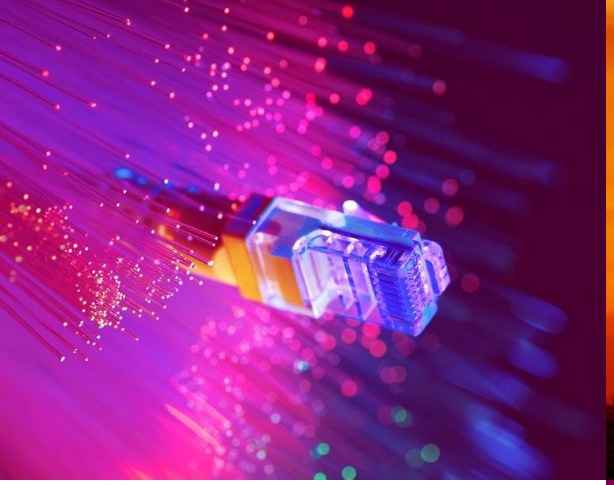


Analysis of bulk RNA-seq data using Chipster

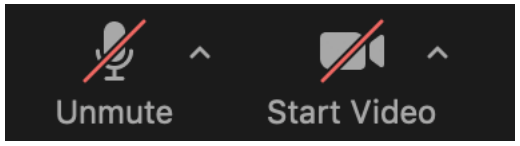
1.-2.3.2023

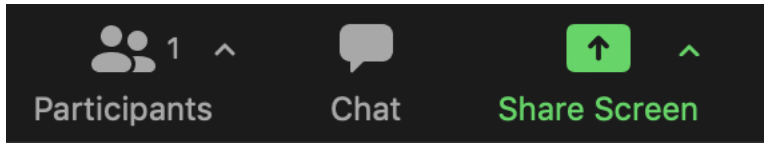
Eija Korpelainen and Lauren Gerber

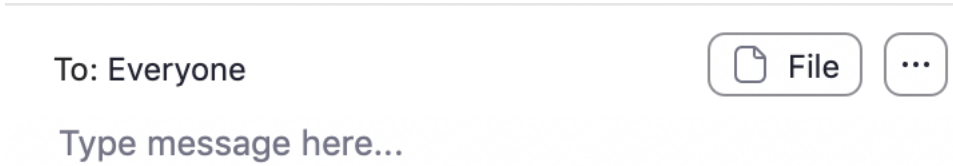
CSC – Suomalainen tutkimuksen, koulutuksen, kulttuurin ja julkishallinnon ICT-osaamiskeskus

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







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

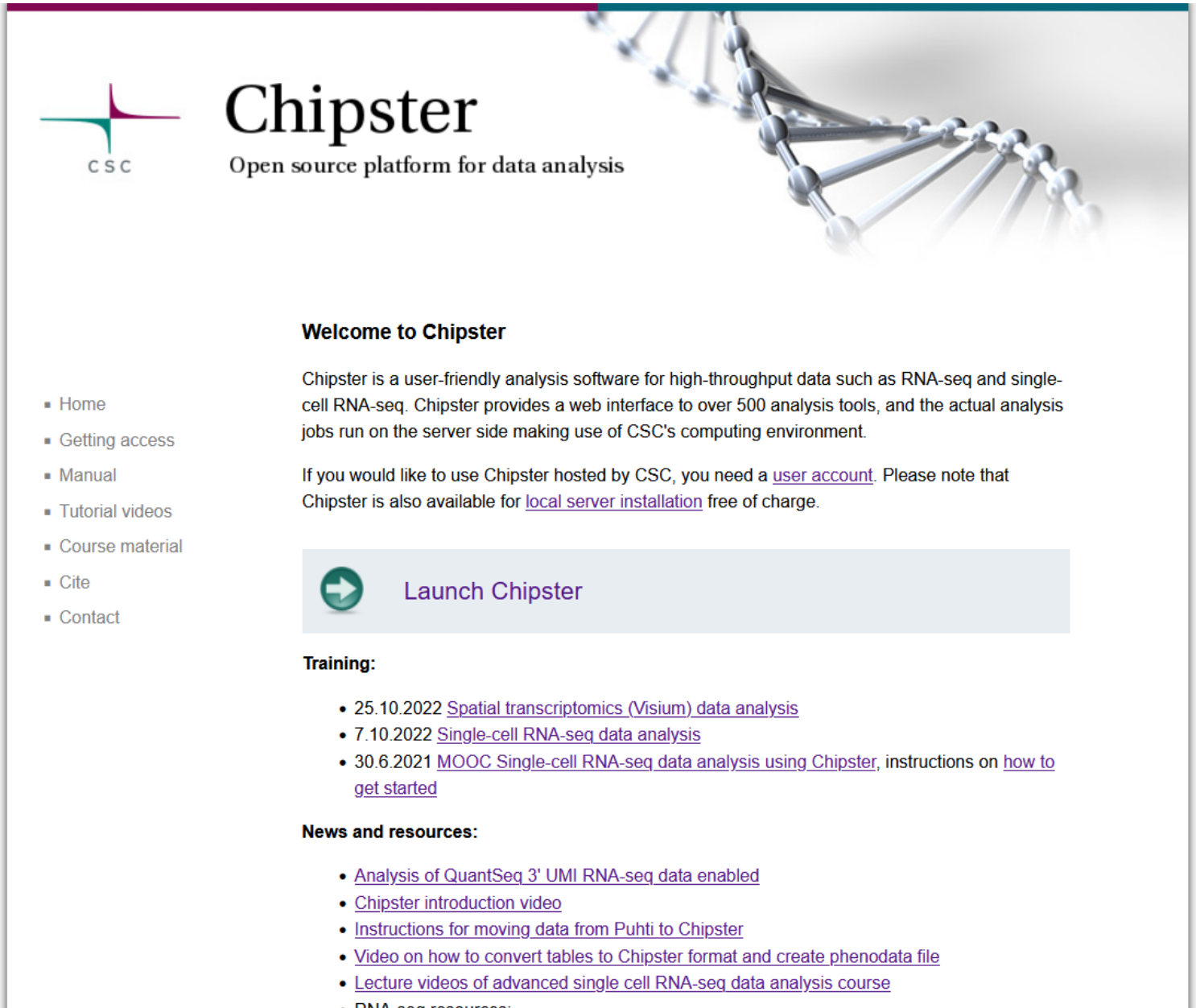
Introduction to Chipster

Chipster

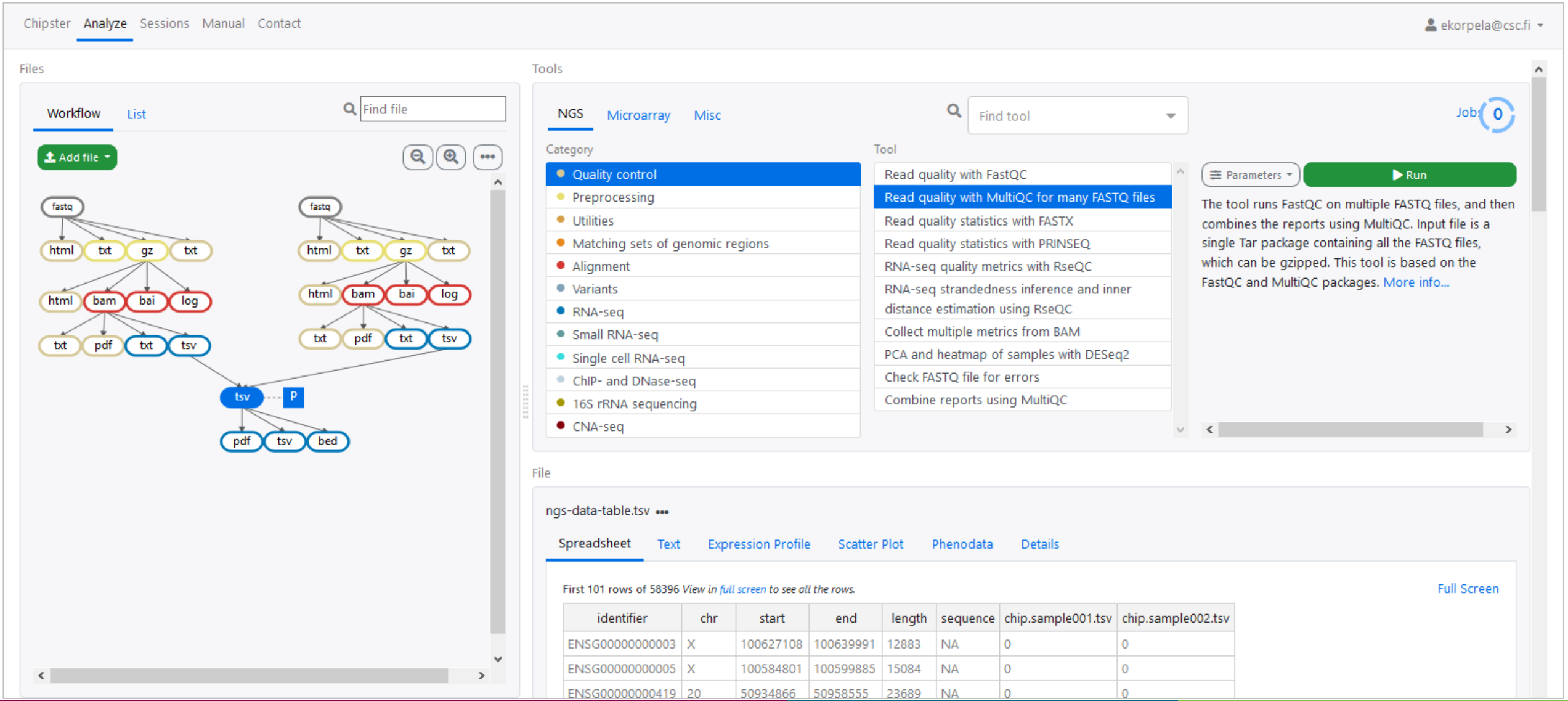
- User-friendly analysis software for high-throughput data
- Provides an easy access to over 500 analysis tools
- Command line tools
- R/Bioconductor packages
- Free, open source software

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

Chipster website (https://chipster.csc.fi/)



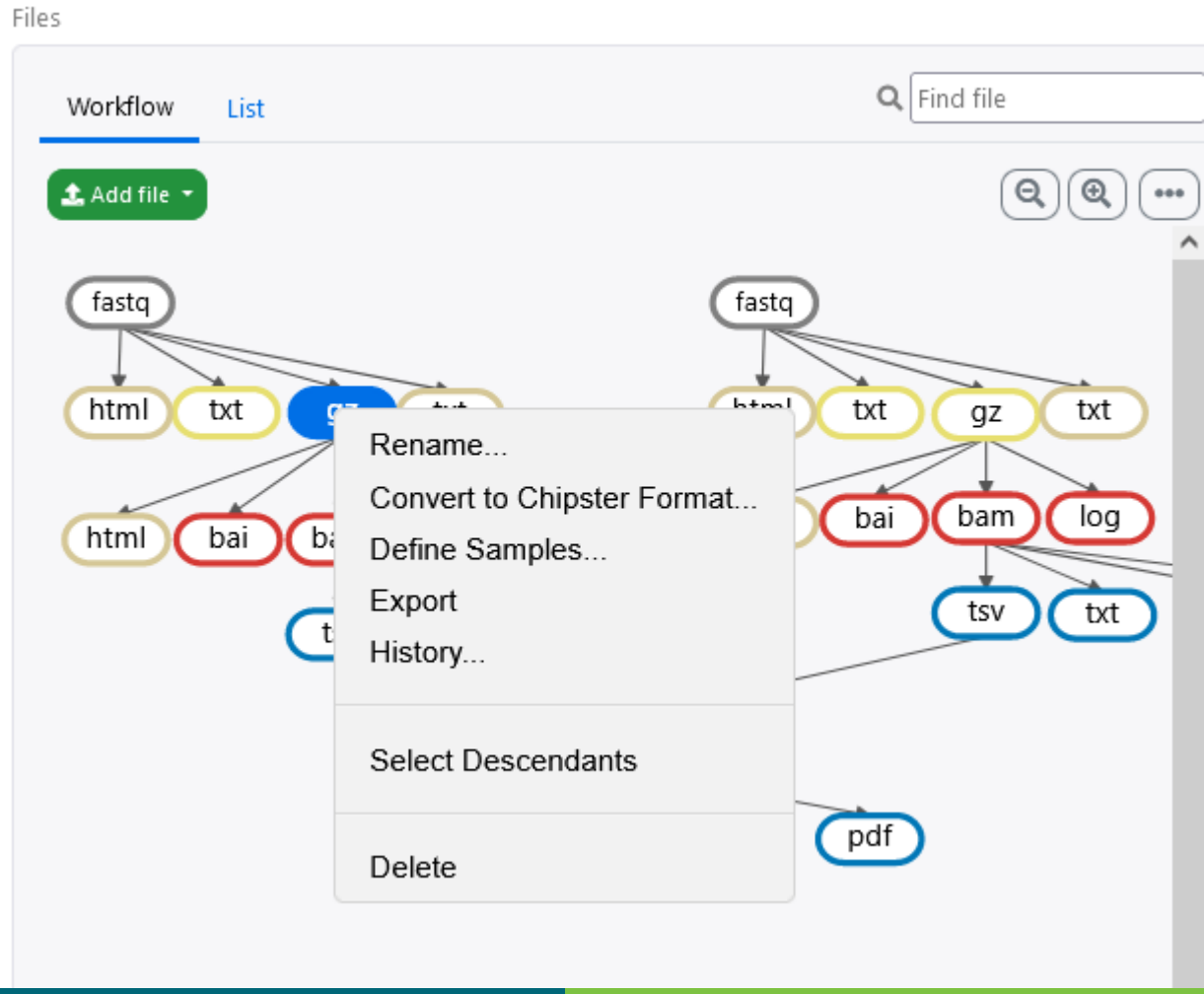
Chipster user interface (chipster.rahtiapp.fi)



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)



Options for importing data to Chipster

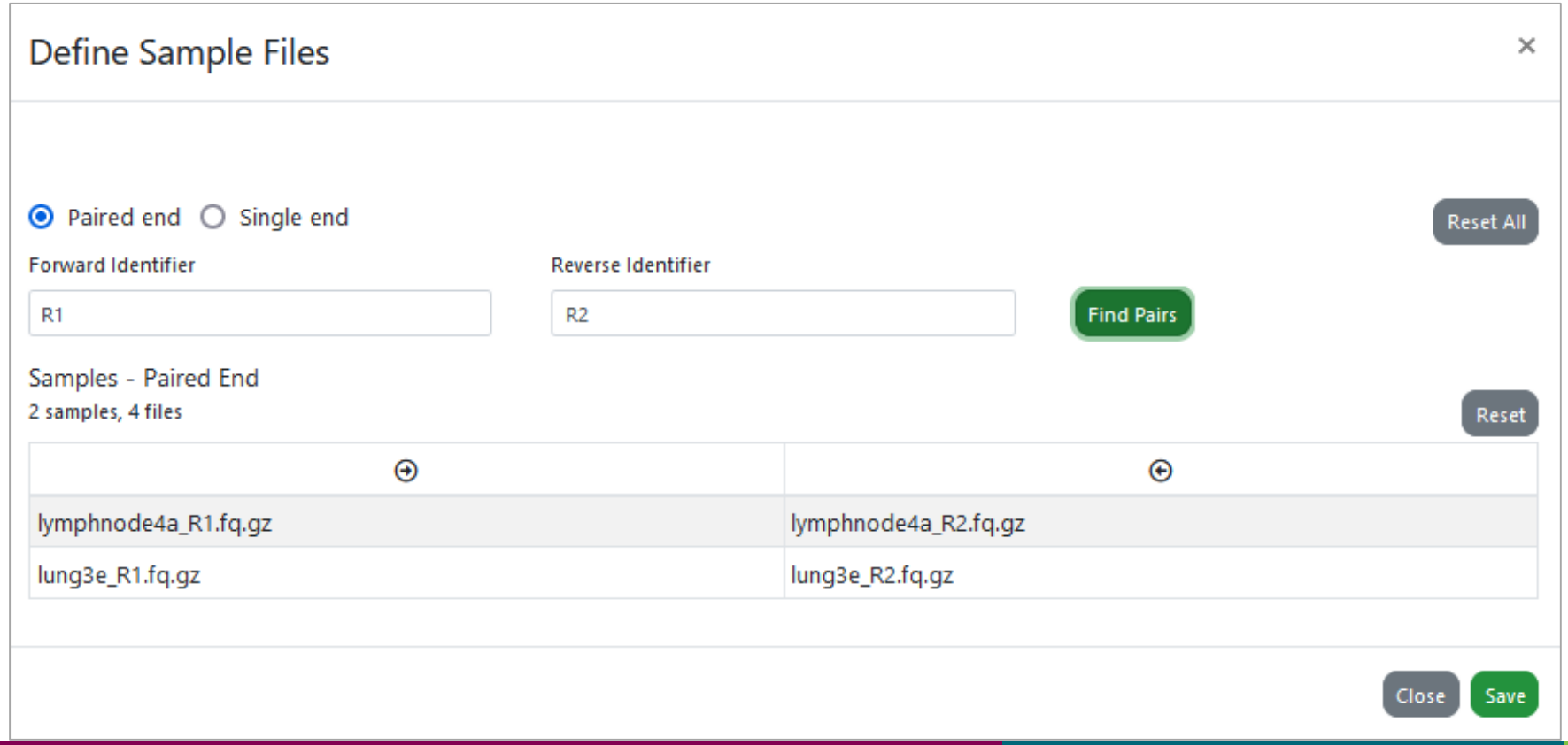
- Add file button
  - Upload files
  - Upload folder
  - Download from URL
- Tools
  - Import from SRA database
    - Utilities / Retrieve data from ENA database
  - Import from Ensembl database
    - Utilities / Retrieve data for a given organism in Ensembl
  - Import from URL
    - Utilities / Download 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

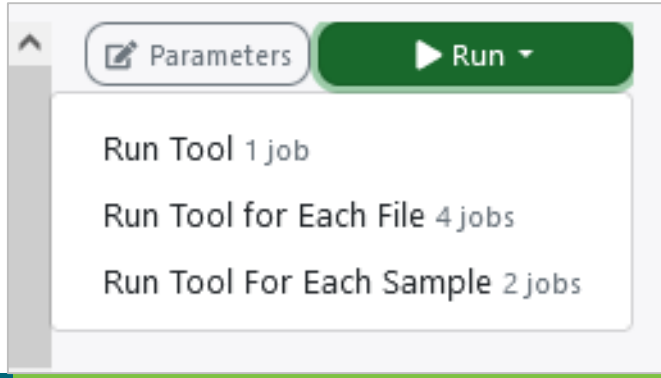


Running many analysis jobs at the same time

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

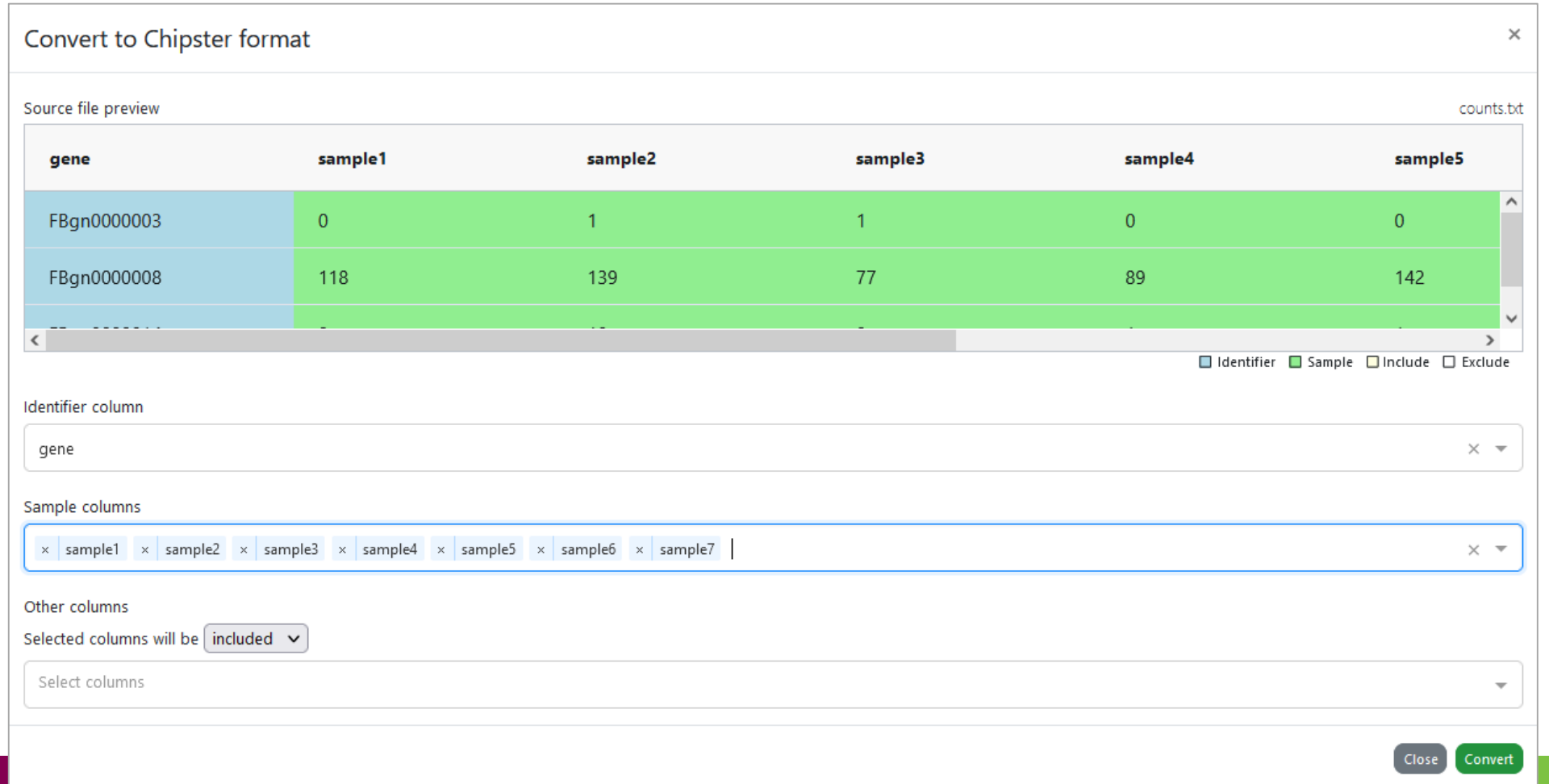
Run button gives several options:

- 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.



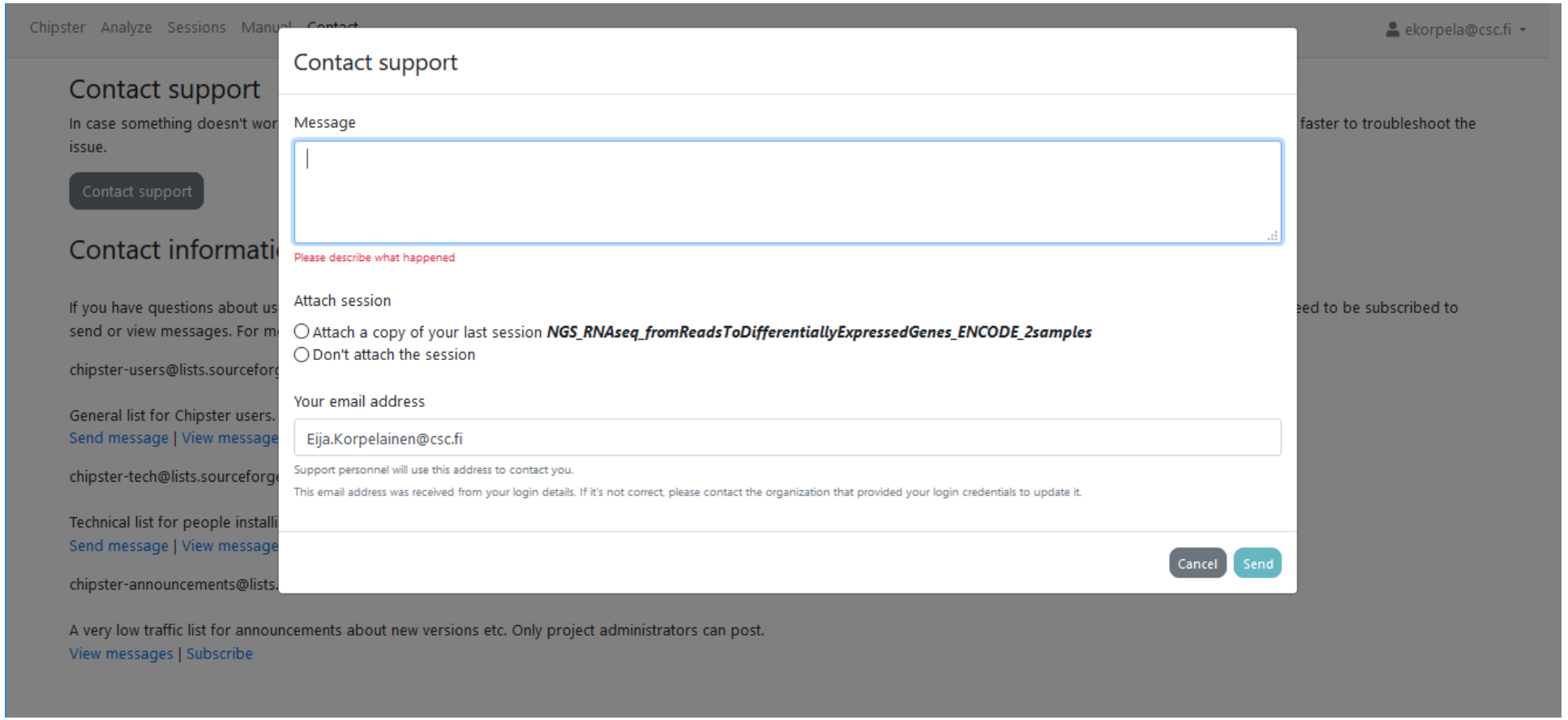
Converting a read count table to Chipster format

- 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



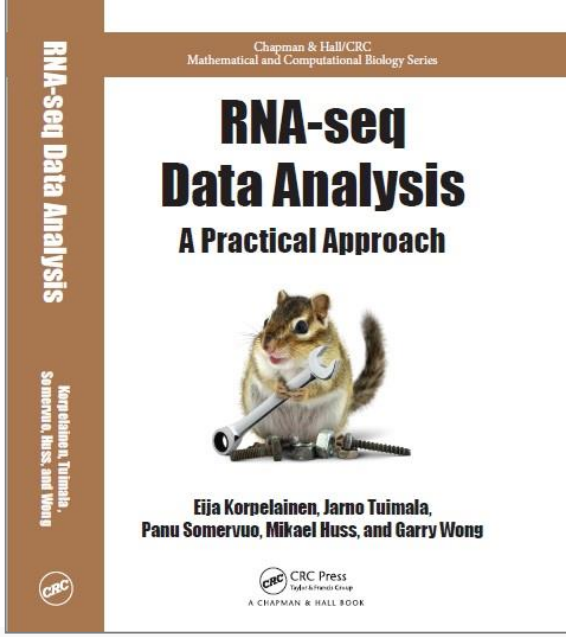
Problems? Send us a support request

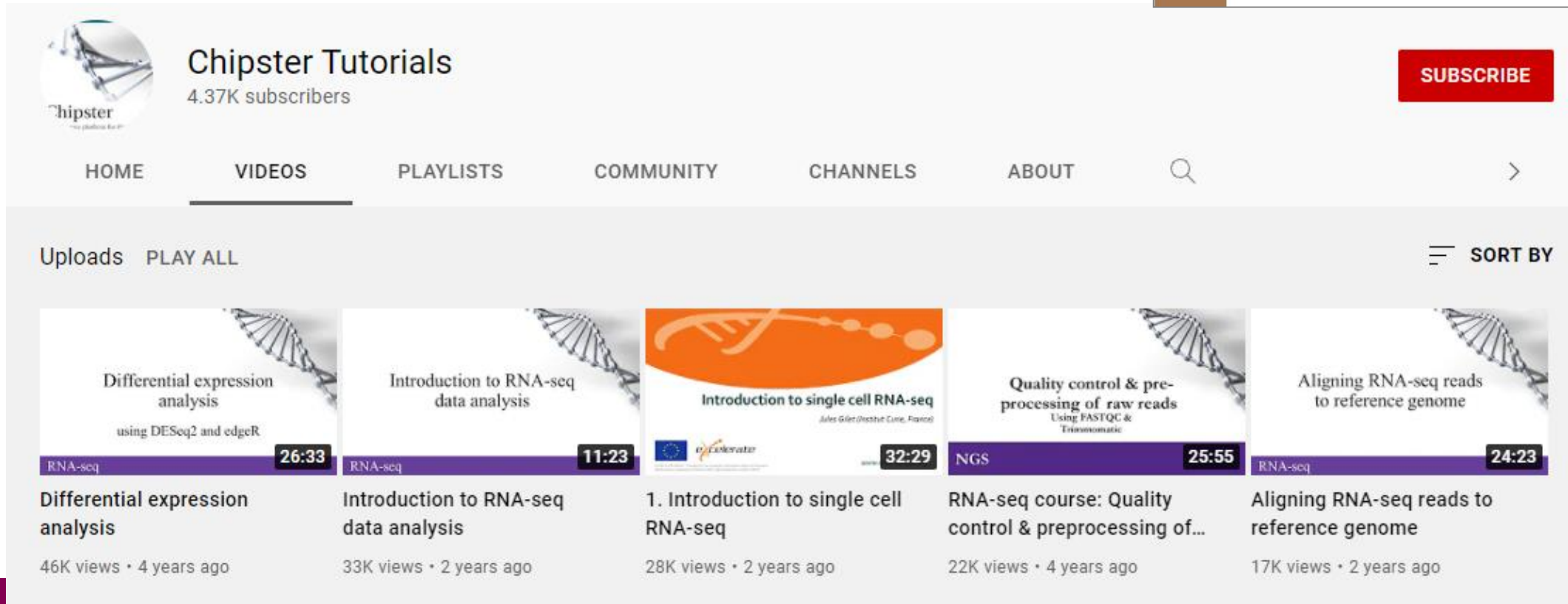
-request includes the error message and link to analysis session (optional)



More info

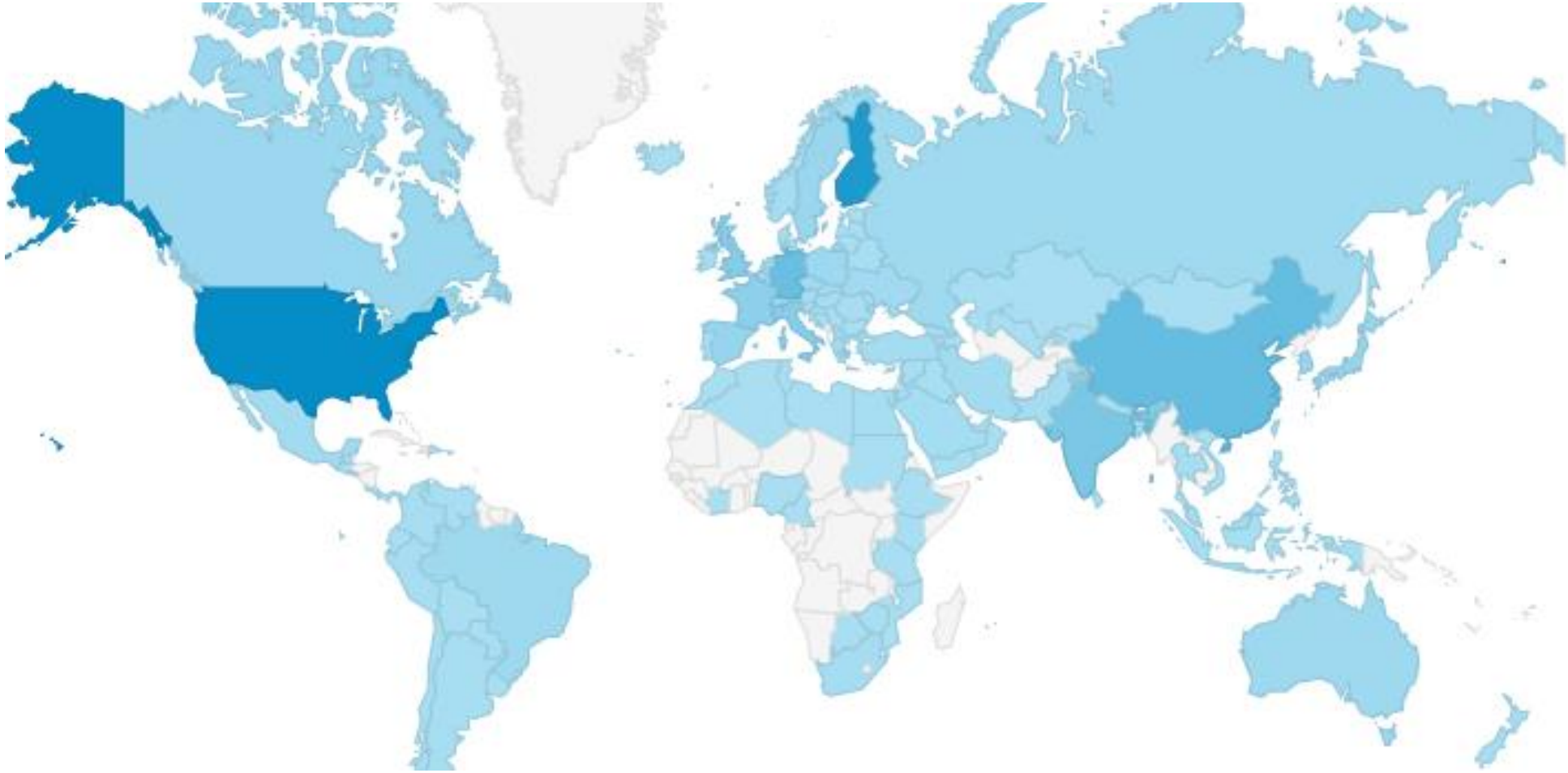
- chipster@csc.fi
- http://chipster.csc.fi
- Chipster tutorials in YouTube
- https://chipster.csc.fi/manual/courses.html





Acknowledgements to Chipster users and contibutors

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