LongTREC

Funded by
the European Union

Bioinformatics Summer School

Long-reads Transcriptomics

Ana Conesa & Fabian Jetzinger

I^{2}SysBio, Valencia, Spain
BioBam Bioinformatics, Valencia, Spain

LongTREC - The Long-reads TRanscriptome European Consortium
Marie Skłodowska-Curie grant agreement No 101072892

Course Contents

1. Biological replicates and transcriptome reconstruction
2. Differential Expression vs. Differential Isoform Usage
3. tappAS

Section 1

Biological replicates and transcriptome reconstruction

Understanding the biological replicates
influence transcriptome reconstruction

Biological replicates capture natural variation and improve reproducibility

- Distinguish general patterns from sample-specific noise
- Detect and discard outliers, low-quality samples
- Increase power of statistical analyses
- Enable a wider range of tools

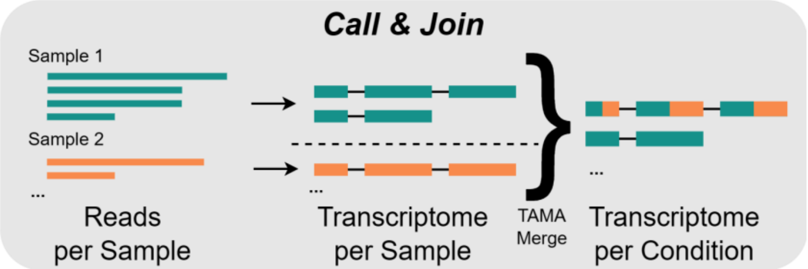
Combination of replicates influences transcriptome reconstruction

Join & Call

- Combine all reads → transcriptome reconstruction
- Detect lowly-expressed transcript variants

Call & Join

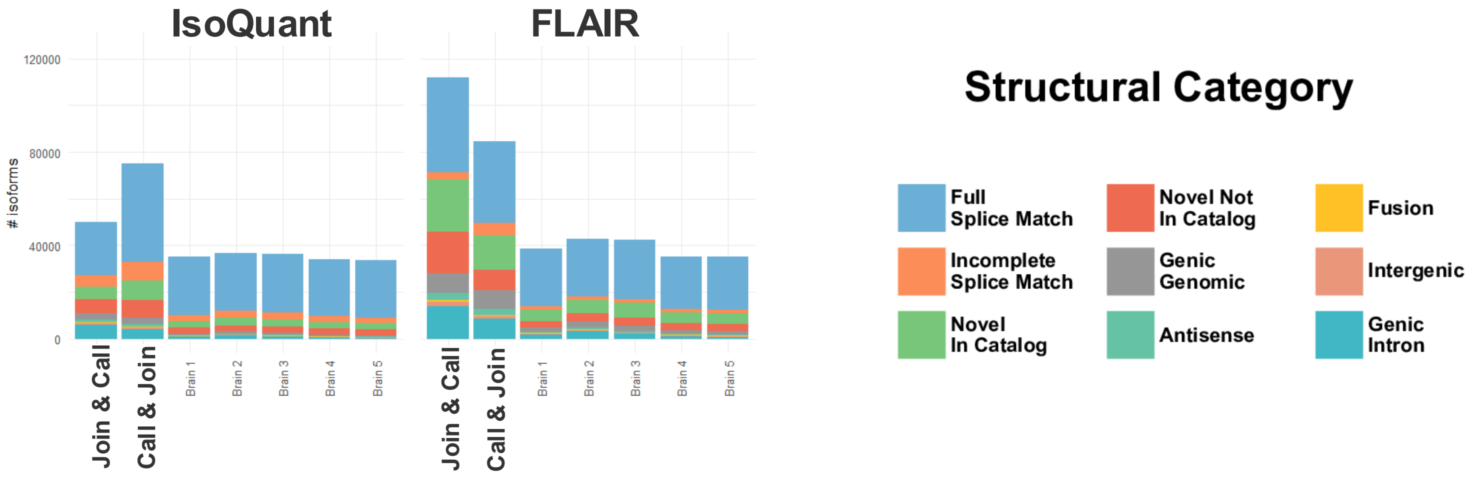
- Transcriptome reconstruction on each sample → combine annotations (e.g. TAMA Merge)
- Detect sample-specific transcript variants



Results of both strategies vary across transcriptome reconstruction tools!

Transcriptome diversity varies based on strategy and reconstruction tool

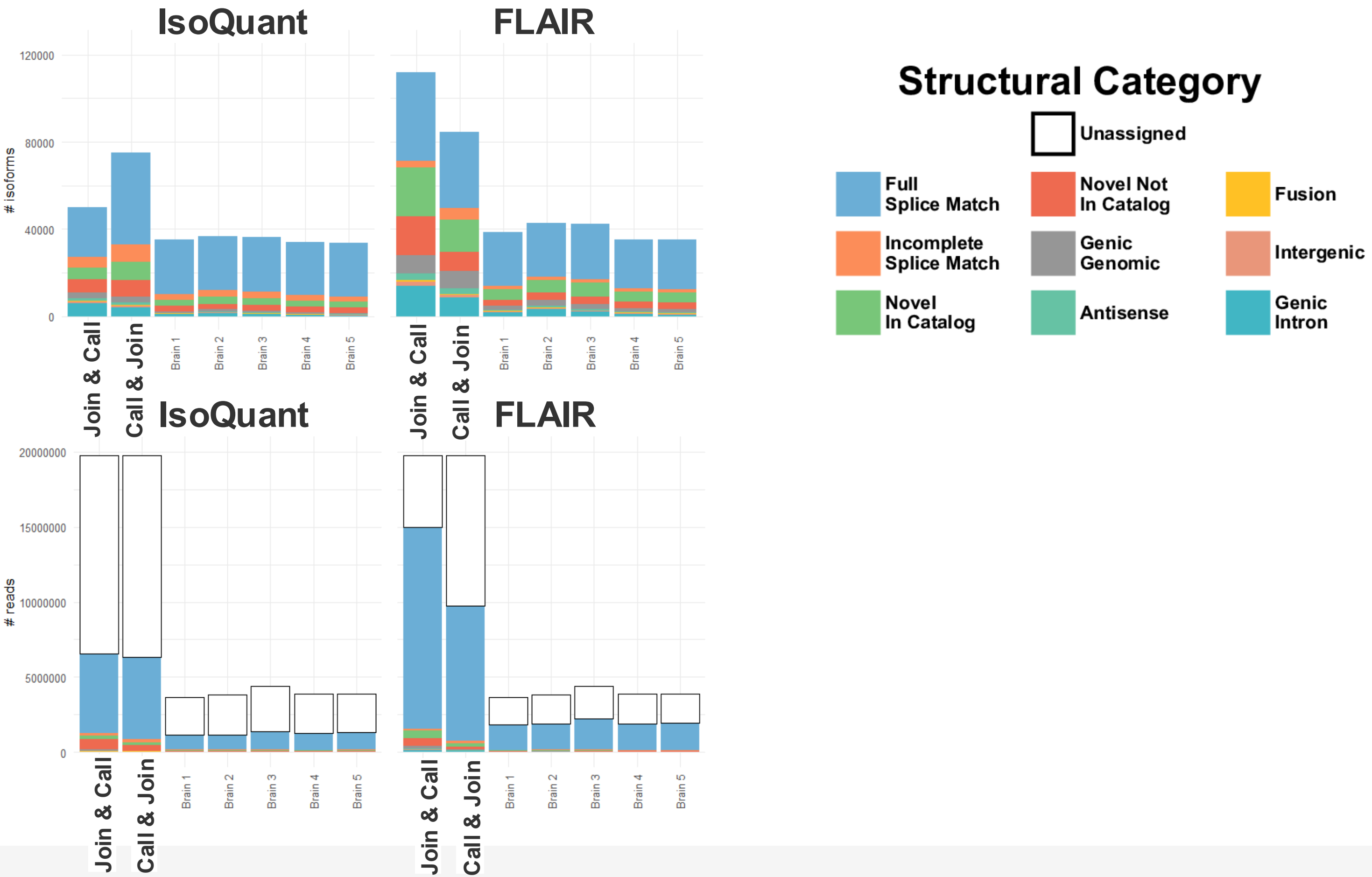
Transcriptome diversity



Transcriptome diversity varies based on strategy and reconstruction tool

Transcriptome diversity

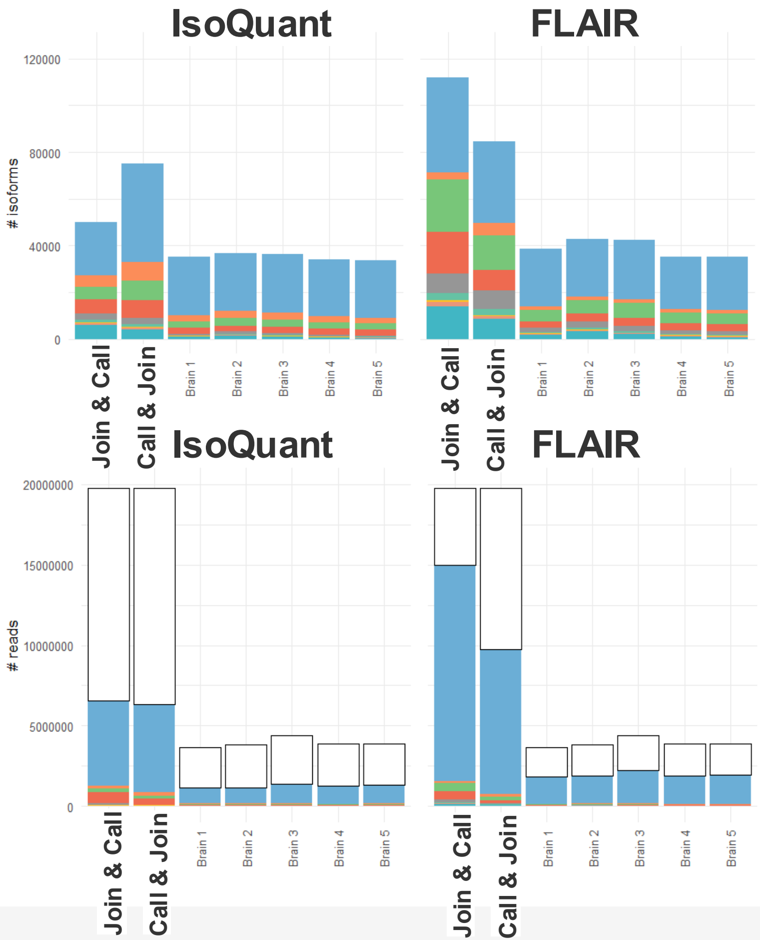
Assignment of reads



Transcriptome diversity varies based on strategy and reconstruction tool

Transcriptome diversity

Assignment of reads



Structural Category

- ☐ Unassigned
- 🟦 Full Splice Match
- 🟧 Incomplete Splice Match
- 🟩 Novel In Catalog
- 🟥 Novel Not In Catalog
- ⬛ Genic Genomic
- 🟩 Antisense
- 🟨 Fusion
- 🟧 Intergenic
- 🟦 Genic Intron

Takeaways:

Performance of strategy depends on reconstruction tool

Majority of reads assigned to FSM (reference, known) transcripts

Consider choice of strategy and tool carefully based on research objectives

Section 2

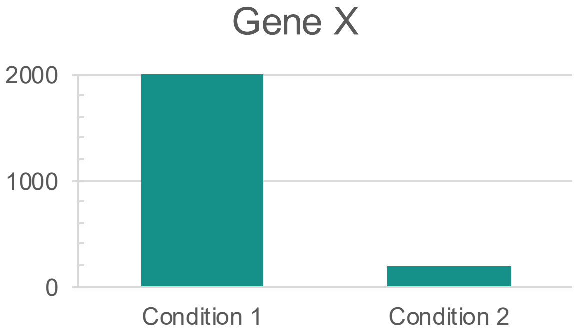
Differential Expression vs. Differential Isoform Usage

Understanding differential expression at isoform resolution

Differential Expression measures total expression levels

Differential Expression

- **Gene- or transcript-level**
- Compares **total expression** across conditions

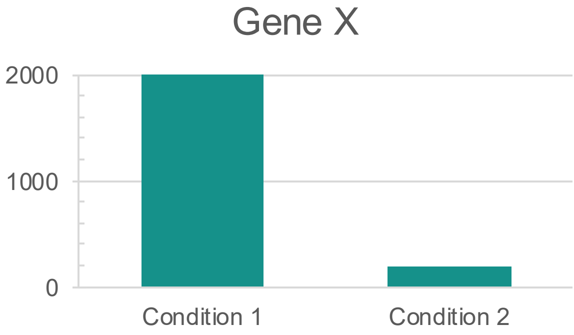


✓ **DGE**

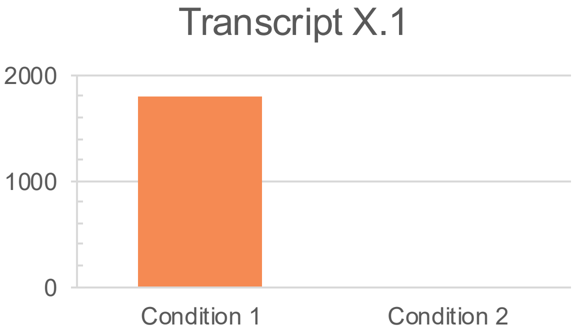
Differential Expression measures total expression levels

Differential Expression

- **Gene- or transcript-level**
- Compares **total expression** across conditions



✓ **DGE**

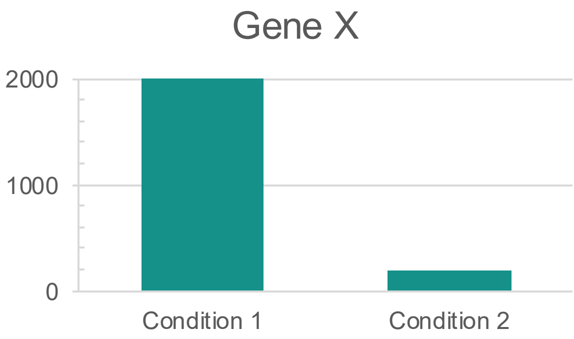


✓ **DTE**

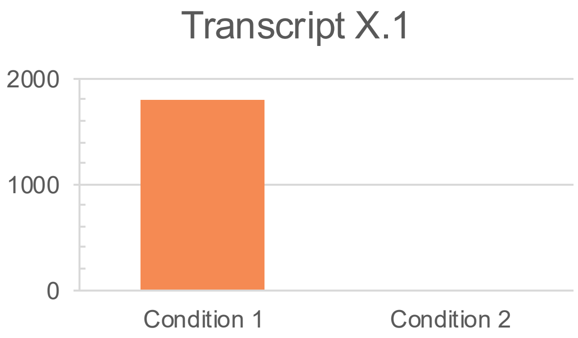
Differential Expression measures total expression levels

Differential Expression

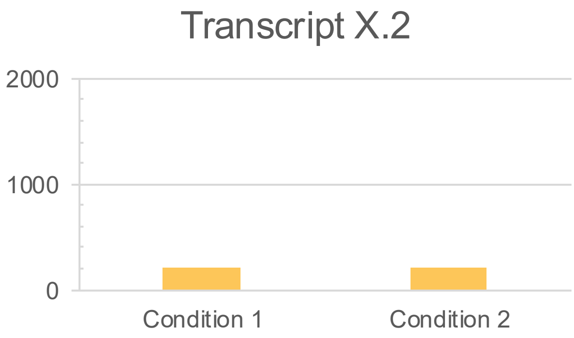
- **Gene- or transcript-level**
- Compares **total expression** across conditions



✓ **DGE**



✓ **DTE**



✗ **DTE**

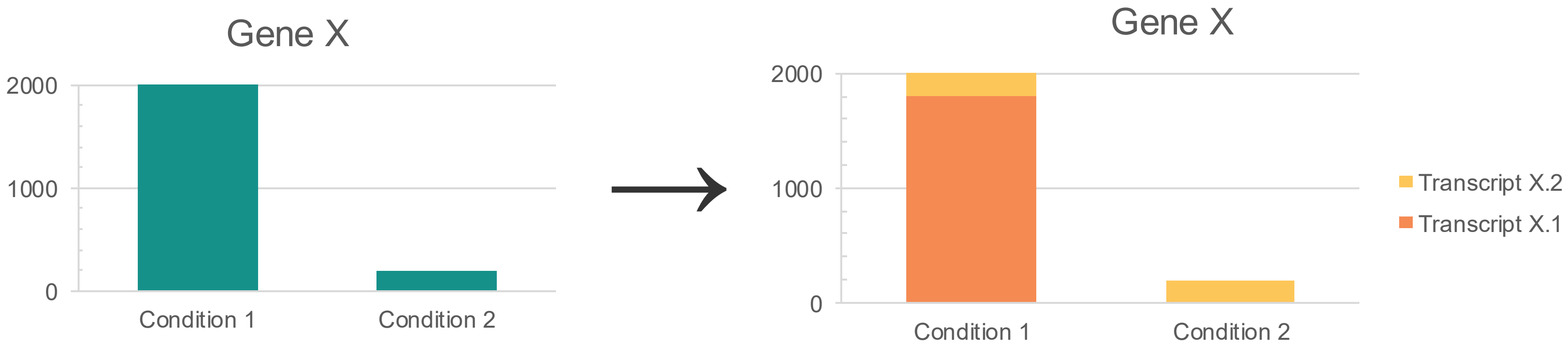
Differential Expression measures total expression levels

Differential Expression

- **Gene- or transcript-level**
- Compares **total expression** across conditions

Differential Isoform Usage

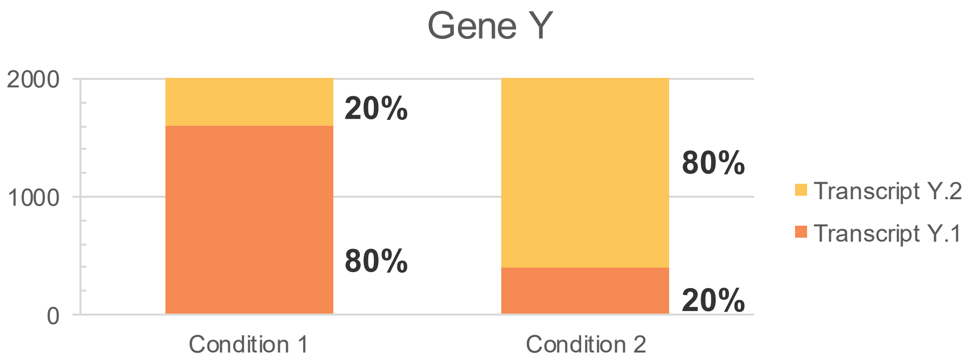
- Compares **relative expression of a gene's isoforms** across conditions



Differential Isoform Usage sheds light on relative differences

Differential Isoform Usage

- Compares **relative expression of a gene's isoforms** across conditions



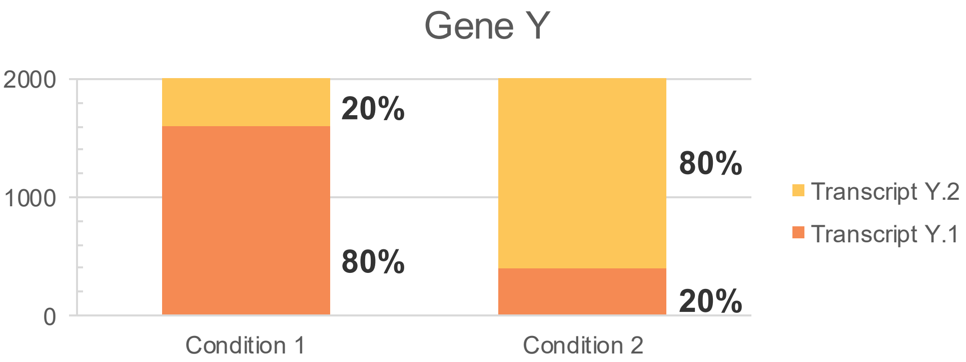
✗ **DGE**

✓ **DIU** (major isoform switch)

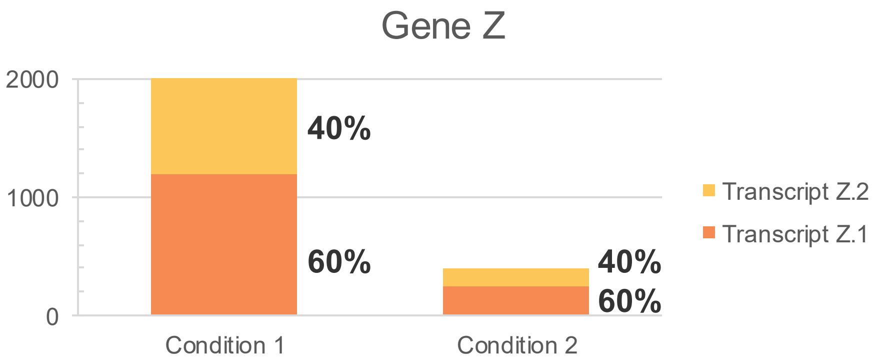
Differential Isoform Usage sheds light on relative differences

Differential Isoform Usage

- Compares **relative expression of a gene's isoforms** across conditions



Gene Y



Gene Z

✗ **DGE**

✓ **DIU** (major isoform switch)

✓ **DGE**

✗ **DIU**

Many tools for DE, DIU, and similar types of analyses exist

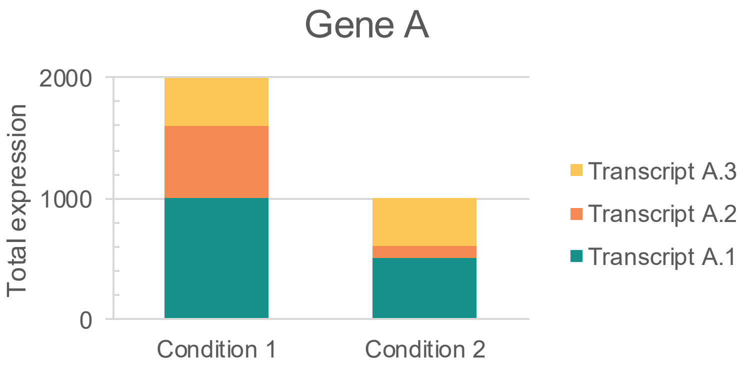
- Other, similar types of analyses exist, e.g. Differential Exon Usage
- Popular tools: edgeR, DESeq2, DEXSeq, NOISeq (can simulate replicates), DRIMSeq, etc.
- Traditional tools (edgeR, DESeq2, etc.) have been shown to also work well in long reads

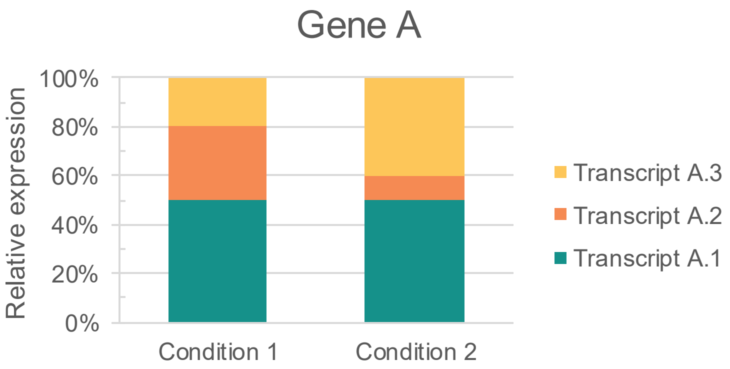
See:
Dong, Xueyi et al. "Benchmarking long-read RNA-sequencing analysis tools using in silico mixtures." Nature methods vol. 20,11 (2023): 1810-1821. doi:10.1038/s41592-023-02026-3

Reflection

How would you describe the following situation of Gene A and its 3 isoforms?

- Differential Gene Expression?
- Differential Transcript Expression?
- Differential Transcript Usage?
- Major Isoform Switch?





	Tr. A.1	Tr. A.2	Tr. A.3
Condition 1	1000	600	400
Condition 2	500	100	400

Section 3

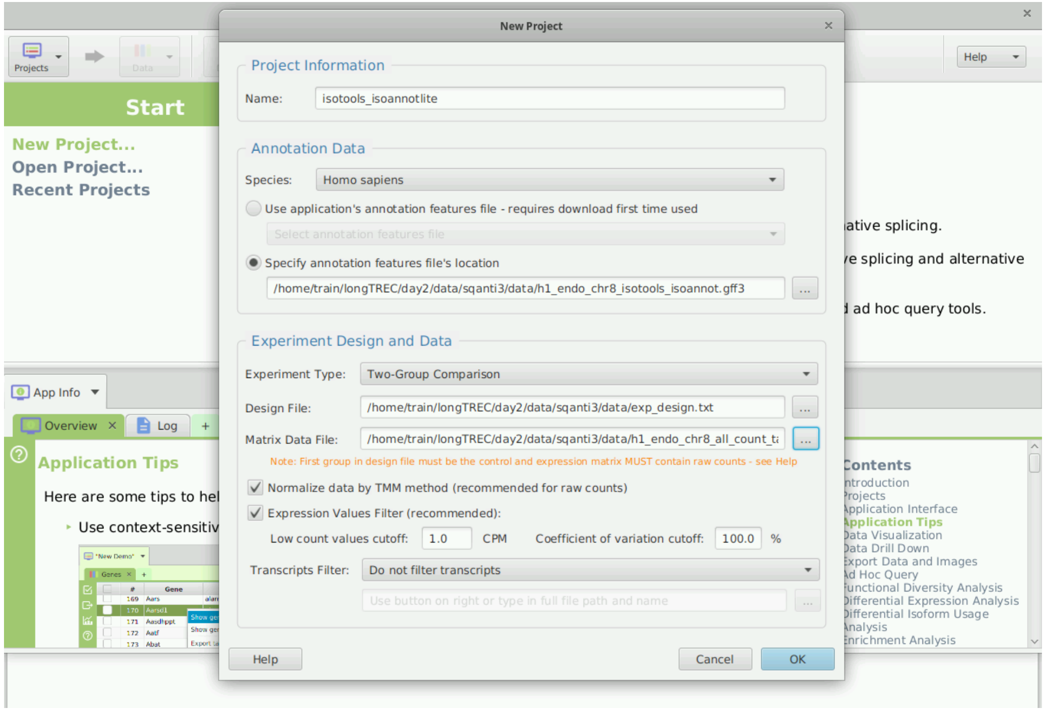
tappAS

Applying differential expression and usage analyses
and much more...

tappAS enables expression and functional analyses and visualization

- Gene-, transcript- or CDS-level differential expression
- Transcript- or CDS-level differential usage
- Gene Set Enrichment Analysis (GSEA)
- Visualization of functional annotation

tappAS project configuration



New Project

Project Information

Name: isotools_isoannotlite

Annotation Data

Species: Homo sapiens

○ Use application's annotation features file - requires download first time used

Select annotation features file

● Specify annotation features file's location

/home/train/longTREC/day2/data/sqanti3/data/h1_endo_chr8_isotools_isoannot.gff3

Experiment Design and Data

Experiment Type: Two-Group Comparison

Design File: /home/train/longTREC/day2/data/sqanti3/data/exp_design.txt

Matrix Data File: /home/train/longTREC/day2/data/sqanti3/data/h1_endo_chr8_all_count_t

Note: First group in design file must be the control and expression matrix MUST contain raw counts - see Help

☑ Normalize data by TMM method (recommended for raw counts)

☑ Expression Values Filter (recommended):

Low count values cutoff: 1.0 CPM   Coefficient of variation cutoff: 100.0 %

Transcripts Filter: Do not filter transcripts

Use button on right or type in full file path and name

Help   Cancel   OK

Thank You!

For more information about the LongTREC Summer School:

https://longtrec.eu