

Analysis of bulk RNA-seq data using Chipster

1.-2.3.2023

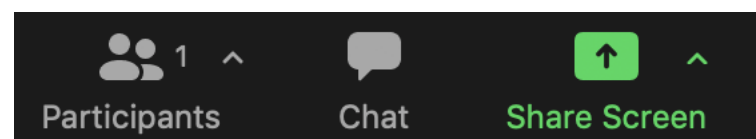
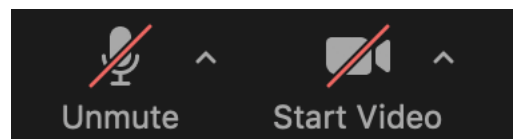
Eija Korpelainen and Lauren Gerber



CSC – Suomalainen tutkimuksen, koulutuksen, kulttuurin ja julkishallinnon ICT-osaamiskeskus

Virtual coffee & get together

- Test your mic & tell us a little bit about yourself
 - Name
 - Group
 - Do you already have RNA-seq data?
 - What is The Sign of Spring to you?
- Zoom etiquette
 - When you are not talking, please keep your mic muted
 - You can find all the controls (mic, video, chat, screen sharing) at the bottom of the Zoom window
 - In chat: write your questions to Everyone, not just to Hosts
 - Please use a headset to avoid the echo



To: Everyone

Type message here...



Understanding data analysis - why?

- Bioinformaticians might not always be available when needed
- Biologists know their own experiments best
 - Biology involved (e.g. genes, pathways, etc)
 - Potential batch effects etc
- Allows you to design experiments better → less money wasted
- Allows you to discuss more easily with bioinformaticians

Introduction to Chipster



Chipster

- User-friendly analysis software for high-throughput data
- Provides an easy access to over 500 analysis tools
- Command line tools
- R/Bioconductor packages
- Free, open source software

- What can I do with Chipster?
 - analyze high-throughput data
 - visualize data efficiently
 - share analysis sessions



Chipster

Open source platform for data analysis



- Home
- Getting access
- Manual
- Tutorial videos
- Course material
- Cite
- Contact

Welcome to Chipster

Chipster is a user-friendly analysis software for high-throughput data such as RNA-seq and single-cell RNA-seq. Chipster provides a web interface to over 500 analysis tools, and the actual analysis jobs run on the server side making use of CSC's computing environment.

If you would like to use Chipster hosted by CSC, you need a [user account](#). Please note that Chipster is also available for [local server installation](#) free of charge.



[Launch Chipster](#)

Training:

- 25.10.2022 [Spatial transcriptomics \(Visium\) data analysis](#)
- 7.10.2022 [Single-cell RNA-seq data analysis](#)
- 30.6.2021 [MOOC Single-cell RNA-seq data analysis using Chipster](#), instructions on [how to get started](#)

News and resources:

- [Analysis of QuantSeq 3' UMI RNA-seq data enabled](#)
- [Chipster introduction video](#)
- [Instructions for moving data from Puhti to Chipster](#)
- [Video on how to convert tables to Chipster format and create phenodata file](#)
- [Lecture videos of advanced single cell RNA-seq data analysis course](#)
- [RNA-seq resources](#)

Chipster user interface (chipster.rahtiapp.fi)



Files

Workflow List

[Add file](#)

```
graph TD; fastq((fastq)) --> html1[html]; fastq --> txt1[txt]; fastq --> gz1[gz]; fastq --> txt2[txt]; html1 --> bam[bam]; txt1 --> bai[bai]; gz1 --> log[log]; txt2 --> log; bam --> txt3[txt]; bam --> pdf1[pdf]; bam --> txt4[txt]; bam --> tsv1[tsv]; bai --> tsv1; log --> tsv1; tsv1 --> pdf2[pdf]; tsv1 --> tsv2[tsv]; tsv1 --> bed[bed];
```

Tools

NGS Microarray Misc Job: 0

Category

- Quality control
- Preprocessing
- Utilities
- Matching sets of genomic regions
- Alignment
- Variants
- RNA-seq
- Small RNA-seq
- Single cell RNA-seq
- ChIP- and DNase-seq
- 16S rRNA sequencing
- CNA-seq

Tool

- Read quality with FastQC
- Read quality with MultiQC for many FASTQ files
- Read quality statistics with FASTX
- Read quality statistics with PRINSEQ
- RNA-seq quality metrics with RseQC
- RNA-seq strandedness inference and inner distance estimation using RseQC
- Collect multiple metrics from BAM
- PCA and heatmap of samples with DESeq2
- Check FASTQ file for errors
- Combine reports using MultiQC

[Parameters](#) [Run](#)

The tool runs FastQC on multiple FASTQ files, and then combines the reports using MultiQC. Input file is a single Tar package containing all the FASTQ files, which can be gzipped. This tool is based on the FastQC and MultiQC packages. [More info...](#)

File

ngs-data-table.tsv ...

[Spreadsheet](#) [Text](#) [Expression Profile](#) [Scatter Plot](#) [Phenodata](#) [Details](#)

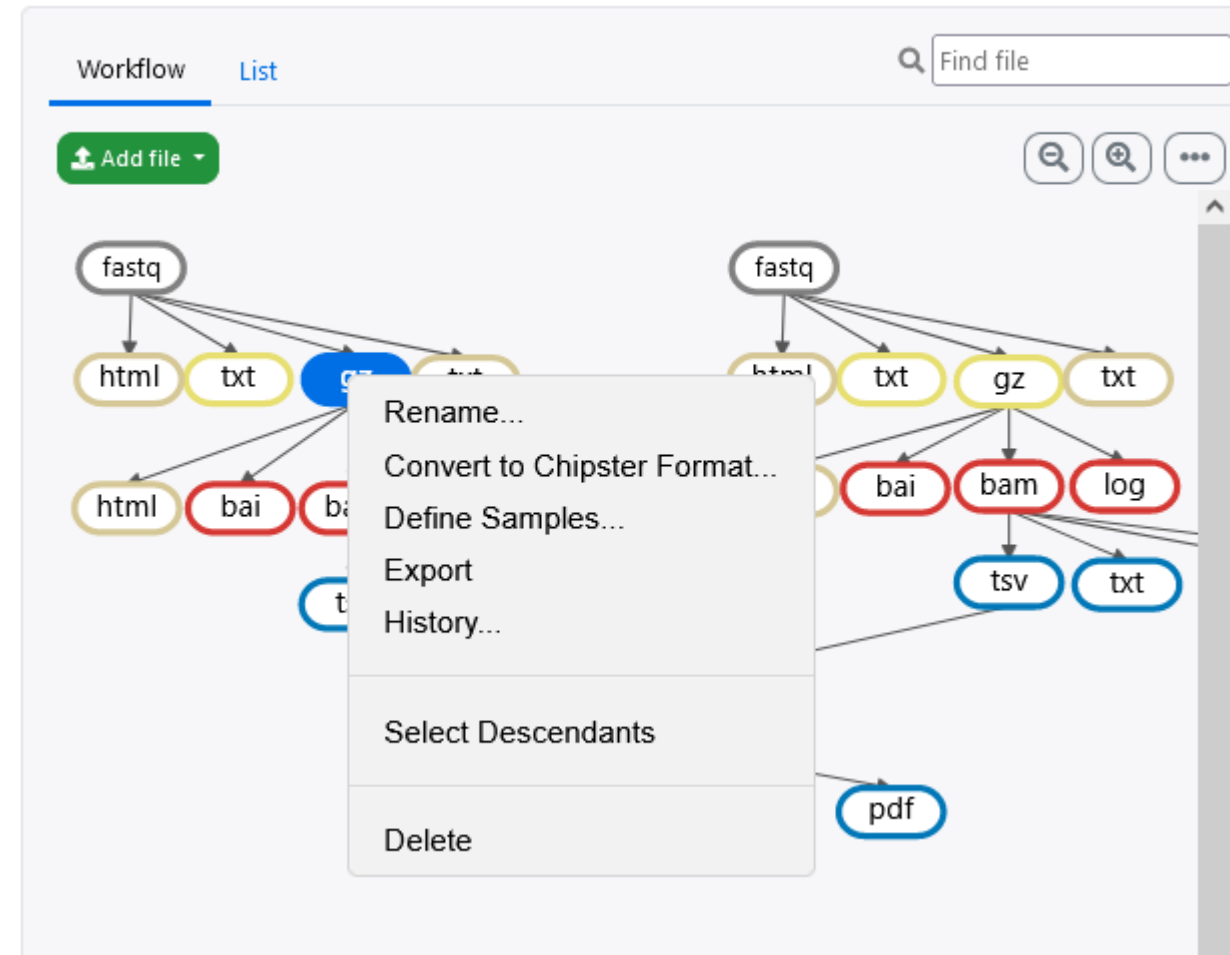
First 101 rows of 58396 [View in full screen to see all the rows.](#) [Full Screen](#)

identifier	chr	start	end	length	sequence	chip.sample001.tsv	chip.sample002.tsv
ENSG000000000003	X	100627108	100639991	12883	NA	0	0
ENSG000000000005	X	100584801	100599885	15084	NA	0	0
ENSG000000000419	20	50934866	50958555	23689	NA	0	0

Workflow view

- Shows the relationships of the files
- You can move the boxes around, and zoom in and out.
- Several files can be selected by
 - keeping the Ctrl/Cmd key down
 - drawing a box around them
- Right clicking a file allows you to
 - Download (“Export”)
 - Delete
 - Rename
 - View history
 - Select descendants
 - Convert to Chipster format (for tables)
 - Define samples (for FASTQ files)

Files



Options for importing data to Chipster



- Add file button
 - Upload files
 - Upload folder
 - Download from URL
- Tools
 - Import from SRA database
 - Utilities / Retrieve data from ENA database
 - Import from Ensembl database
 - Utilities / Retrieve data for a given organism in Ensembl
 - Import from URL
 - Utilities / Download file from URL directly to server
 - Import from Illumina BaseSpace
 - Utilities / Retrieve data from Illumina BaseSpace
 - Access token needed
- Sessions tab
 - Import session file

Analysis sessions

- Your analysis is saved automatically in the cloud
 - Session includes all the files, their relationships and metadata (what tool and parameters were used to produce each file).
 - Session is a single .zip file.
 - Note that cloud sessions are not stored forever! Remember to download the session when ready.
- You can share sessions with other Chipster users
 - You can give either read-only or read-write access
- If your analysis job takes a long time, you don't need to keep Chipster open
 - Once you have clicked Run you can close Chipster
 - Open Chipster later and the results will be there

Define samples: assign FASTQ files to samples

- If you have paired end data, you can assign the R1 and R2 files to samples
 - Select the files to be paired, right click, and select **Define samples**
- This allows you to e.g. align all the samples with just one click: **Run for each sample**

Define Sample Files ×

Paired end Single end Reset All

Forward Identifier Reverse Identifier Find Pairs

Samples - Paired End
2 samples, 4 files Reset

⊕	⊕
lymphnode4a_R1.fq.gz	lymphnode4a_R2.fq.gz
lung3e_R1.fq.gz	lung3e_R2.fq.gz

Close Save

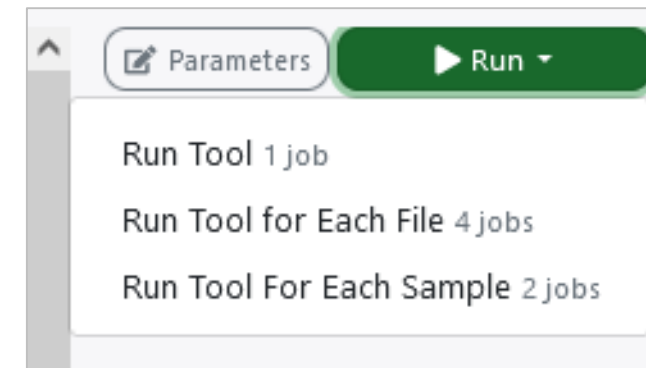
Running many analysis jobs at the same time



- You can have many analysis jobs running at the same time
 - No need to wait that one finishes before starting a new one

Run button gives several options:

- Run tool
 - Runs the selected analysis tool once (all the selected input files are analyzed together)
- Run tool for each file
 - Runs the selected analysis tool for each of the input files individually
- Run tool for each sample
 - If you have grouped paired end FASTQ files to samples using the Define samples –option, you can run the selected analysis tool for the input files in a sample specific manner.



Converting a read count table to Chipster format



- Save the file as txt or tsv and import it to Chipster
- Select the file, right-click, and select **Convert to Chipster format**
- Indicate the identifier column and the sample columns

Convert to Chipster format

Source file preview counts.txt

gene	sample1	sample2	sample3	sample4	sample5
FBgn0000003	0	1	1	0	0
FBgn0000008	118	139	77	89	142

Identifier Sample Include Exclude

Identifier column:

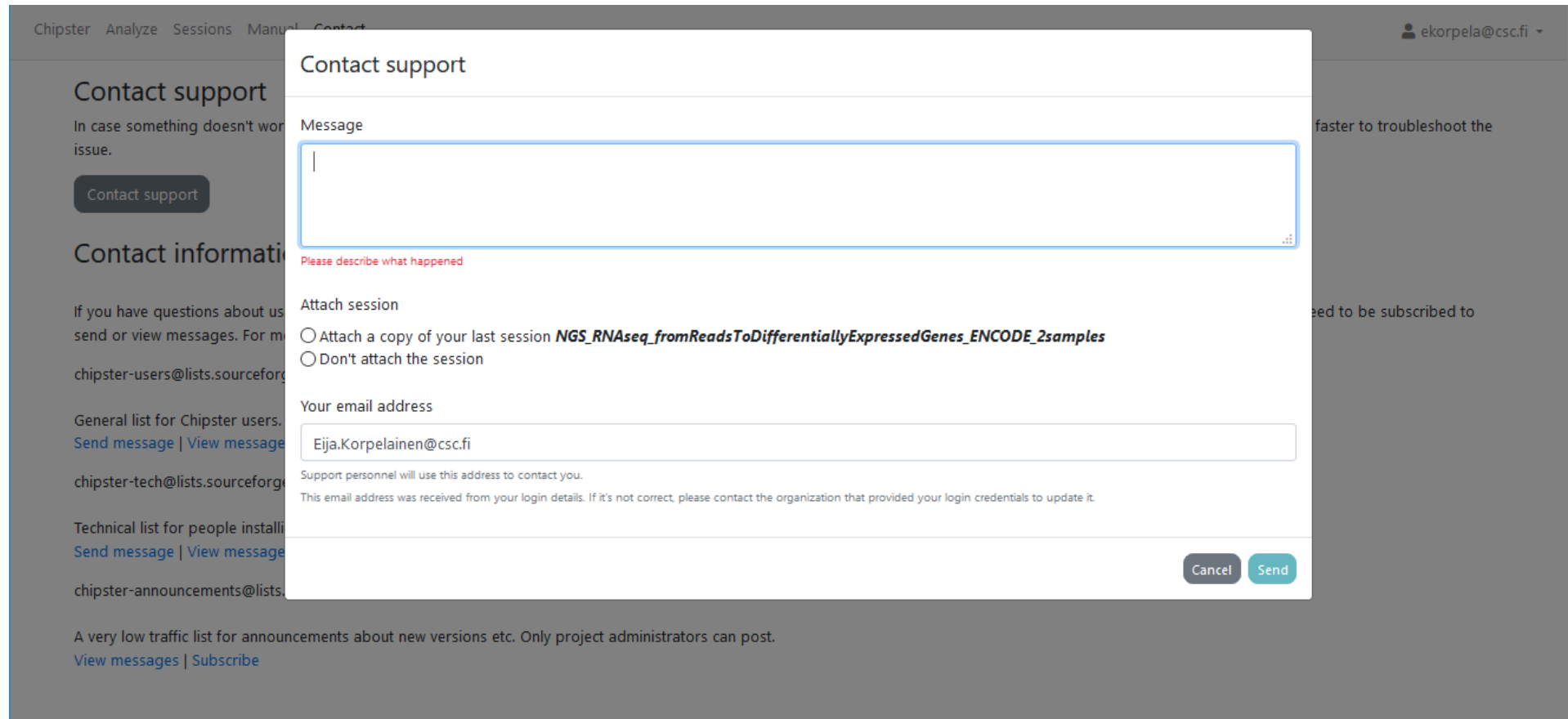
Sample columns:

Other columns: Selected columns will be

Select columns:

Problems? Send us a support request

-request includes the error message and link to analysis session (optional)



Chipster Analyze Sessions Manual Contact

ekorpela@csc.fi

Contact support

In case something doesn't work or you have an issue.

Contact support

Contact information

If you have questions about using Chipster, you can send or view messages. For more information, see the [Chipster user guide](#).

chipster-users@lists.sourceforge.net
General list for Chipster users.
[Send message](#) | [View message](#)

chipster-tech@lists.sourceforge.net
Technical list for people installing Chipster.
[Send message](#) | [View message](#)

chipster-announcements@lists.sourceforge.net
A very low traffic list for announcements about new versions etc. Only project administrators can post.
[View messages](#) | [Subscribe](#)

Contact support

Message

Please describe what happened

Attach session

Attach a copy of your last session *NGS_RNAseq_fromReadsToDifferentiallyExpressedGenes_ENCODE_2samples*

Don't attach the session

Your email address

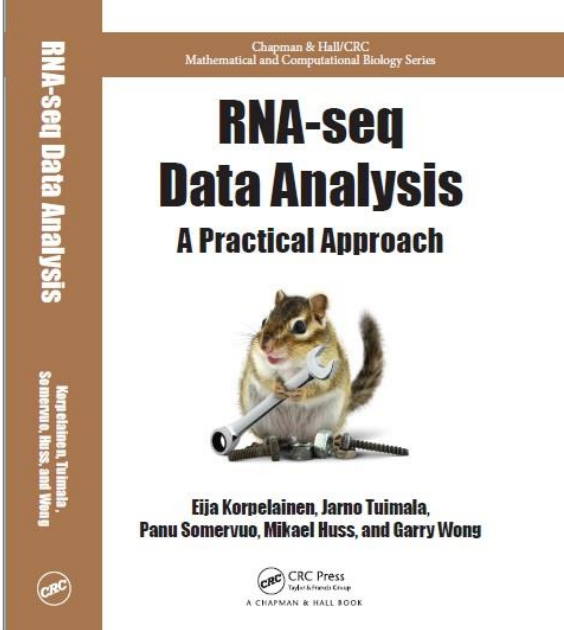
Eija.Korpelainen@csc.fi

Support personnel will use this address to contact you.
This email address was received from your login details. If it's not correct, please contact the organization that provided your login credentials to update it.

Cancel Send

More info

- chipster@csc.fi
- <http://chipster.csc.fi>
- Chipster tutorials in YouTube
- <https://chipster.csc.fi/manual/courses.html>



Chipster Tutorials
4.37K subscribers

HOME VIDEOS PLAYLISTS COMMUNITY CHANNELS ABOUT

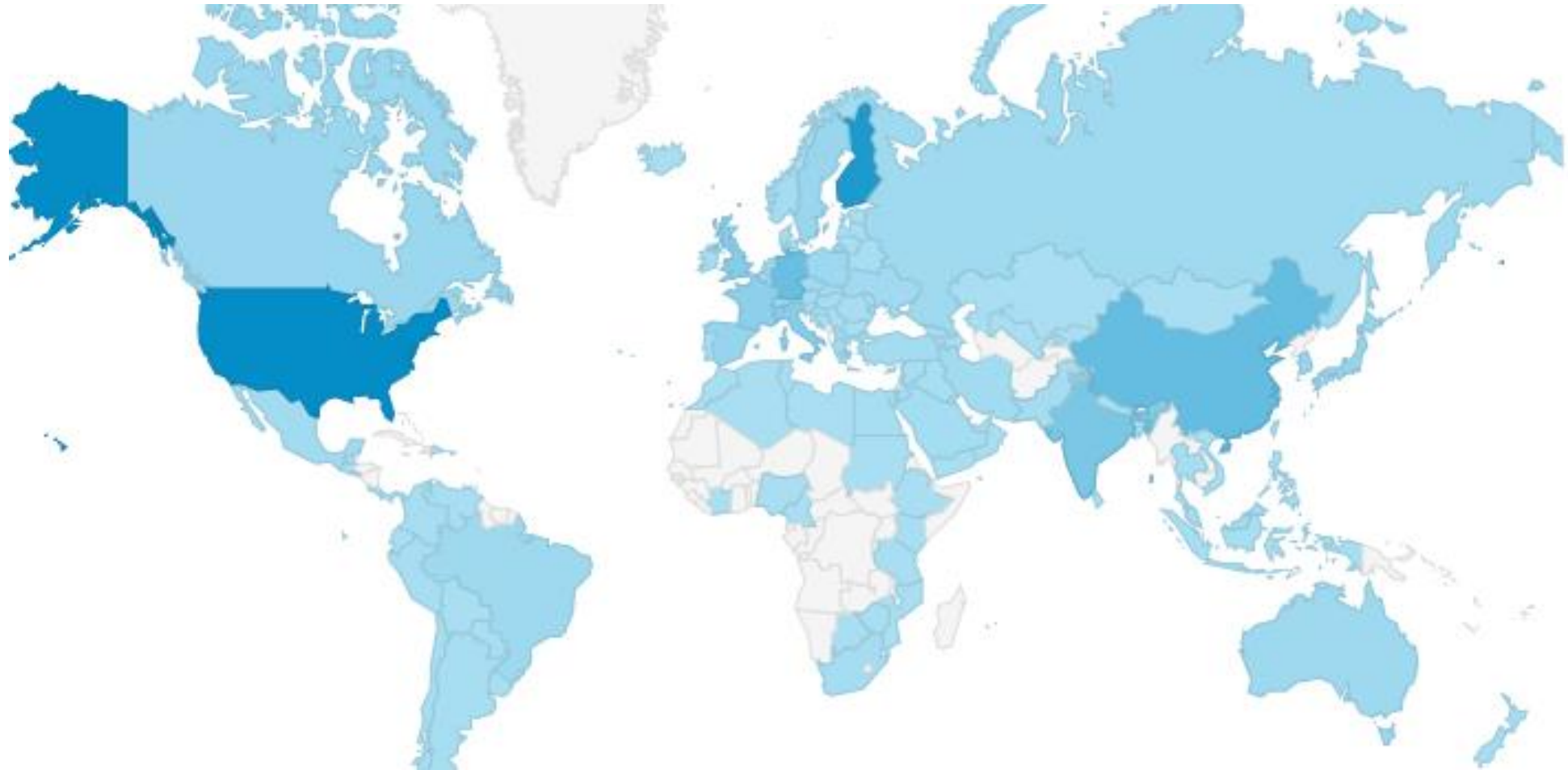
Uploads PLAY ALL SORT BY

Video Title	Duration	Views	Time Ago
Differential expression analysis using DESeq2 and edgeR	26:33	46K views	4 years ago
Introduction to RNA-seq data analysis	11:23	33K views	2 years ago
1. Introduction to single cell RNA-seq	32:29	28K views	2 years ago
RNA-seq course: Quality control & preprocessing of raw reads Using FASTQC & Trimmomatic	25:55	22K views	4 years ago
Aligning RNA-seq reads to reference genome	24:23	17K views	2 years ago

Acknowledgements to Chipster users and contributors



Users' feedback and ideas have helped us to shape the software over the years.
Let us know what needs to be improved!



What are we going to do during this course?

Recap the main points of the theory

Practise with different data sets:

- Analysis starting from FASTQ files: QC, alignment, counting, combining, DE analysis
 - Human lung and lymph node sample (2 samples)
- DE analysis starting from a count table
 - Human lung and lymph node samples (10 samples)
- DE analysis when there is a batch effect
 - Drosophila knock-down experiment
- DE analysis when there are big differences between individuals
 - Human adenocarcinoma samples before and after treatment