Gene Expression Profiling of Colon Cancer Metastatic Cells by RNA-Sequencing

Students: Aleksey Aleev
Aleksandr Khudiakov
Margarita Akseshina
Igor Evsyukov

Scientific advisors: Olga Bajenova
Simonov S., Lapidus A., Komissarov A., Makunin A., Tamazian G.

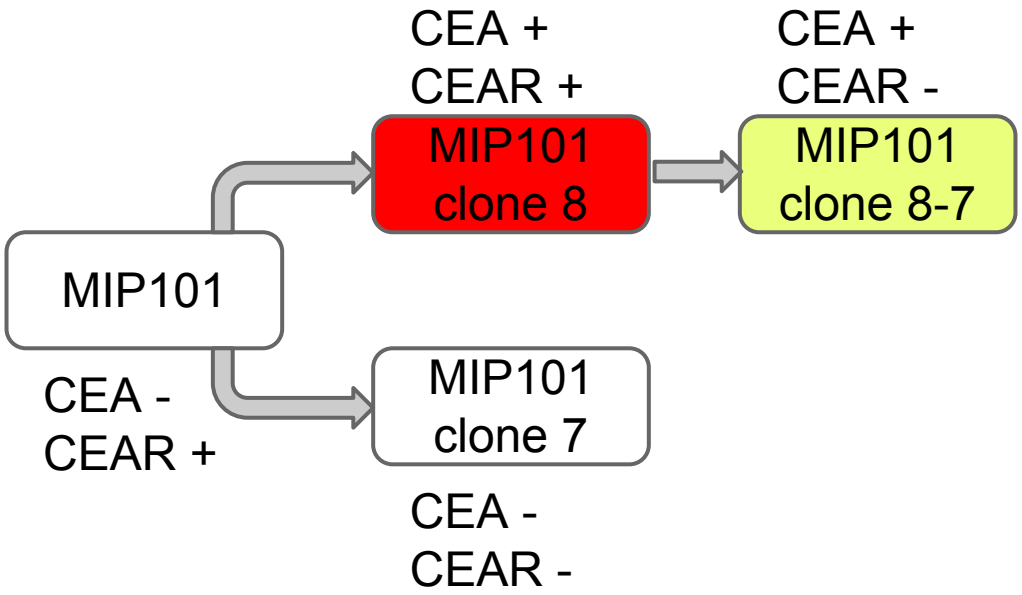
- The project is connected with cancer research
    - One of the main problems in cancer research is metastasis

The aim of the project

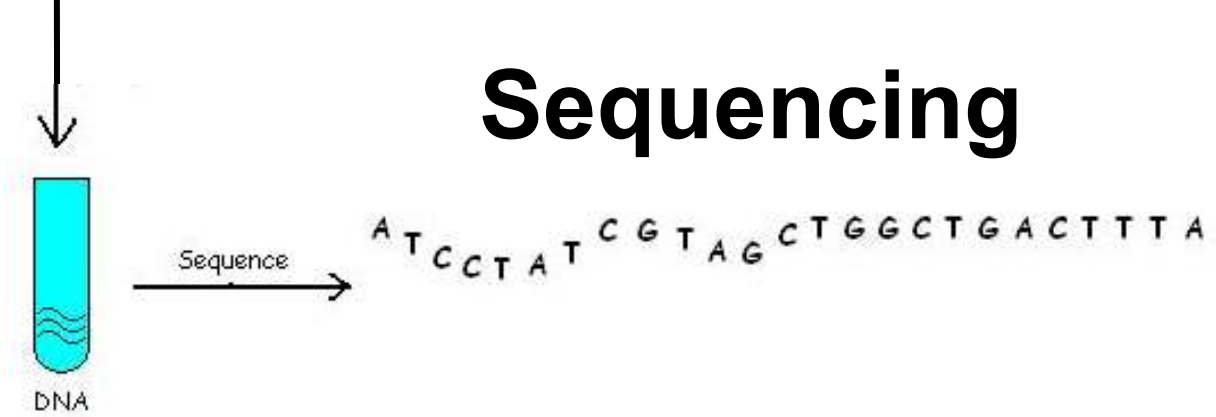
is a comparative analysis of the transcriptome of colorectal cancer cells with varying ability to form metastases for the purpose of determination of genes and signaling pathways involved in the formation of metastases

What was already done

1. 4 cell lines with different expression level of CEA and CEAR
2. RNA extraction and cDNA library preparing



Sequencing



MiSeq

illumina®

Project goals

1. Sequence cDNA library and perform quality control of the reads
2. Perform differential expression analysis of given 4 cell lines
3. Compare results of performed analysis to find genes expressed in high-metastatic cell lines
4. Identify global signaling pathways and gene clusters which are probably responsible for metastatic process
5. Compare different pipelines and corresponding tools for differential expression analysis

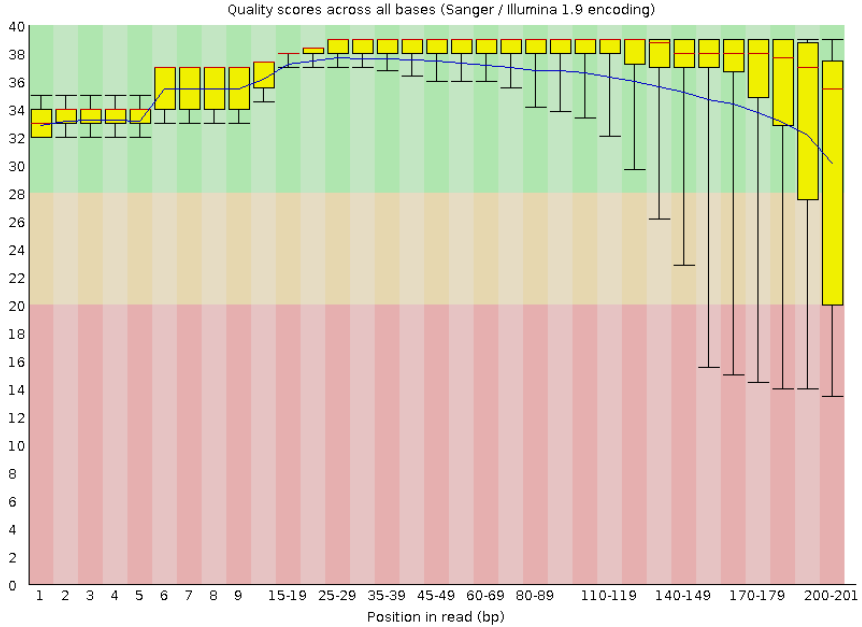
Initial analysis and correction

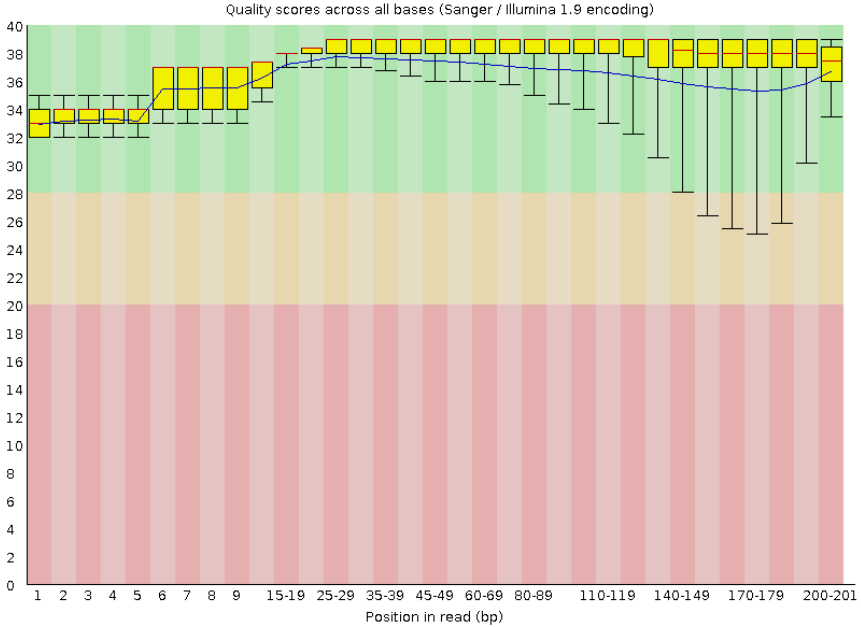
- Throwing out reads with Illumina adapters
  is done with the help of A.Komissarov.
  95% reads remained for the further work

- Cutting off low quality read tails

Quality control of the reads

Quality before and after cleaning:

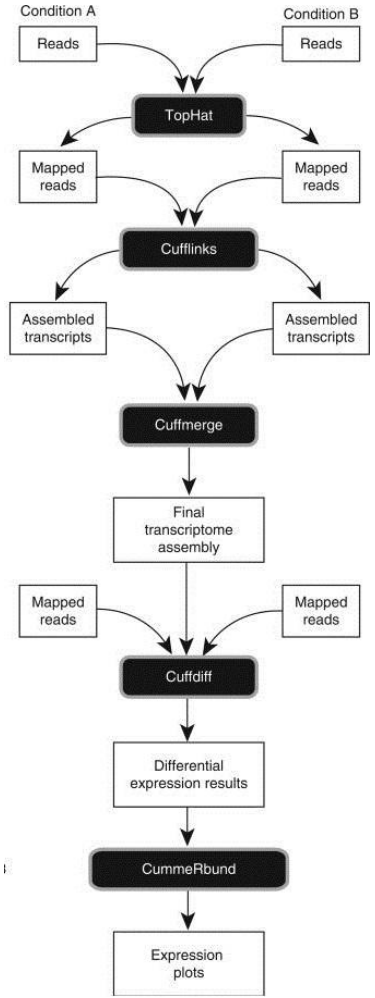




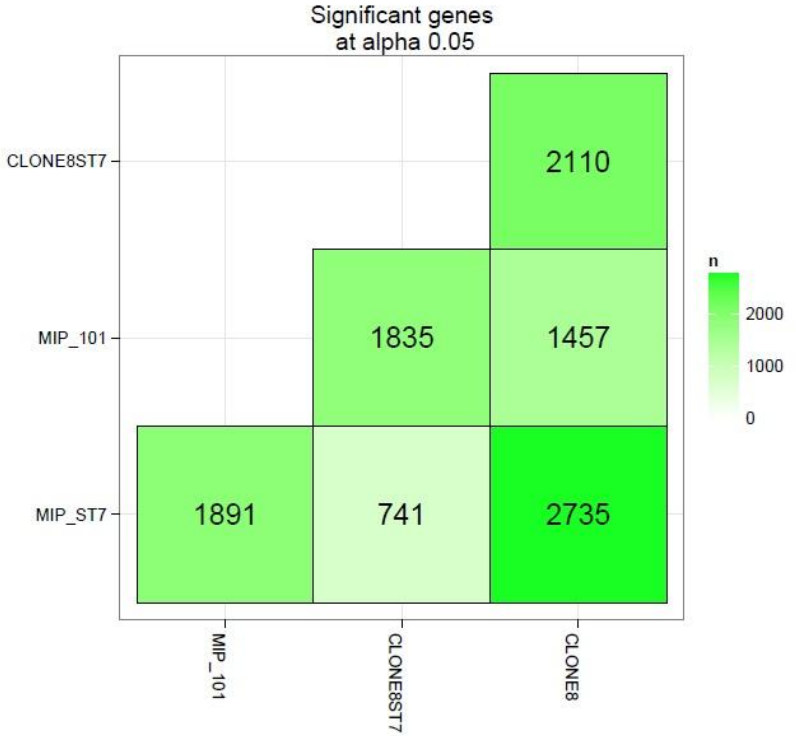
TopHat + Cufflinks

Required:

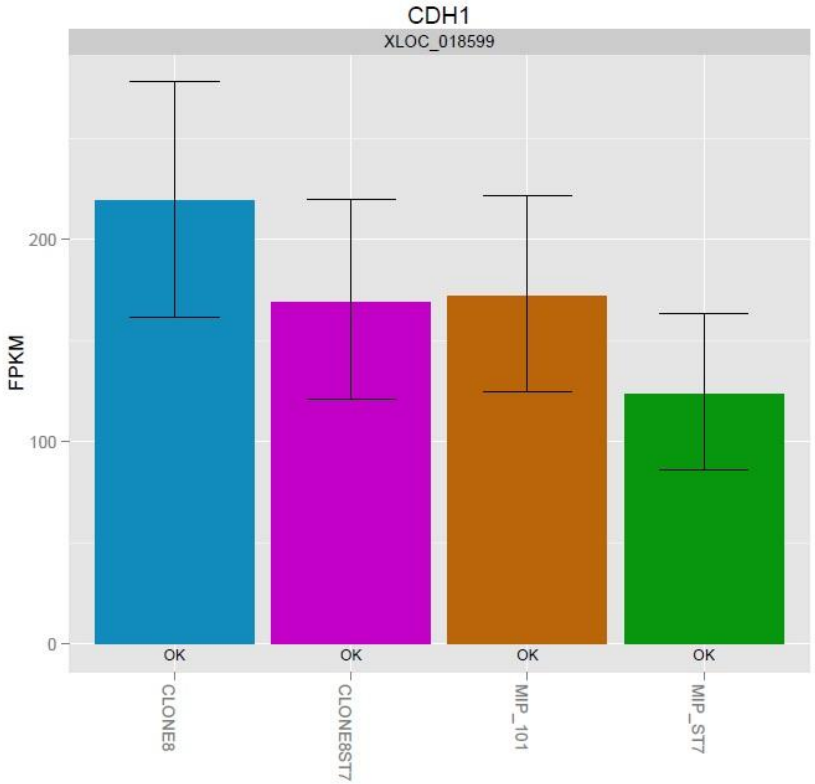
- tools (Samtools, Bowtie, TopHat, Cufflinks, R, CummerBund)
- raw reads
- reference human genome
- genes annotation for that reference

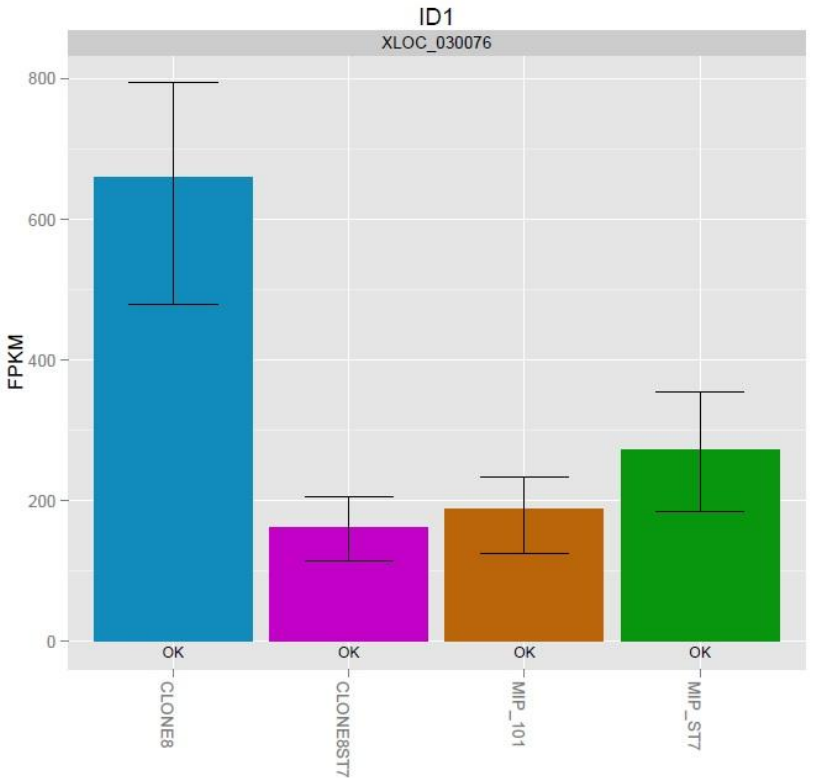


Data exploration

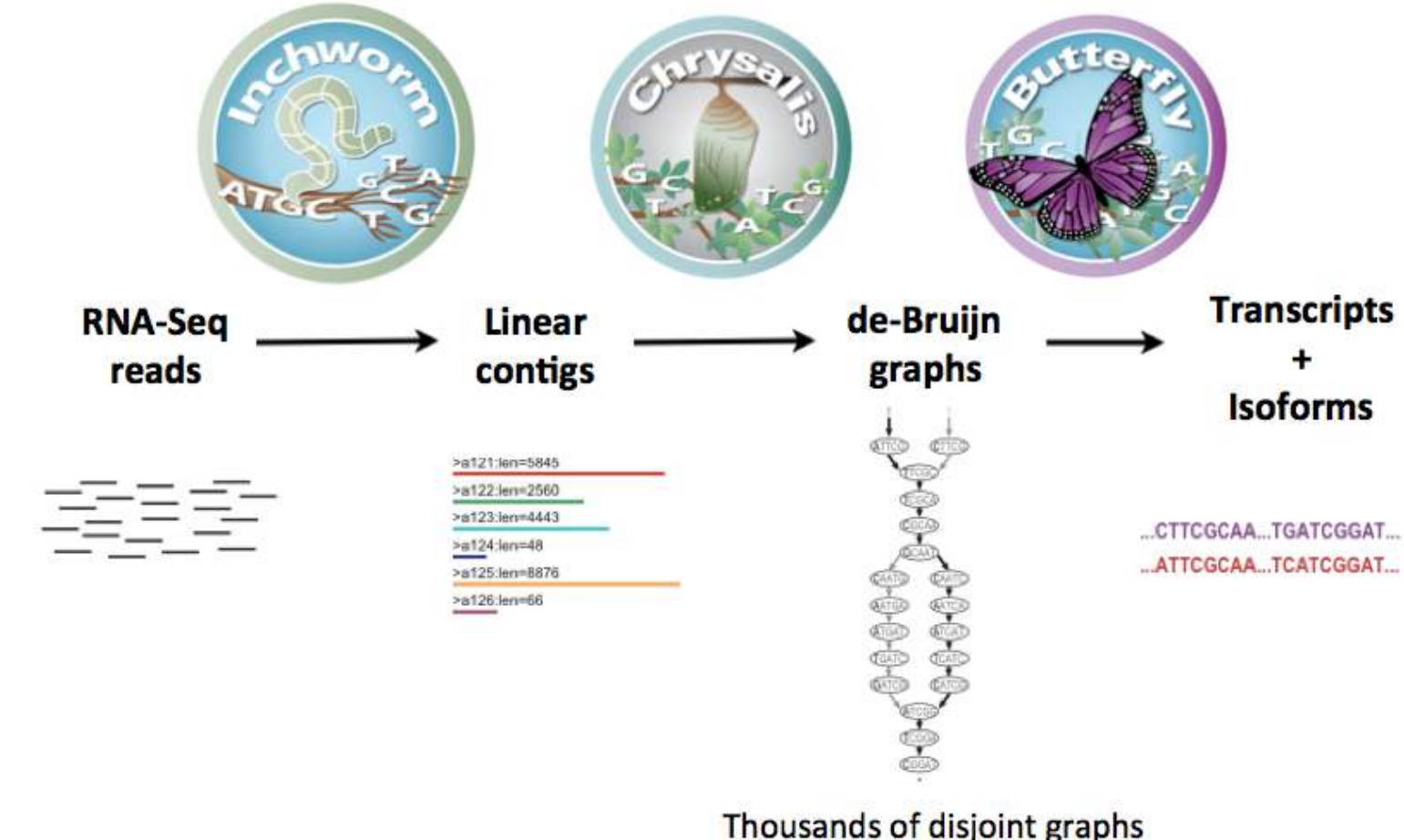


Genes analysis





Trinity pipeline: de novo transcriptome assembly



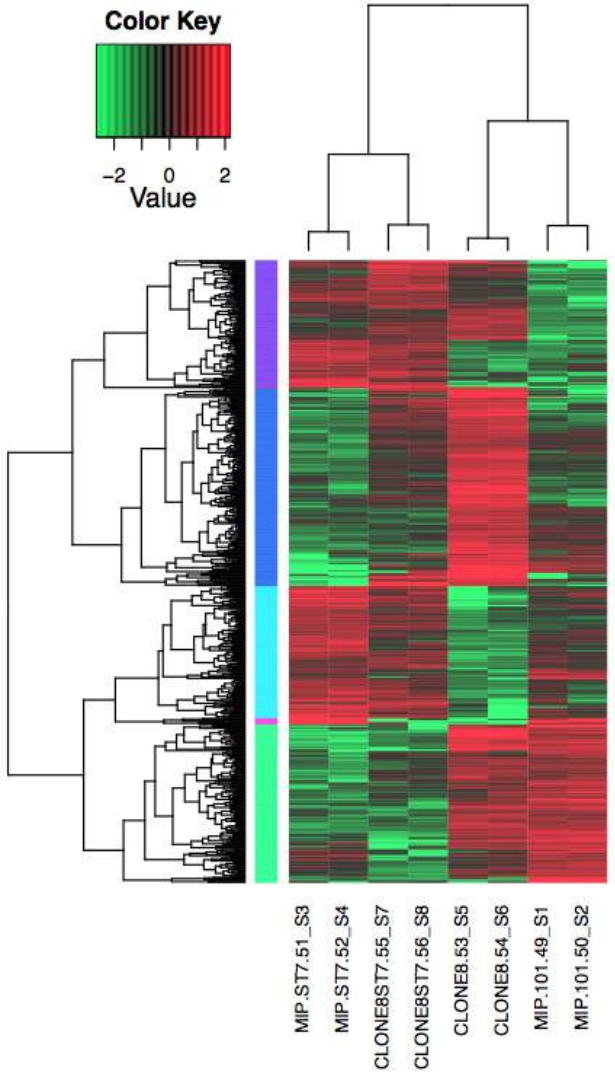
Trinity quality control

- 96.5% of reads were aligned to assembled transcriptome with their pairs
- 2.51% - only left reads were aligned
- 0.84% - only right reads were aligned

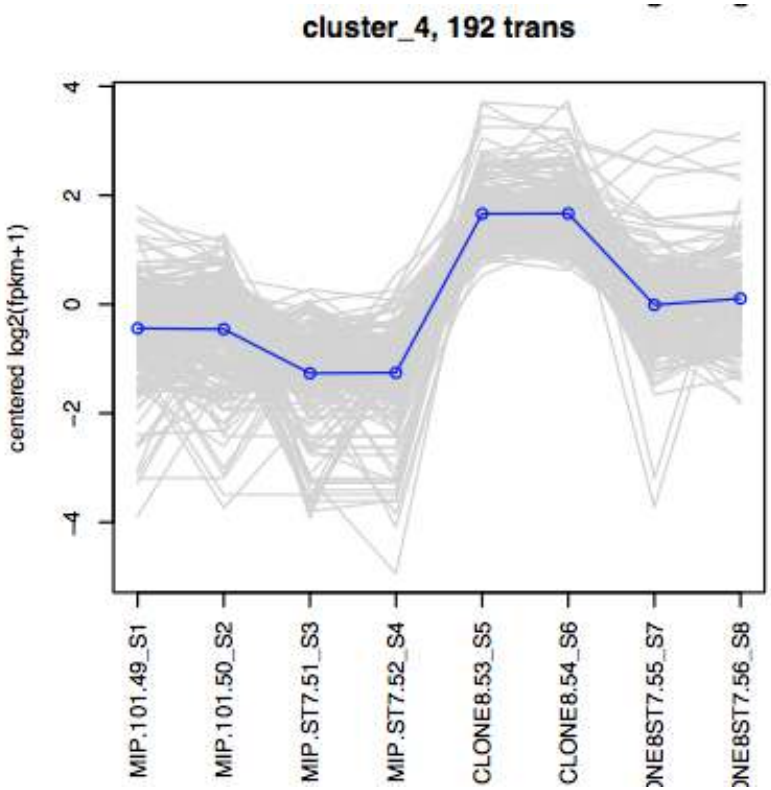
Trinity pipeline: differential expression analysis

- Abundance estimation using RSEM
- Identifying differentially expressed transcripts using EdgeR Bioconductor package
- Perform functional annotation of transcriptomes using Trinotate

Genes expression heatmap



Possible CEA-connected transcripts

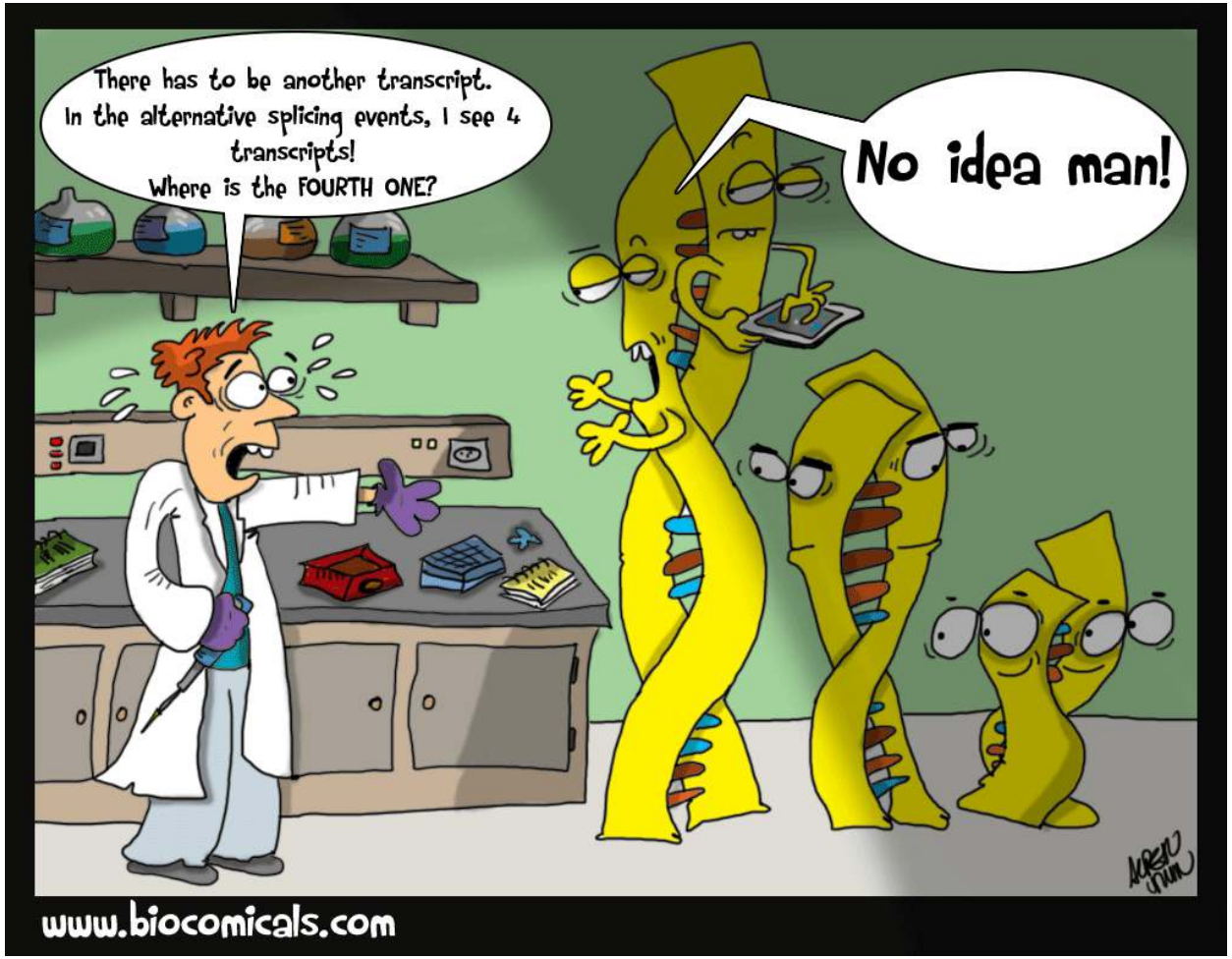


Tophat+Cufflinks vs. Trinity

- Several genes' expression levels calculated via different pipelines were compared
  - The tendency of expression levels is the same for both pipelines
- Tools' outputs provide similar capabilities for further analysis

Project results

- ✓ cDNA library was sequenced and quality control of the reads was performed
- ✓ Differential expression analysis of given 4 cell lines was performed
- ✓ Results of performed analysis were compared, some genes expressed in high-metastatic cell lines were found
- ✓ De novo transcriptome assembly pipeline with corresponding tools for differential expression analysis was compared with reference-assisted transcriptome assembly pipeline


There has to be another transcript. In the alternative splicing events, I see 4 transcripts! Where is the FOURTH ONE?
No idea man!
www.biocomicals.com