

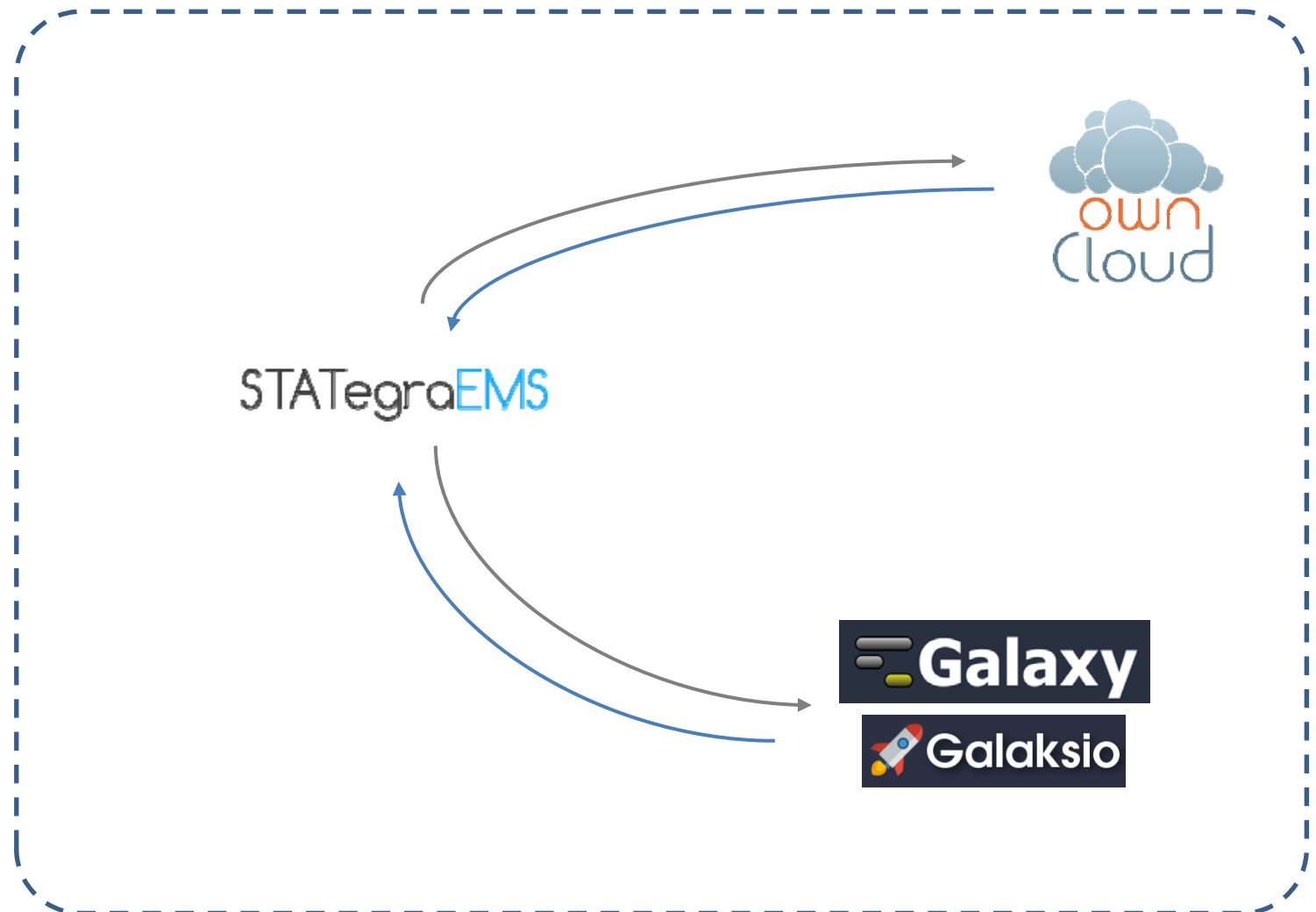
Overview

- **Galaxy** : workflow management system.
- **Galaksio**: a simplified interface for Galaxy.
- **STATegra EMS**: a *log-book* for biomedical experiments.
OwnCloud: a personal cloud that stores the files produced or used in the experiments.

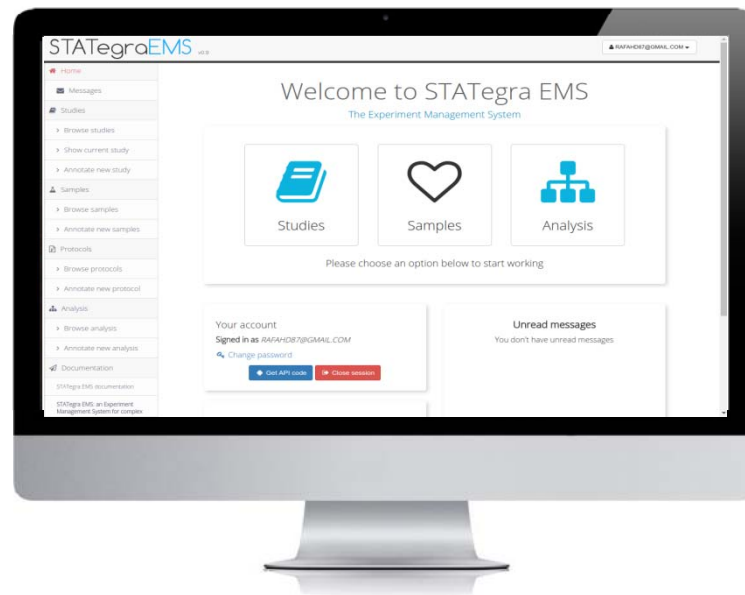
Overview



eB3Kit



The STATegra EMS



The STATegra EMS

What for?

- *Easy-to-use* for tracking the **provenance** of files.
- An simple use case:
 - Your lab participates in different genomic projects.
- Using the EMS, users can store the files together with a complete description of how they were produced and stored.

The STATegra EMS

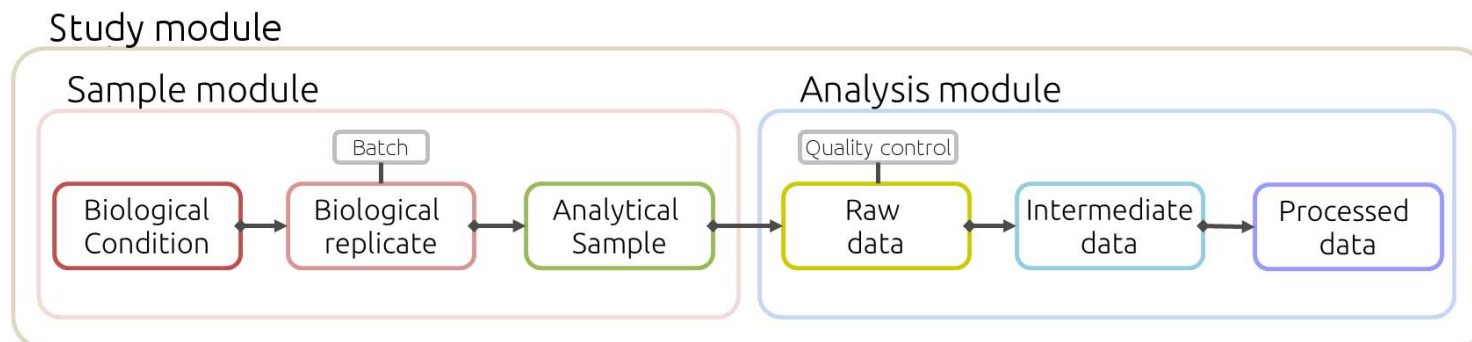
Features

- Data storage + meta-data annotation for sequencing, proteomics, and metabolomics experiments.
- Allows to track the **provenance** of files.
- Enables **reproducibility** of your experiments.
- Uses data standards for meta-data annotation to facilitate interoperability (MIAME, MINSEQE, MIABIS, MIAPE,...).
- Centralized web-service.
- Integrated with other tools such as Galaxy.
- Free to use, open-source.

The STATegra EMS

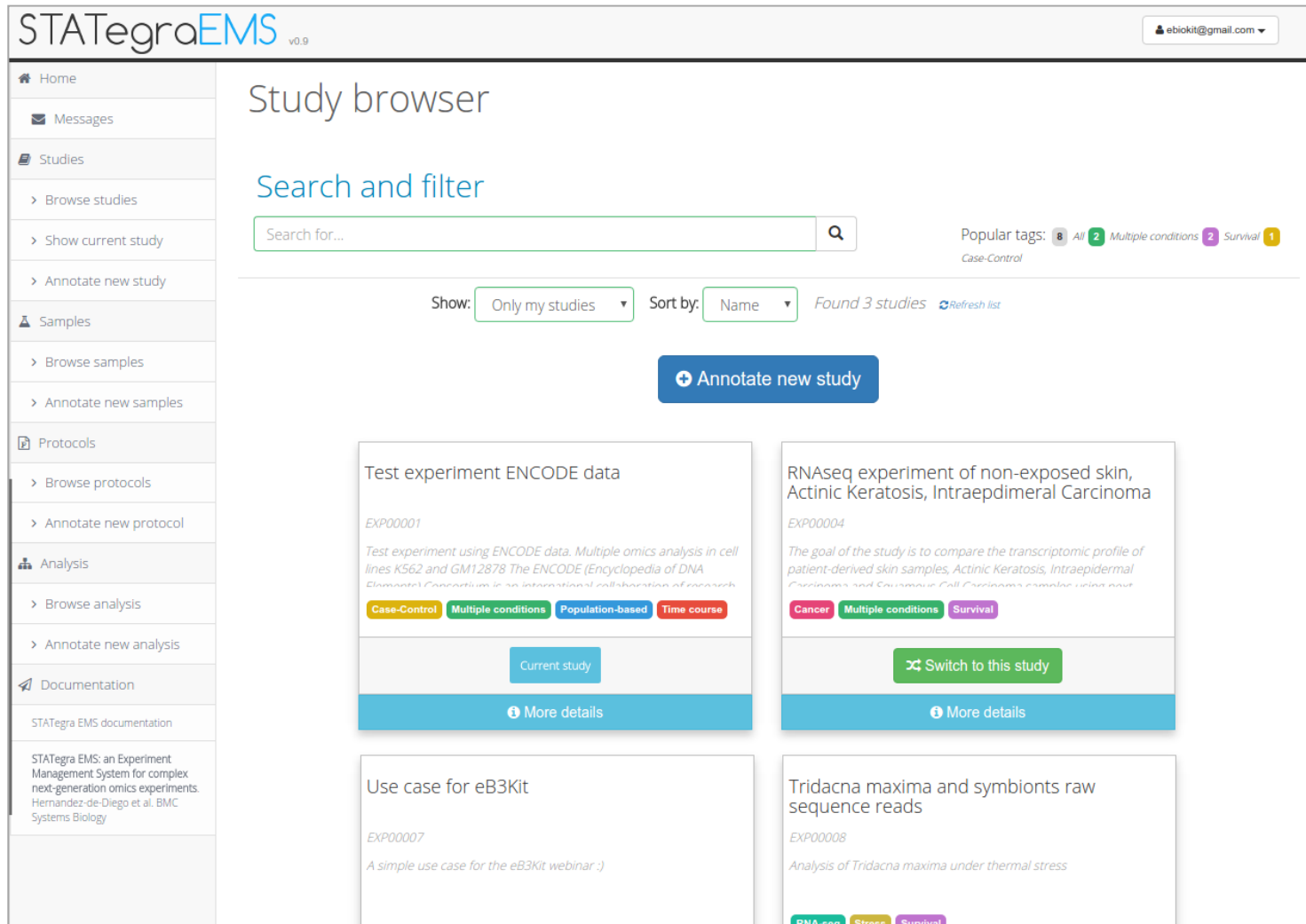
Some quick concepts

- ▶ Meta-data organized around 3 information units (IU):
 - ♦ Study
 - ♦ Samples
 - ♦ Analysis



The STATegra EMS

Study annotation (i)



STATegraEMS v0.9

Home Messages Studies Browse studies Show current study Annotate new study Samples Browse samples Annotate new samples Protocols Browse protocols Annotate new protocol Analysis Browse analysis Annotate new analysis Documentation

STATegra EMS documentation

STATegra EMS: an Experiment Management System for complex next-generation omics experiments. Hernandez de-Diego et al. BMC Systems Biology

Study browser

Search and filter

Search for...

Popular tags: 8 All 2 Multiple conditions 2 Survival 1 Case-Control

Show: Only my studies Sort by: Name Found 3 studies Refresh list

Annotate new study

Test experiment ENCODE data

EXP00001

Test experiment using ENCODE data. Multiple omics analysis in cell lines K562 and GM12878. The ENCODE (Encyclopedia of DNA Elements) Consortium is an international collaboration of researchers.

Case-Control Multiple conditions Population-based Time course

Current study

More details

RNAseq experiment of non-exposed skin, Actinic Keratosis, Intraepidermal Carcinoma

EXP00004

The goal of the study is to compare the transcriptomic profile of patient-derived skin samples, Actinic Keratosis, Intraepidermal Carcinoma, and Cutaneous Cell Carcinoma, similar using next.

Cancer Multiple conditions Survival

Switch to this study

More details

Use case for eB3Kit

EXP00007

A simple use case for the eB3Kit webinar :)

Tridacna maxima and symbionts raw sequence reads

EXP00008

Analysis of Tridacna maxima under thermal stress

RNA-seq Stress Survival

The STATegra EMS

Study annotation (ii)

Study details.

Study

General details

Study ID

EXP00001

Title

Test experiment ENCODE data

Description

Test experiment using ENCODE data. Multiple omics analysis in cell lines K562 and GM12878

The ENCODE (Encyclopedia of DNA Elements) Consortium is an international collaboration of research groups funded by the National Human Genome Research Institute (NHGRI). The goal of

Public references

<https://www.encodeproject.org>
<https://genome.ucsc.edu/ENCODE/>
<http://www.genome.gov/encode/>

Data storage

Data directory

FTP server

Host

hgdownload.soe.ucsc.edu

Port number

21

User

ucscftp

Password

.....

Path

/data/ucsc/encode_project/

Experimental design

Study tags

Case-Control

Multiple conditions

Time course

Population-based

Other information

General details

Title of the study, main objectives, URLs,...

Data storage

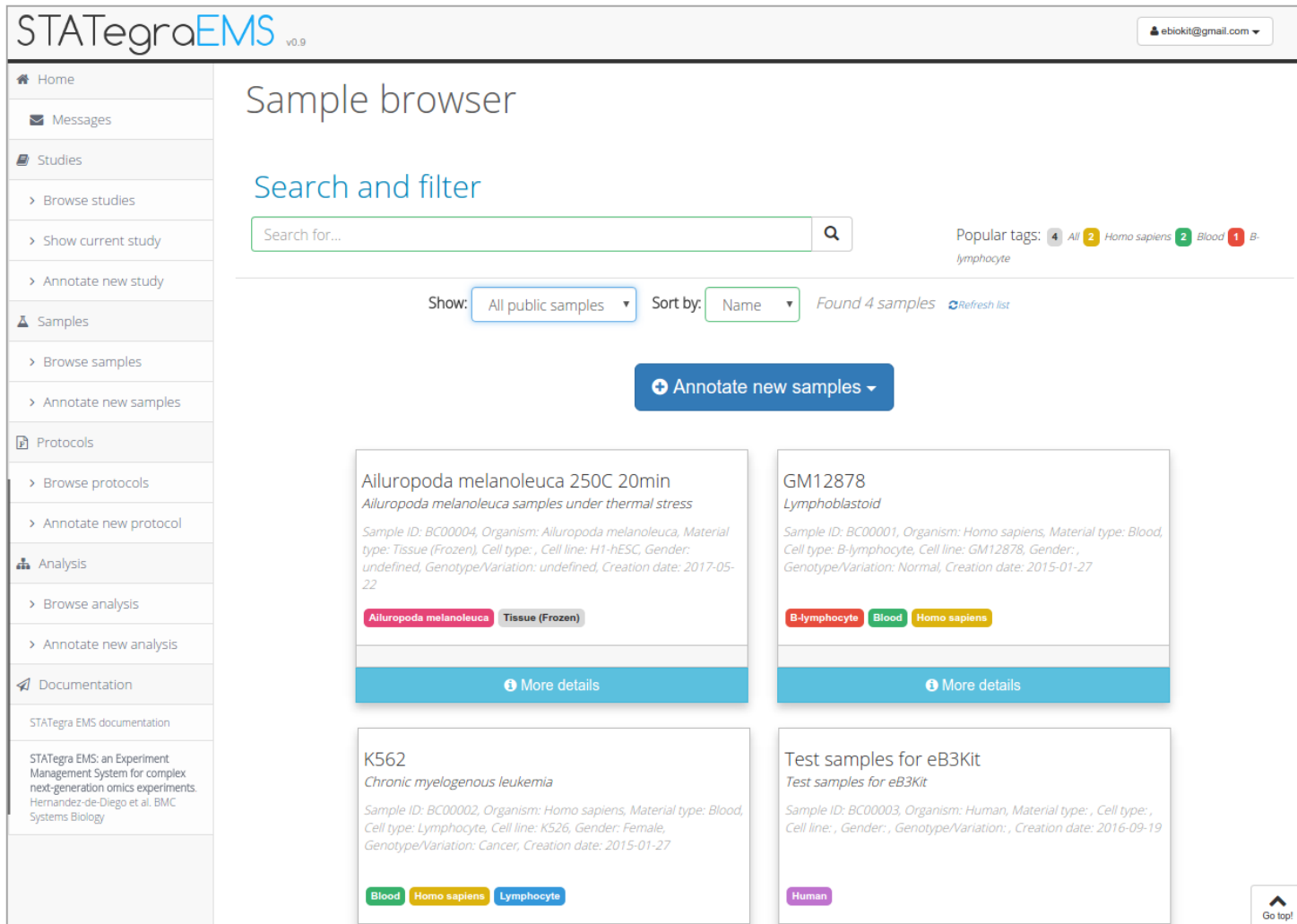
Support for external and local storage systems (e.g. FTP)

Experimental design

(type of study, planned measurements, etc.)

The STATegra EMS

Samples annotation (i)



The screenshot displays the STATegra EMS v0.9 interface. On the left is a sidebar with navigation links: Home, Messages, Studies (with sub-links for Browse studies, Show current study, and Annotate new study), Samples (with sub-links for Browse samples and Annotate new samples), Protocols (with sub-links for Browse protocols and Annotate new protocol), Analysis (with sub-links for Browse analysis and Annotate new analysis), and Documentation. The main content area is titled 'Sample browser' and includes a 'Search and filter' section with a search input field and a 'Popular tags' section showing counts for 'All' (4), 'Homo sapiens' (2), 'Blood' (2), and 'B-lymphocyte' (1). Below this, filters are set to 'Show: All public samples' and 'Sort by: Name', indicating 'Found 4 samples'. A button 'Annotate new samples' is present. Four sample cards are displayed, each with a title, description, sample ID, organism, material type, cell type, and creation date, along with tags and a 'More details' button.

Sample 1: Ailuropoda melanoleuca 250C 20min
Ailuropoda melanoleuca samples under thermal stress
Sample ID: BC00004, Organism: Ailuropoda melanoleuca, Material type: Tissue (Frozen), Cell type: , Cell line: H1-hESC, Gender: undefined, Genotype/Variation: undefined, Creation date: 2017-05-22
Tags: Ailuropoda melanoleuca, Tissue (Frozen)

Sample 2: GM12878
Lymphoblastoid
Sample ID: BC00001, Organism: Homo sapiens, Material type: Blood, Cell type: B-lymphocyte, Cell line: GM12878, Gender: , Genotype/Variation: Normal, Creation date: 2015-01-27
Tags: B-lymphocyte, Blood, Homo sapiens

Sample 3: K562
Chronic myelogenous leukemia
Sample ID: BC00002, Organism: Homo sapiens, Material type: Blood, Cell type: Lymphocyte, Cell line: K562, Gender: Female, Genotype/Variation: Cancer, Creation date: 2015-01-27
Tags: Blood, Homo sapiens, Lymphocyte

Sample 4: Test samples for eB3Kit
Test samples for eB3Kit
Sample ID: BC00003, Organism: Human, Material type: , Cell type: , Cell line: , Gender: , Genotype/Variation: , Creation date: 2016-09-19
Tags: Human

The STATegra EMS

Samples annotation (ii)

Sample details.

[Biological conditions](#)[Samples](#)

Biological conditions details

General details

BioCondition ID BC00001

Name GM12878

Title Lymphoblastoid

Sample tags

Homo sapiens

Blood

B-lymphocyte

Biomaterial details

Organism Homo sapiens

Material type Blood

Cell type B-lymphocyte

Cell line GM12878

Gender Not specified

Genotype/Variation Normal

Other details

International HapMap Project - CEPH/Utah - European Caucasian, Epstein-Barr Virus

Experimental conditions

Treatment Not specified

General details

Name, Title, tags, ...

Details for biomaterial

Organism, material type, cell line and type, ...

Experimental conditions

Treatments, dose, time,...

The STATegra EMS

Samples annotation (iii)

Sample details.

[♥ Biological conditions](#)[🔥 Samples](#)

Biological samples grown in these conditions.

Sample GM12878_R1

Sample ID BR00001.001



Sample name GM12878_R1



Aliquots

1 aliquots were extracted from this sample using 1 different extraction protocols.

Aliquots for protocol Prot_In-gel digestion

| Aliquot ID | Aliquot name |
|-----------------|--------------|
| AR00001.001.001 | GM12878_R1 |

Sample GM12878_R2

Sample ID BR00001.002



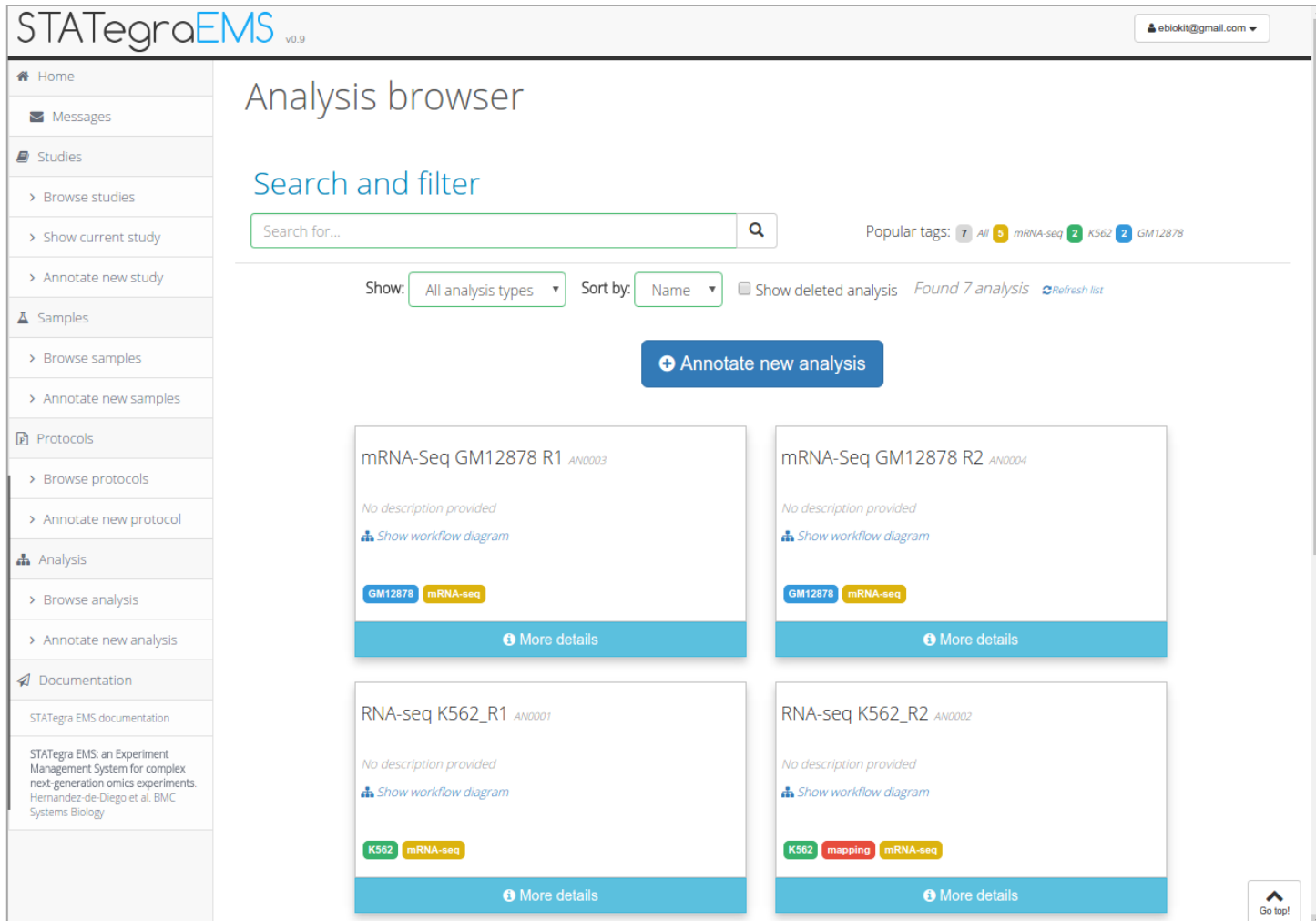
Sample name GM12878_R2



Details for samples grown
under the biological condition
Replicates, aliquots, protocols
for sample isolation...

The STATegra EMS

Analysis annotation (i)



STATegraEMS v0.9

Home Messages Studies Browse studies Show current study Annotate new study Samples Browse samples Annotate new samples Protocols Browse protocols Annotate new protocol Analysis Browse analysis Annotate new analysis Documentation STATegra EMS documentation

STATegra EMS: an Experiment Management System for complex next-generation omics experiments. Hernandez-de-Diego et al. BMC Systems Biology

Analysis browser

Search and filter

Search for...

Popular tags: 7 All 5 mRNA-seq 2 K562 2 GM12878

Show: All analysis types Sort by: Name Show deleted analysis Found 7 analysis Refresh list

Annotate new analysis

mRNA-Seq GM12878 R1 AN0003

No description provided

Show workflow diagram

GM12878 mRNA-seq

More details

mRNA-Seq GM12878 R2 AN0004

No description provided

Show workflow diagram

GM12878 mRNA-seq

More details

RNA-seq K562_R1 AN0001

No description provided

Show workflow diagram

K562 mRNA-seq

More details

RNA-seq K562_R2 AN0002

No description provided

Show workflow diagram

K562 mapping mRNA-seq

More details

Go top!

The STATegra EMS

Analysis annotation (ii)

Analysis details.

[Analysis overview](#)[Workflow steps](#)

Analysis details

Analysis ID AN0001

Analysis name RNA-seq K562_R1

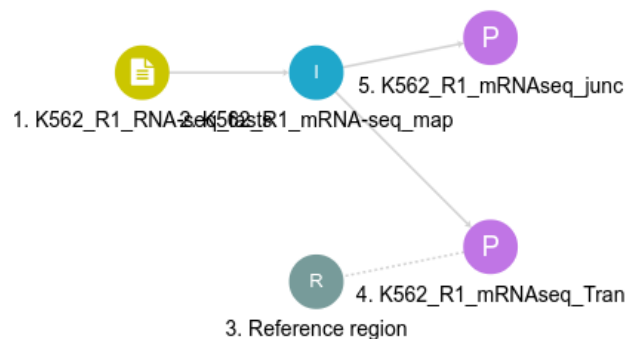
Analysis type mRNA-seq

Analysis tags

mRNA-seq

K562

Workflow overview

[Refresh diagram](#)[Analysis details](#)[Pipeline overview](#)

The STATegra EMS

Analysis annotation (iii)

Analysis details.

 Analysis overview

 Workflow steps

 K562_R1_RNA-seq_fasts ✕

mRNA-seq details

General details

Step name K562_R1_RNA-seq_fasts ?

Technology mRNA-seq ?

Output files <http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeCaltechRnaSeq/wgEncodeCaltec> ?

Step ID ST0001.001 ?

Submission date 2013/07/11 ?

Last edition 2015/01/27 ?

Owners

testuser1 ?

Biomaterial details

Biological condition title: Chronic myelogenous leukemia

Sample name: K562_R1

Aliquot name: K562_RNA_R1

Equipment details

Platform Family Solexa Illumina ?

Platform Model Solexa Illumina Genome Analyzer II ?

Steps details

Name, sequencing platform details, software used for processing, files location,...

The STATegra EMS

More information

Hernández de Diego et al. *BMC Systems Biology* 2014, **8**(Suppl 2):S9
<http://www.biomedcentral.com/1752-0509/8/S2/S9>



RESEARCH

Open Access

STATegra EMS: an Experiment Management System for complex next-generation omics experiments

Rafael Hernández de Diego¹, Noemi Boix-Chova², David Gómez-Cabrero², Jesper Tegner², Imad Abugessaisa², Ana Conesa^{3*}

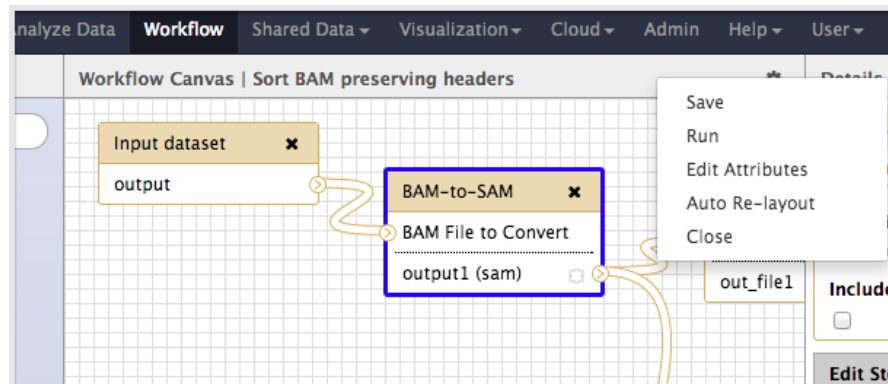
From High-Throughput Omics and Data Integration Workshop
Barcelona, Spain. 13-15 February 2013

Hernandez-de-Diego et al. *BMC Systems Biology* 2014 **8**(Suppl 2):S9
DOI: 10.1186/1752-0509-8-S2-S9

Galaxy

Some quick concepts

- In Galaxy, a workflow is a series of tools and dataset actions that run in sequence as a batch operation.
- Workflows can be created from scratch using the workflow editor.
- Bioinformatics skills are required in order to define the workflow and choose the right values for parameters.



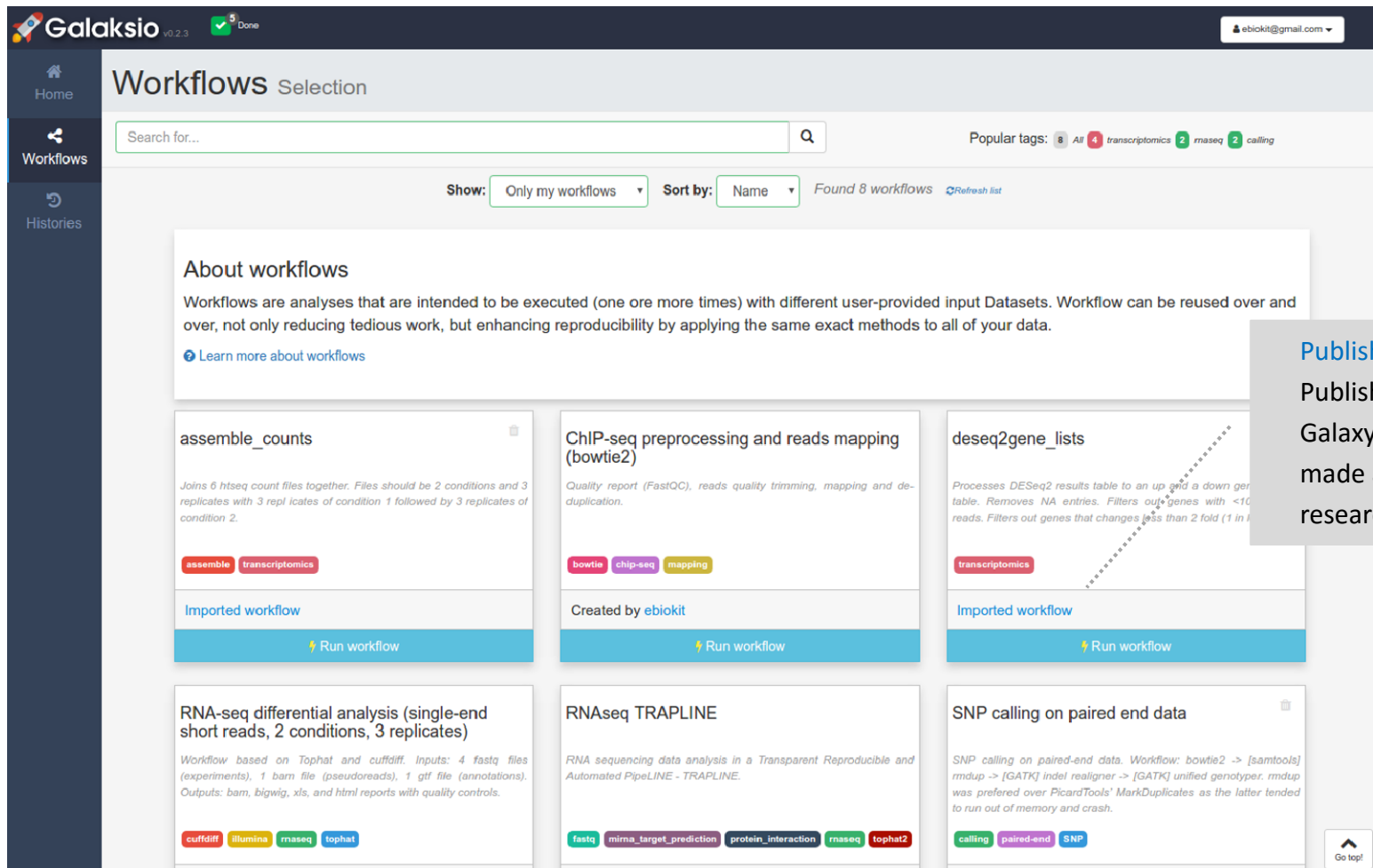
Galaksio

Some quick concepts

- Workflow-centric interface for using Galaxy
- Tailored for researchers who require bioinformatics workflows to complete research tasks without extensive training in programming or computational biology.

Galaksio

Some quick concepts



The screenshot shows the Galaksio interface for workflow selection. The top navigation bar includes 'Home', 'Workflows', and 'Histories'. The main content area is titled 'Workflows Selection' and features a search bar, a 'Popular tags' section with filters for 'All', 'transcriptomics', 'maseq', and 'calling', and a 'Show' dropdown set to 'Only my workflows'. Below this, there are six workflow cards arranged in a 2x3 grid. Each card includes a title, a brief description, a list of associated tags, the creator's name, and a 'Run workflow' button. The workflows shown are: 'assemble_counts', 'ChIP-seq preprocessing and reads mapping (bowtie2)', 'deseq2gene_lists', 'RNA-seq differential analysis (single-end short reads, 2 conditions, 3 replicates)', 'RNAseq TRAPLINE', and 'SNP calling on paired end data'.

About workflows

Workflows are analyses that are intended to be executed (one or more times) with different user-provided input Datasets. Workflow can be reused over and over, not only reducing tedious work, but enhancing reproducibility by applying the same exact methods to all of your data.

[Learn more about workflows](#)

Published workflows

Published workflows in Galaxy are automatically made accessible to researchers in Galaksio.

Galaksio

Some quick concepts

The screenshot displays the Galaksio v0.2.3 web interface. The top navigation bar includes the Galaksio logo, version information, a 'Done' status, and a user profile dropdown for 'ebiokit@gmail.com'. The left sidebar contains links for 'Home', 'Workflows', and 'Histories'. The main content area is titled 'Workflows Run workflow' and features a progress bar with six steps: 1. Workflow overview (active), 2. History selection, 3. Input selection, 4. Custom parameters, 5. Run workflow, and 6. Results.

Below the progress bar, the 'Details' section for the workflow 'Running ChIP-seq preprocessing and reads mapping (bowtie2)' is shown, with author 'ebiokit' and description 'Quality report (FastQC), reads quality trimming, mapping and de-duplication.' The 'Workflow overview' section displays a flowchart with five steps: 1. Input dataset (red circle), 2. FASTQ Groomer (blue circle), 3. FastQC (green circle), 4. FASTQ Quality Trimmer (purple circle), and 5. Bowtie2 (orange circle). Arrows indicate the flow from step 1 to 2, 2 to 3, 2 to 4, and 4 to 5. A dotted line connects step 5 to the 'Next step' button.

At the bottom of the interface, there are two buttons: '< Cancel and go back' and '> Next step'.

Quick and easy

Workflows can be executed easily in just few clicks. Researchers only need to provide the input data and, optionally, adapt the parameters for the steps in the workflow.