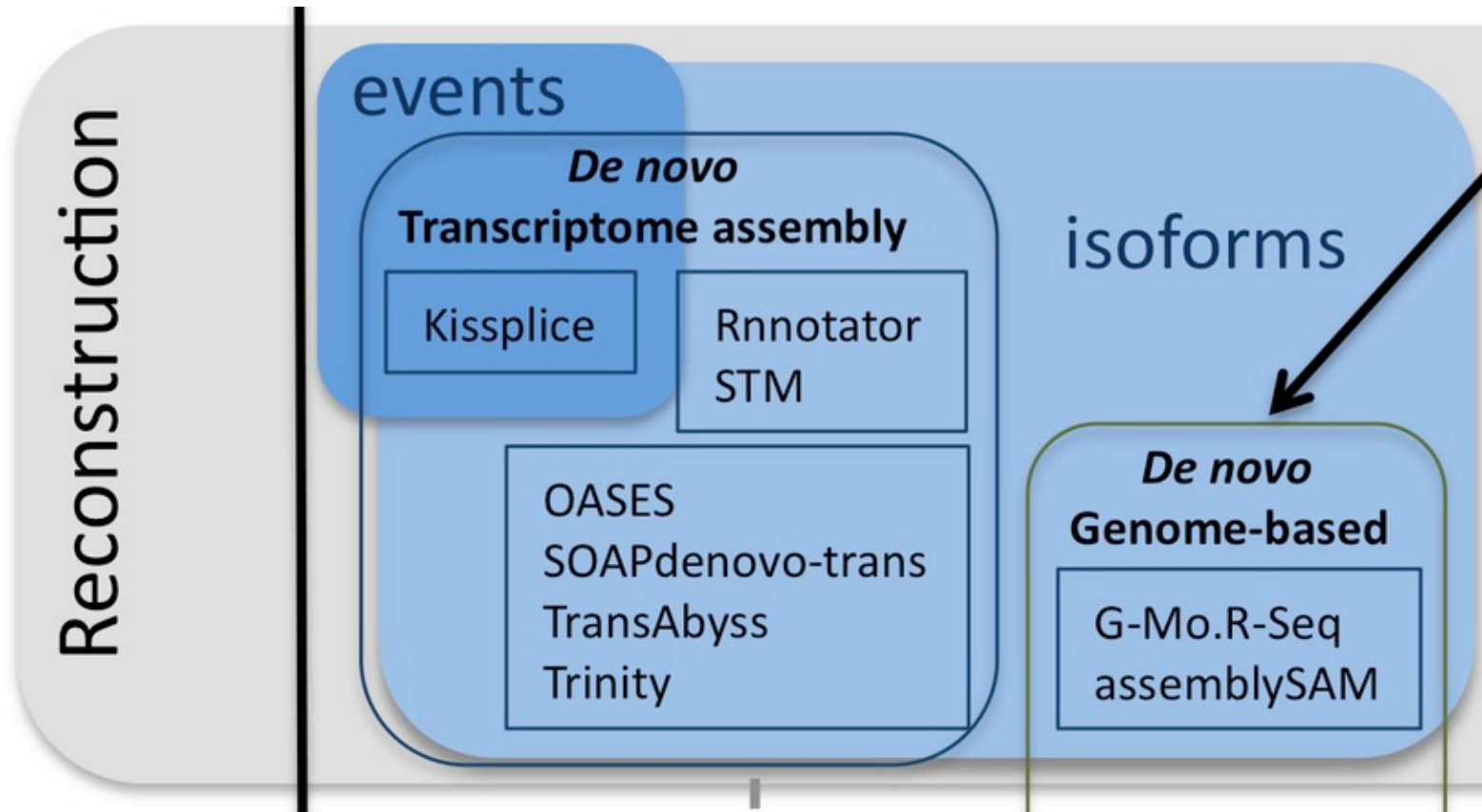




# What if I don't have a reference genome for my species?

- Have you considered sequencing the genome of your species?
- If that is not practical or you simply prefer a transcript discovery approach that does not rely on prior knowledge of the genome or transcriptome there are some tools available ...
  - Unfortunately de novo transcriptome assembly is beyond the scope of this workshop
  - The good news is that the skills you learn here will help you figure out how to install and run those tools yourself

# Methods to study splicing by RNA-seq



<http://www.rna-seqblog.com/data-analysis/splicing-junction/methods-to-study-splicing-from-rna-seq/>  
<http://arxiv.org/ftp/arxiv/papers/1304/1304.5952.pdf>

# We are on a Coffee Break & Networking Session



**bioinformatics.ca**



Cold  
Spring  
Harbor  
Laboratory

1890  
125  
2015

