

PSIchomics

Shiny application for the integrated analysis of alternative splicing from large transcriptomic datasets

Nuno Agostinho
6 Dec. 2016

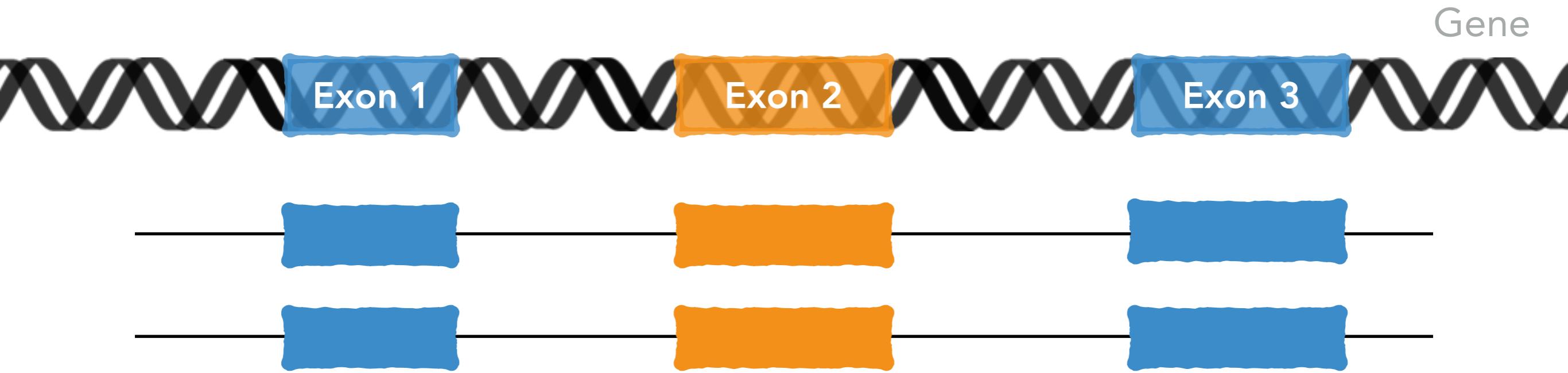
Nuno Morais
laboratory

EuroBioC
2016



Instituto
de Medicina
Molecular

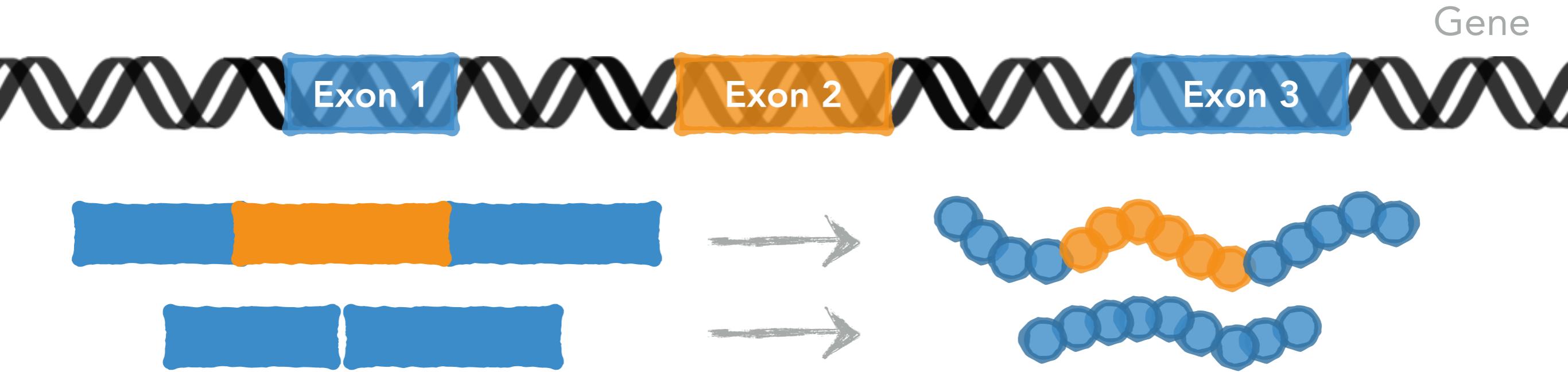
Alternative Splicing



Alternative Splicing

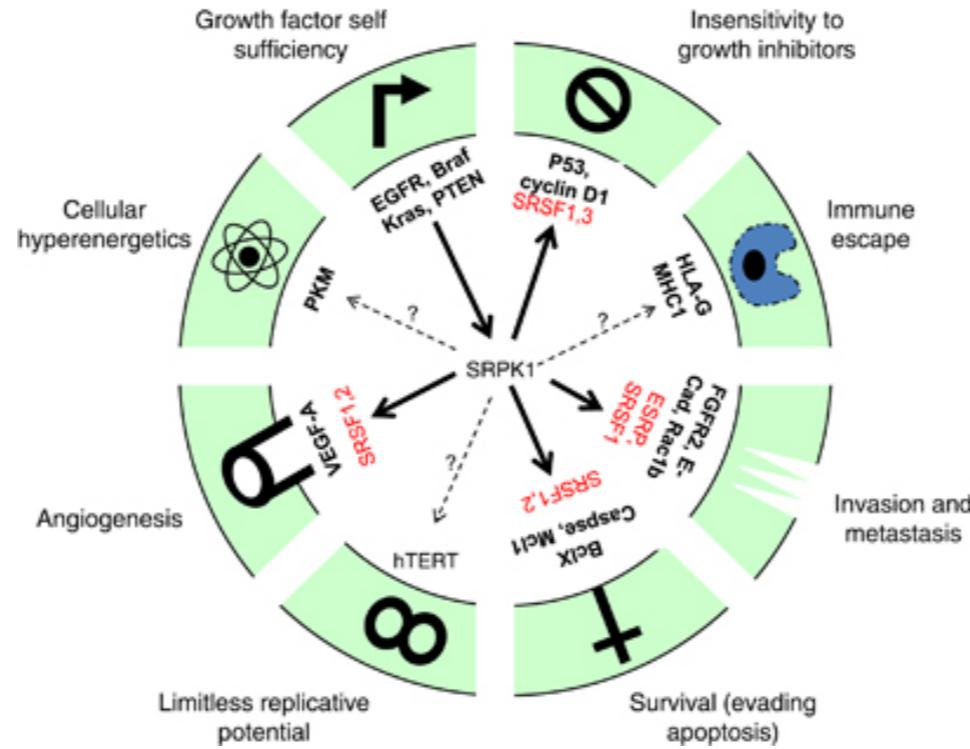


Alternative Splicing



- Occurs in 95% of human multi-exon genes (Pan et al., 2009)
- Involved in the control of many cellular processes (Oltean & Bates, 2014)

Alternative Splicing

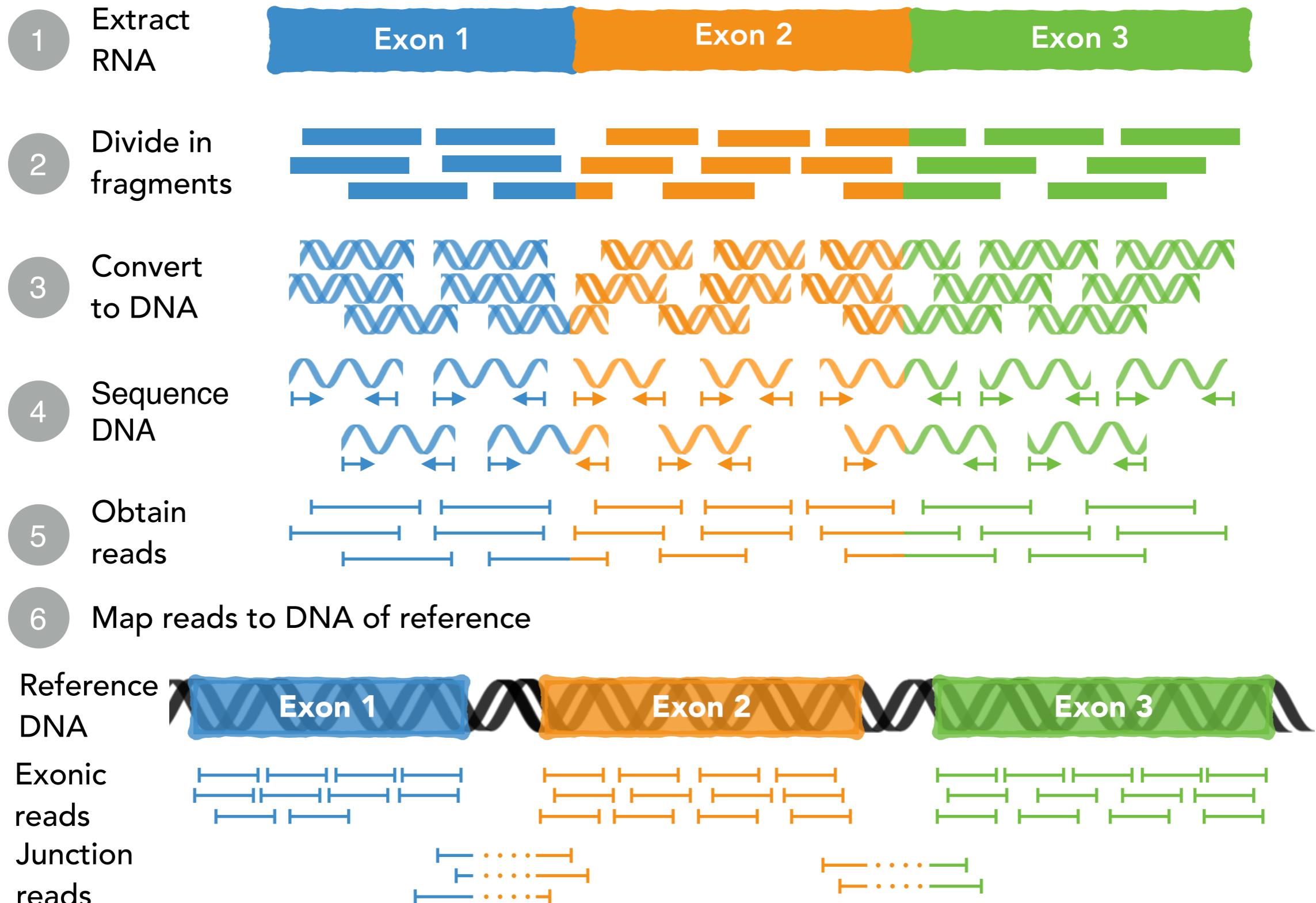


- Occurs in 95% of human multi-exon genes (Pan et al., 2009)
- Involved in the control of many cellular processes (Oltean & Bates, 2014)
- Alternative splicing deregulation is linked with cancer development (Oltean & Bates, 2014)

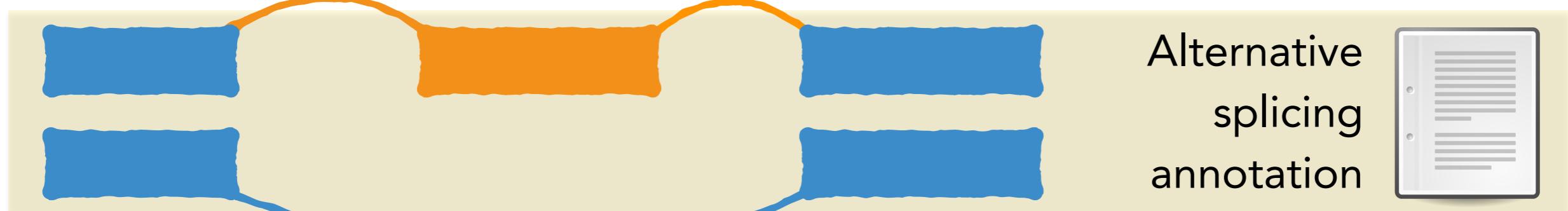
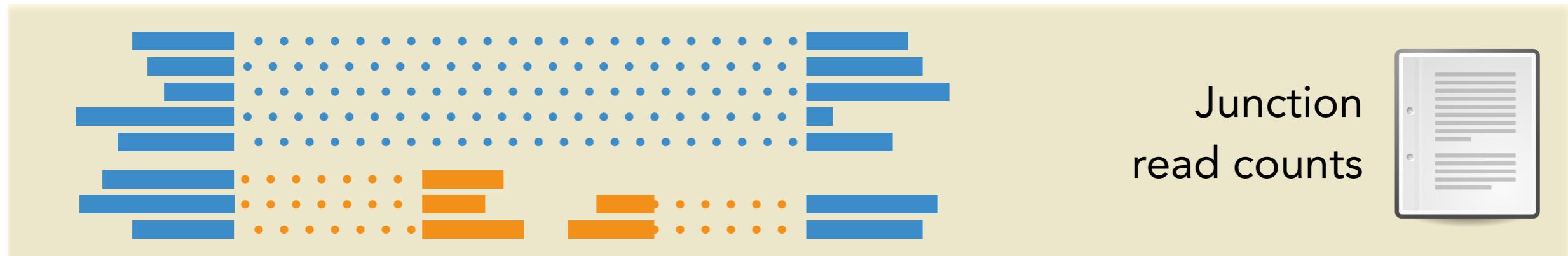


Studying alternative splicing changes may allow to identify prognostic factors and therapeutic targets

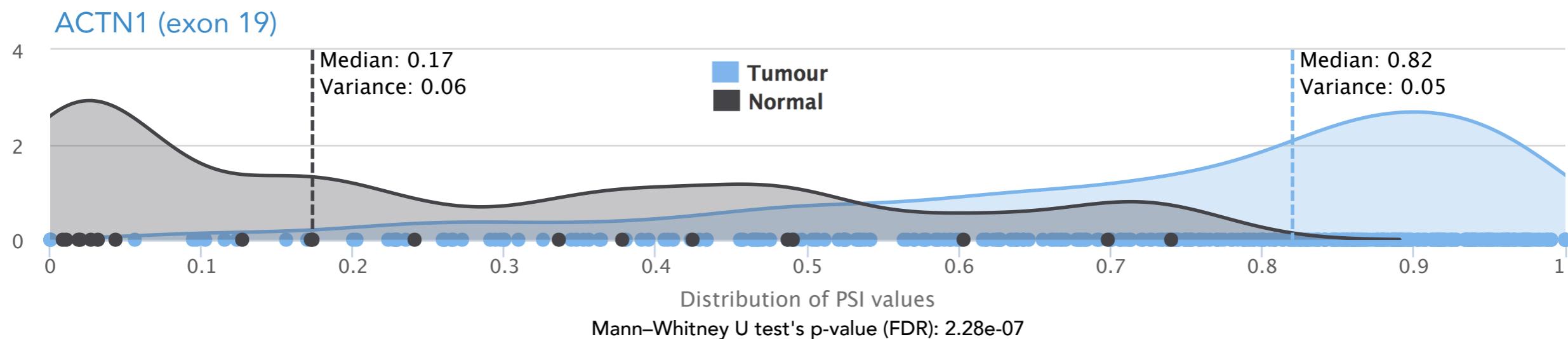
RNA Sequencing



Alternative Splicing Quantification



$$\text{Percent Spliced-In (PSI)} = \frac{\text{inclusion reads}}{\text{inclusion} + \text{exclusion reads}}$$



Quantification and Analytical Tools

- Many programs quantify, analyse and visualise alternative splicing data



Quantification and Analytical Tools

- Many programs quantify, analyse and visualise alternative splicing data
- No standard pipeline
- Their problems range from...

Over-simplistic analyses or focus in the quantification step

Time-consuming quantification of alternative splicing

No user-friendly interfaces in most tools

No incorporation of clinical information

PSIchomics

Quantify, analyse and visualise alternative splicing in cancer data

Modular architecture to easily modify and extend the program

Visual and command-line interfaces

Incorporate clinical information

Choosing as the language

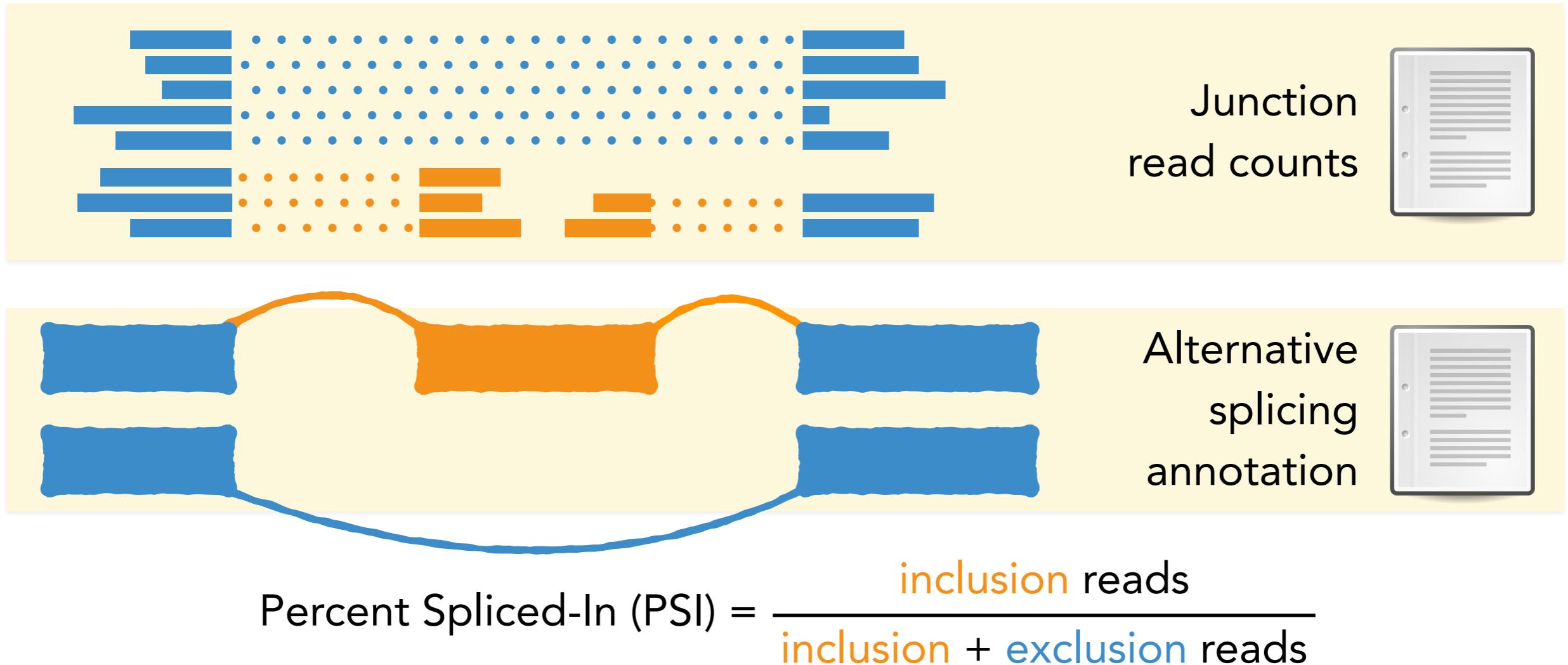
- Open-source, free and cross-platform functional language
- Ideal for statistics, data manipulation and graphical computation
- Used by the scientific community



- Web app framework using R, HTML, CSS and JavaScript
- Reactive programming model (Excel-like reactivity)

Workflow

Alternative Splicing Quantification





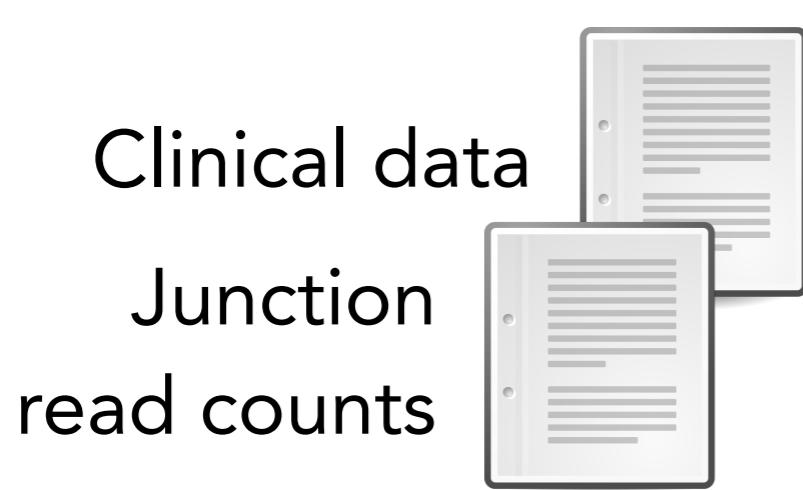
Data Retrieval



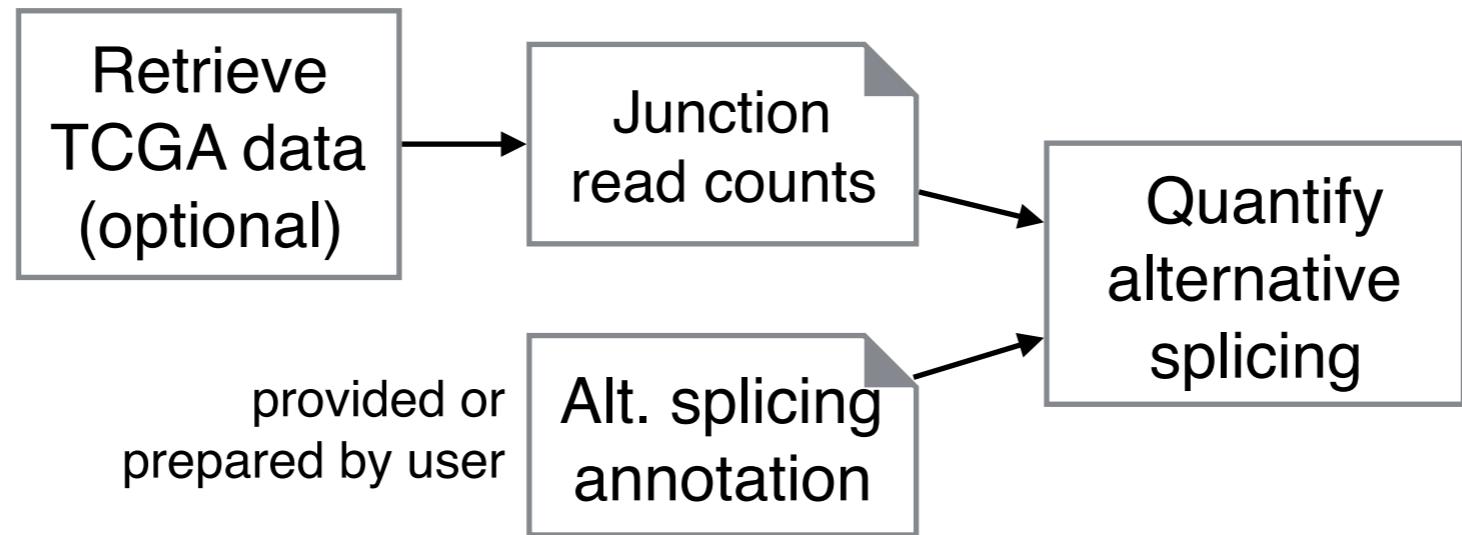
Alt. Splicing Quantification



Analyses and Visualisation



- The Cancer Genome Atlas  (data from human tumours)
 - Data is downloaded in-app using Firebrowse web API
-
- Human (hg19 assembly) annotation is available
 - Custom annotation files may be created

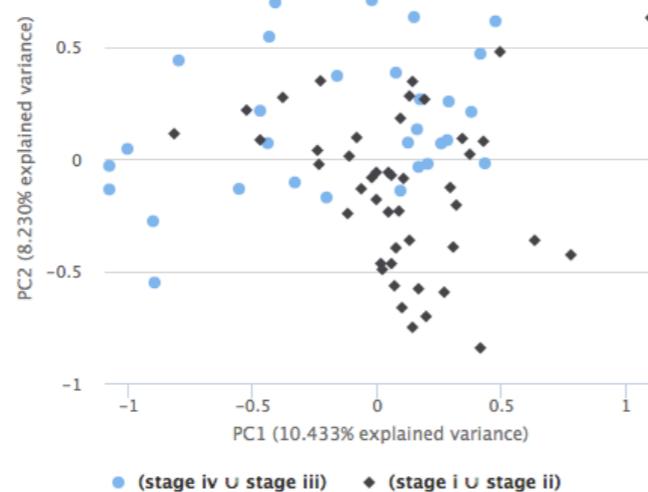


$$\text{Percent Spliced-In (PSI)} = \frac{\text{inclusion reads}}{\text{inclusion} + \text{exclusion reads}}$$

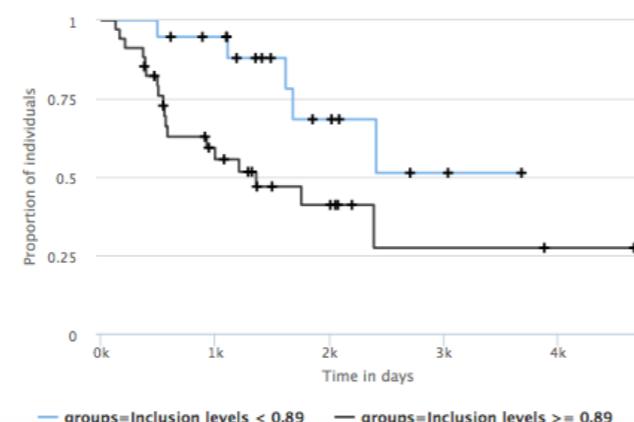


Data Retrieval

Principal component analysis

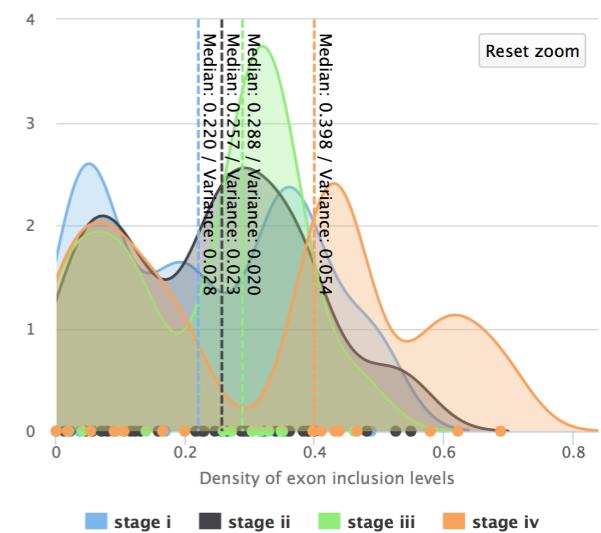


Survival analysis

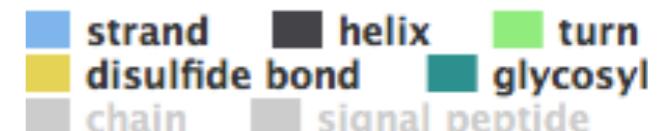
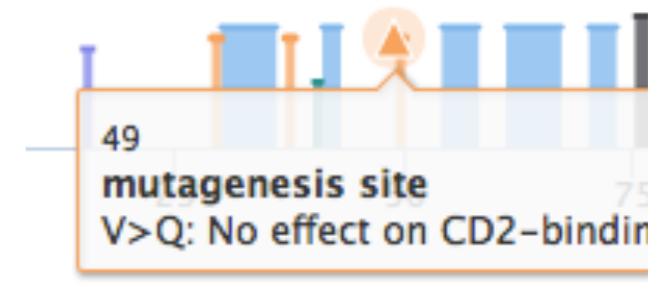


Alt. Splicing Quantification

Differential splicing analysis



Gene, RNA and protein information

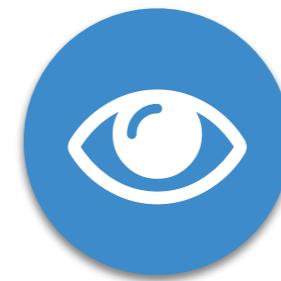




Data Retrieval

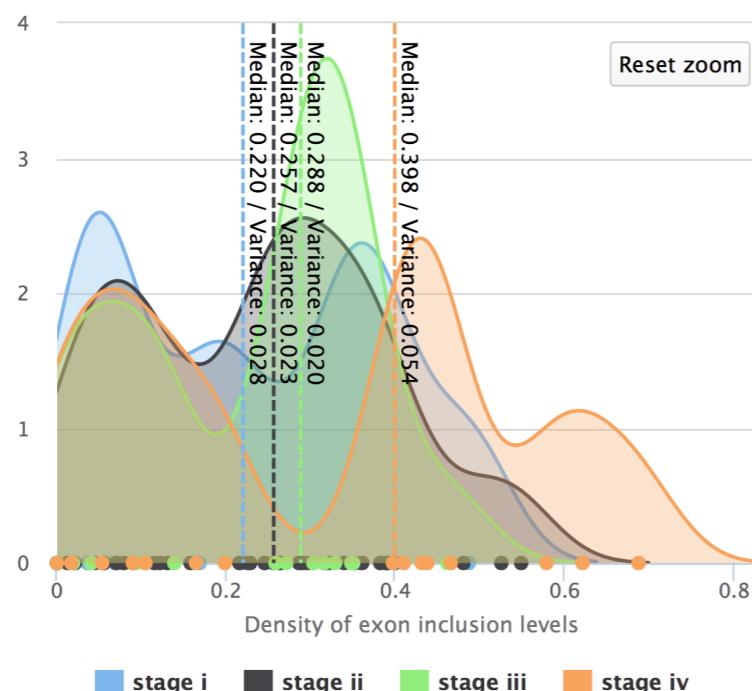


Alt. Splicing Quantification

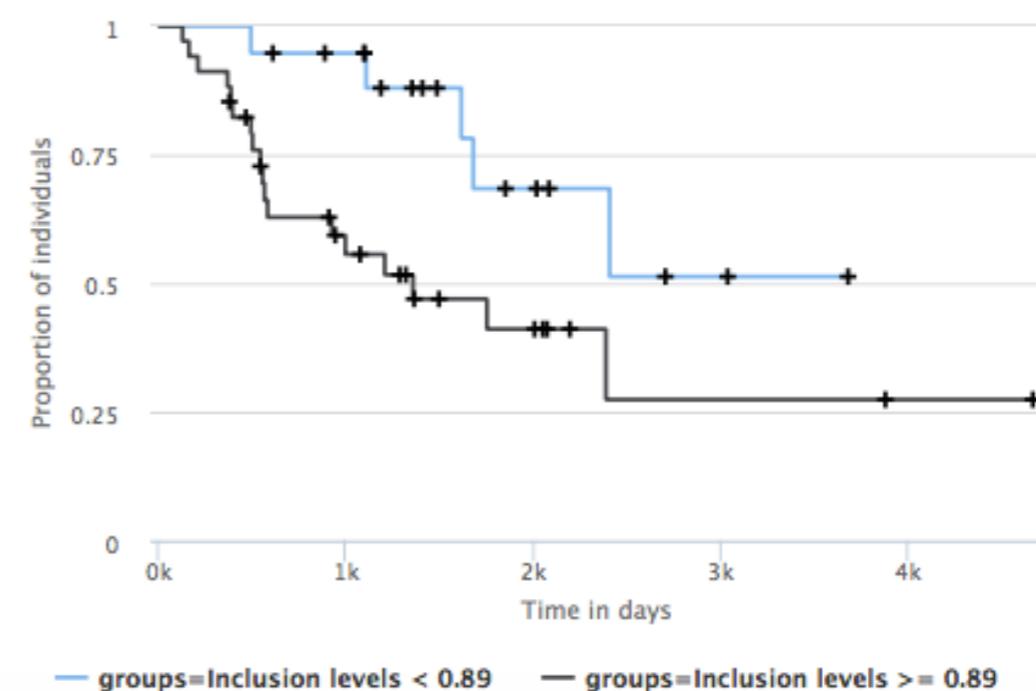


Analyses and Visualisation

- Vast array of customisable and interactive plots
- Features tooltips, zooming and plot exporting



Density plots



Survival curves

Case Study

Introduction

Workflow

Case Study

Testing

Conclusions

PSΨchomics Data Analyses Settings 127.0.0.1 Selected data category Change... No data loaded Selected splicing event Change... No data loaded

Load local files

Load TCGA/Firehose data

Tumour type
Glioma (GBMLGG)

Date
2016-01-28 (most recent)

Data type
Clinical data Junction quantification

Folder to store the data
/Users/Daniel/Downloads/

Load data

Quantify alternative splicing events



Welcome

Analyse alternative splicing based on transcriptomic and clinical data from The Cancer Genome Atlas (TCGA).

- 1** Load clinical data and alternative splicing junction quantification from TCGA .
More data types will soon be supported.
- 2** Quantify alternative splicing events based on the values from the percentage splicing index (PSI) metric.
Note: as TCGA does not include exon-intron junction quantification, intron retention events are not measurable.
- 3** Explore statistically significant events or individual events of interest through the following analyses:

Differential splicing analysis	Gene, transcript and protein information	Principal component analysis (PCA)	Survival analysis
Analyse alternative splicing quantification based on variance and median statistical tests. The groups available for differential analysis comprise sample types (e.g. normal versus tumour) and clinical attributes of patients (e.g. tumour stage).	For a given splicing event, examine its gene's annotation and corresponding transcripts and proteins. Related research articles are also available.	Explore alternative splicing quantification groups using associated clinical attributes.	Analyse survival based on clinical attributes (e.g. tumour stage, gender and race). Additionally, study the impact of the quantification of a single alternative splicing event on patient survivability.

Nuno Morais Lab, iMM (Nuno Agostinho , 2015-2016)
Special thanks to my lab colleagues for their work-related support and supporting chatter.

PSΨchomics Data Analyses Settings 127.0.0.1 Selected data category Change... Glioma 2016-01-28 Selected splicing event Change... No data loaded



Load local files

Load TCGA/Firehose data

Quantify alternative splicing events

Exon inclusion levels are measured from junction quantification using the Percent Spliced-In (PSI) metric.

Alternative splicing junction quantification

Junction quantification (Illumina HiSeq)

Alternative splicing event annotation

Human (hg19/GRCh37)

Event type(s)

Skipped exon

Minimum read counts threshold 10

Load quantification from file

Quantify events

Clinical data Junction quantification (Illumina HiSeq)

Table description: Clinical data of the patients [Save table](#)

Visible columns: patient.days_to_death × patient.days_to_last_followup × patient.ethnicity × patient.gender × patient.race_list.race × patient.radiation_therapy × patient.vital_status ×

Show 10 entries Search:

patient.days_to_death	patient.days_to_last_followup	patient.ethnicity	patient.gender
All	All	All	All
tcga-19-a6j4	121	not hispanic or latino	male
tcga-cs-6665		378	not hispanic or latino female
tcga-cs-6670		201	not hispanic or latino male
tcga-db-a4xc		840	not hispanic or latino male
tcga-dh-5141		968	hispanic or latino male
tcga-dh-a66b		901	not hispanic or latino male
tcga-du-7300	1886	not hispanic or latino	female
tcga-du-a76o		217	not hispanic or latino male
tcga-e1-a7yn	727	not hispanic or latino	female
tcga-fg-8187		215	not hispanic or latino male

Showing 1 to 10 of 1,109 entries Previous 1 2 3 4 5 ... 111 Next

127.0.0.1

PSΨchomics Data Analyses Settings Selected data category Change... Glioma 2016-01-28 Selected splicing event Change... SE 1 + 244999058 245005246 245005360 245005497 COX20

Differential splicing analysis

Gene, transcript and protein information

Principal Component Analysis (PCA)

Survival analysis

+ Load local files

+ Load TCGA/Firehose data

Quantify alternative splicing events

Exon inclusion levels are measured from junction quantification using the Percent Spliced-In (PSI) metric.

Alternative splicing junction quantification

Junction quantification (Illumina HiSeq)

Alternative splicing event annotation

Human (hg19/GRCh37)

Event type(s)

Skipped exon

Minimum read counts threshold 10

Load quantification from file

Quantify events

Junction quantification (Illumina HiSeq) **Inclusion levels**

... Clinical data of the patients 

Visible columns

patient.days_to_death × patient.days_to_last_followup × patient.ethnicity × patient.gender ×
patient.race_list.race × patient.radiation_treatment × patient.vital_status ×

Show 10 entries Search:

patient.days_to_death	patient.days_to_last_followup	patient.ethnicity	patient.gender
All	All	All	All
tcga-19-a6j4	121	not hispanic or latino	male
tcga-cs-6665		378	not hispanic or latino female
tcga-cs-6670		201	not hispanic or latino male
tcga-db-a4xc		840	not hispanic or latino male
tcga-dh-5141		968	hispanic or latino male
tcga-dh-a66b		901	not hispanic or latino male
tcga-du-7300	1886	not hispanic or latino	female
tcga-du-a76o		217	not hispanic or latino male
tcga-e1-a7yn	727	not hispanic or latino	female
tcga-fg-8187		215	not hispanic or latino male

Showing 1 to 10 of 1,109 entries

Previous 1 2 3 4 5 ... 111 Next

Go to #tab-7499-3 on this page

PSΨchomics Data Analyses Settings Selected data category Change... Glioma 2016-01-28 Selected splicing event Change... SE 1 + 244999058 245005246 245005360 245005497 COX20

127.0.0.1

Data to perform PCA on
Inclusion levels

Preprocessing
 Center values
 Scale values

Percentage of missing values to tolerate per event ?
0% 100%

Filter data groups
Click on 'Groups' to create or edit Groups

Calculate PCA

Choose X axis
PC1 (23.756% explained variance)

Choose Y axis
PC2 (9.945% explained variance)

Clinical groups to colour the PCA
Click on 'Groups' to create or edit Groups

Show variance plot Plot PCA

PSΨchomics Data Analyses Settings Selected data category Change... Selected splicing event Change... 127.0.0.1 45005246 245005360 245005497 COX20

Groups

Subset by

Column Rows Subset expression Regular expression

Automatically create groups according to the unique values of the selected column. For instance, to create groups by tumour stage, type **tumor_stage** and select the first suggestion that appears.

Select column

Start typing to search for columns

Create group

Search:

Names	Rows	Subset	Input
(g3 ∩ lgg)	265	(Column ∩ Column)	(patient.neoplasm_histologic_grade ∩ admin.di
(g2 ∩ lgg)	249	(Column ∩ Column)	(patient.neoplasm_histologic_grade ∩ admin.di
gbm	594	Column	admin.disease_code

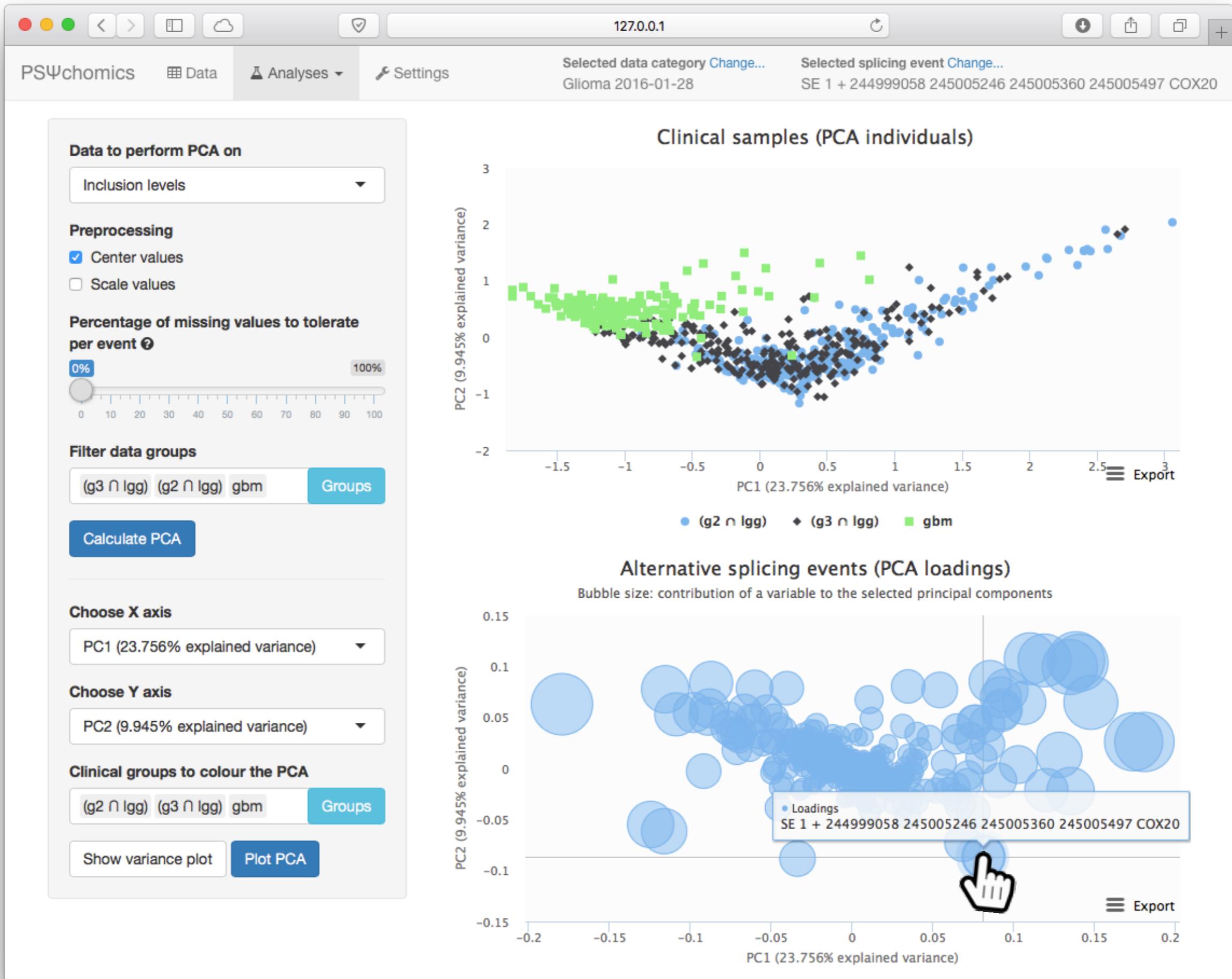
Showing 1 to 3 of 3 entries

Previous 1 Next

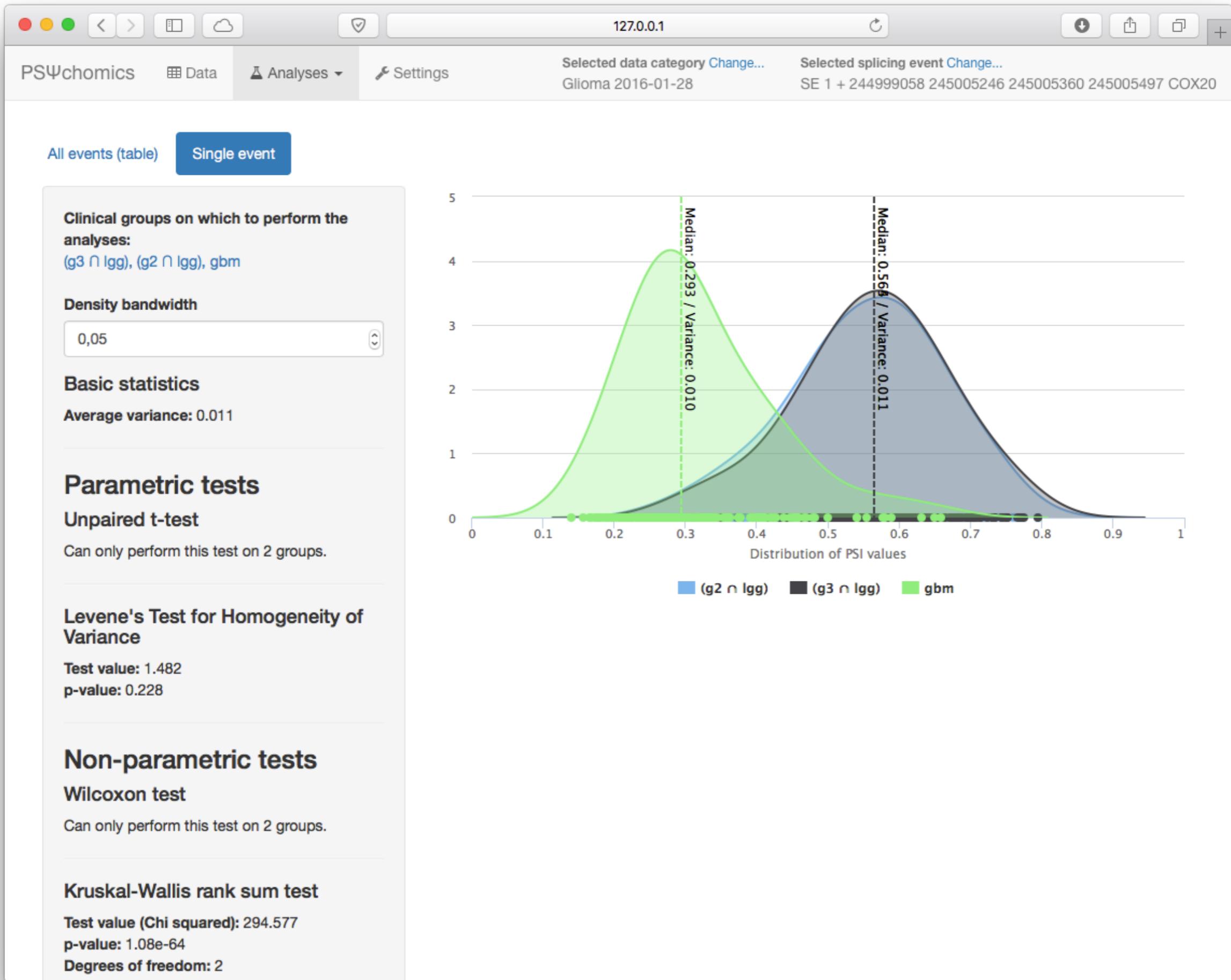
Select groups by clicking on them in order to merge, intersect or remove selected groups.

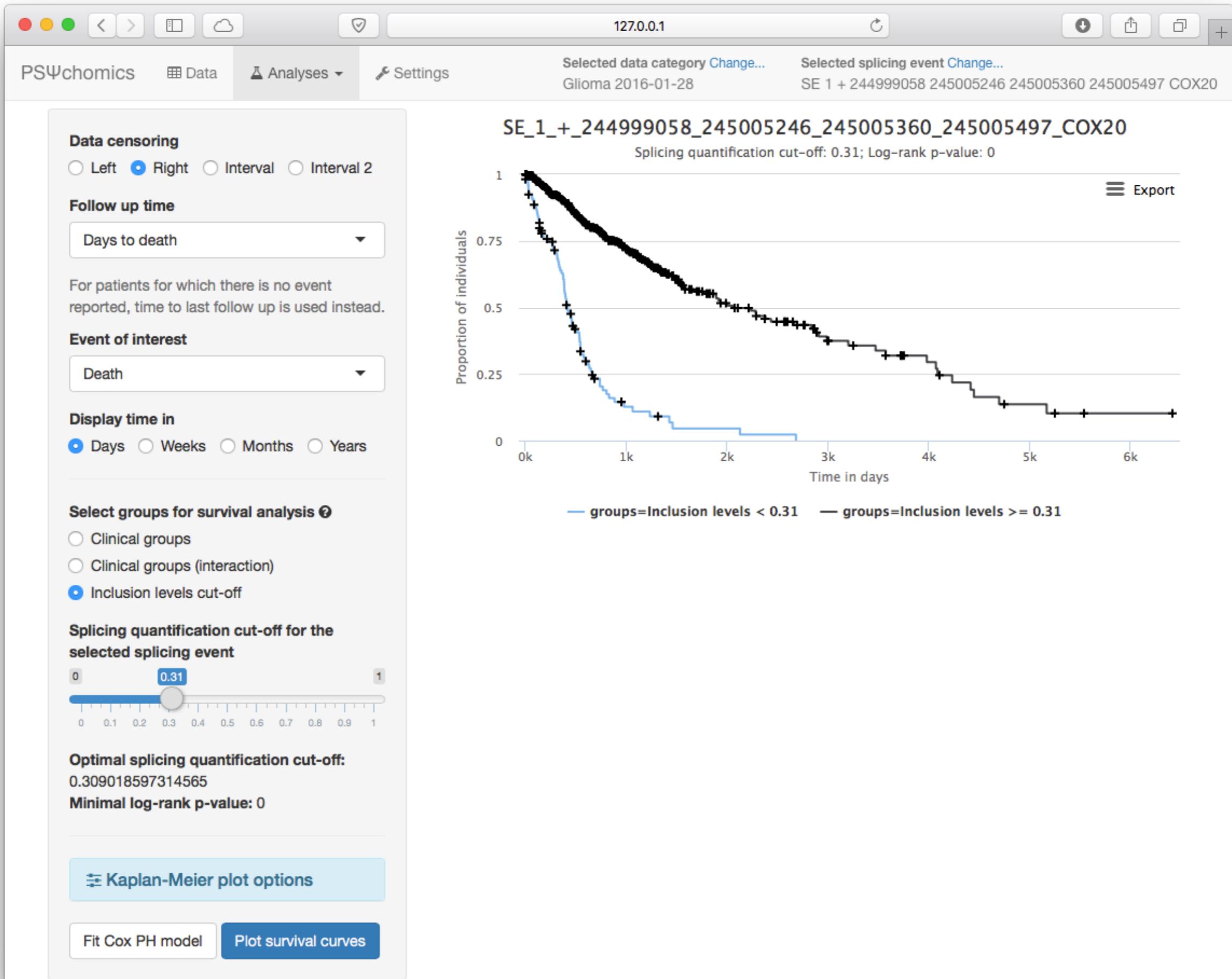
Merge Intersect Remove Remove all groups

Close



PCA: dimensionality reduction by selecting the main directions of variance





PSΨchomics Data Analyses Settings Selected data category Change... Glioma 2016-01-28 Selected splicing event Change... SE 1 + 244999058 245005246 245005360 245005497 COX20

COX20 ENSG00000203667

Species human (hg19 assembly)
Location Chromosome 1: 244,998,624-245,008,359 (forward strand)
Description COX20 cytochrome C oxidase assembly factor [Source:HGNC Symbol;Acc:26970] (protein_coding)
Links Ensembl UCSC GeneCards Human Protein Atlas (Cancer Atlas) VAST-DB

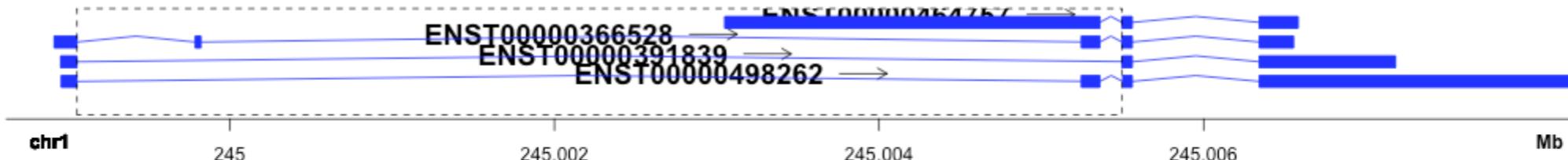
Relevant PubMed articles [Show more articles](#)

Human COX20 cooperates with SCO1 and SCO2 to mature COX2 and promote the assembly of cytochrome c oxidase Bourens M et al. (2014). *Hum Mol Genet*, 23(11).

Multiple Roles of the Cox20 Chaperone in Assembly of *Saccharomyces cerevisiae* Cytochrome c Oxidase Elliott LE et al. (2012). *Genetics*, 190(2).

Expression of Mitochondrial Cytochrome C Oxidase Chaperone Gene (COX20) Improves Tolerance to Weak Acid and Oxidative Stress during Yeast Fermentation Kumar V et al. (2015). *PLoS One*, 10(10).

Transcripts



chr1 245 245.002 245.004 245.006 Mb

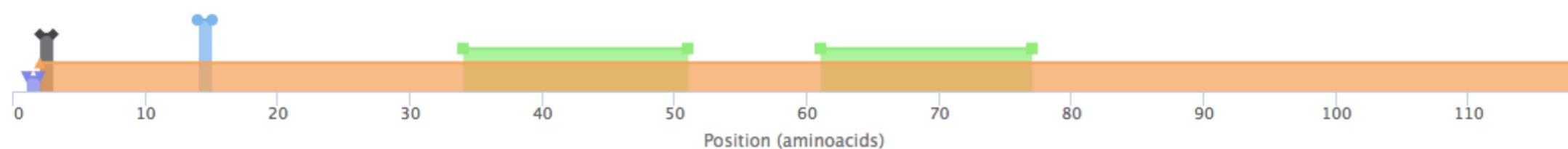
Select transcript

ENST00000411948

Select protein

COX20_HUMAN (UniProtKB/Swiss-Prot)

Ensembl Uniprot



Position (aminoacids)

Legend: splice variant, modified residue, transmembrane region, chain, initiator methionine



Continuous and
Unit Testing



Usability
Testing



Performance
Benchmarking

Testing



Continuous and Unit Testing

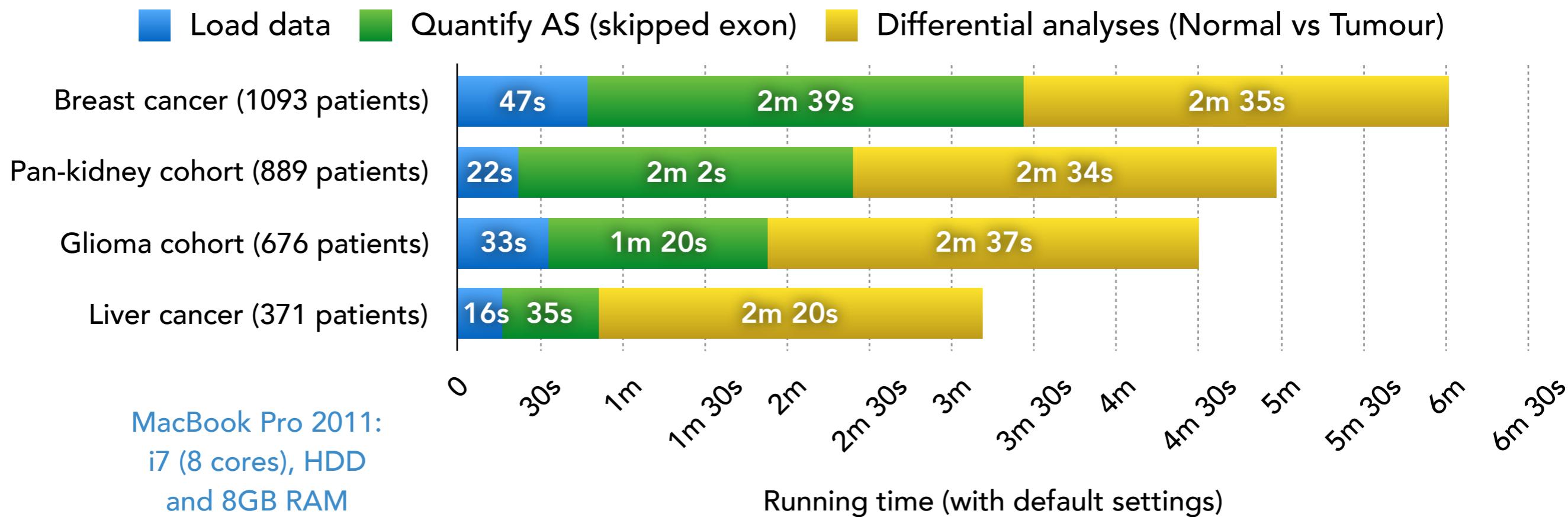


Usability Testing



Performance Benchmarking

- Common tumour types were selected from TCGA
- Average time was collected over 10 consecutive runs



Conclusions

Introduction

Workflow

Case Study

Testing

Conclusions

Conclusions

Quantify, analyse and visualise alternative splicing in cancer data

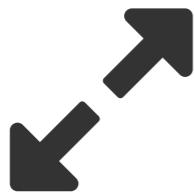
Modular architecture to easily modify and extend the program

Command-line and easy-to-use graphical interface

6 minutes using processed data with the highest number of patients in TCGA

Incorporates clinical information

Future work



Extend available data sources



Quantify alternative splicing from raw data

Less dependence on processed data



Deploy to the web

No installation required

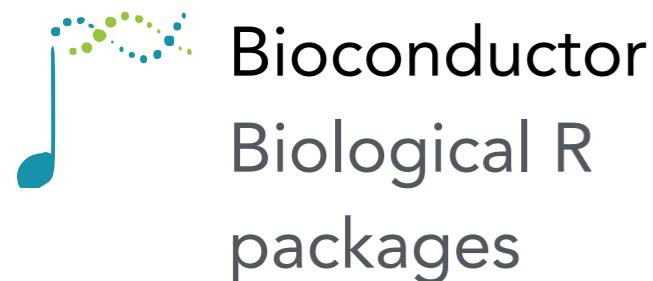
Always up-to-date

PSIchomics

(MIT license)



Code hosting



Biological R packages

127.0.0.1

PSIchomics Data Analyses Settings Selected data category Change... Adrenocortical carcinoma 2016-01

Clinical data Junction quantification (Illumina HiSe

Table description: Clinical data of the patients

Visible columns

- patient.days_to_death × patient.days_to_last_follow
- patient.race_list.race × patient.radiation_therapy ×
- patient.stage_event.pathologic_stage_tumor_stage ×

Show 10 entries

patient.days_to_death	patient.
All	All
tcga-or-a5kp	
tcga-or-a5l5	
tcga-or-a5lb	1204
tcga-p6-a5og	383
tcga-pk-a5hb	
tcga-or-a5j1	1355
tcga-or-a5j2	1677
tcga-or-a5j3	

+ Load local files

+ Load TCGA/Firehose data

Quantify alternative splicing events

Measure exon inclusion levels from junction quantification. The Percent Spliced-In (PSI) metric is used.

Alternative splicing junction quantification

Junction quantification (Illumina HiSeq)

Alternative splicing event annotation

Human (hg19/GRCh37)

Event type(s)

Skipped exon (SE)

Minimum read counts threshold ⓘ

10

Calculate inclusion levels

Thanks to you all!



Nuno Morais Lab



André Falcão



Ana Rita Grosso



Ciências
ULisboa