

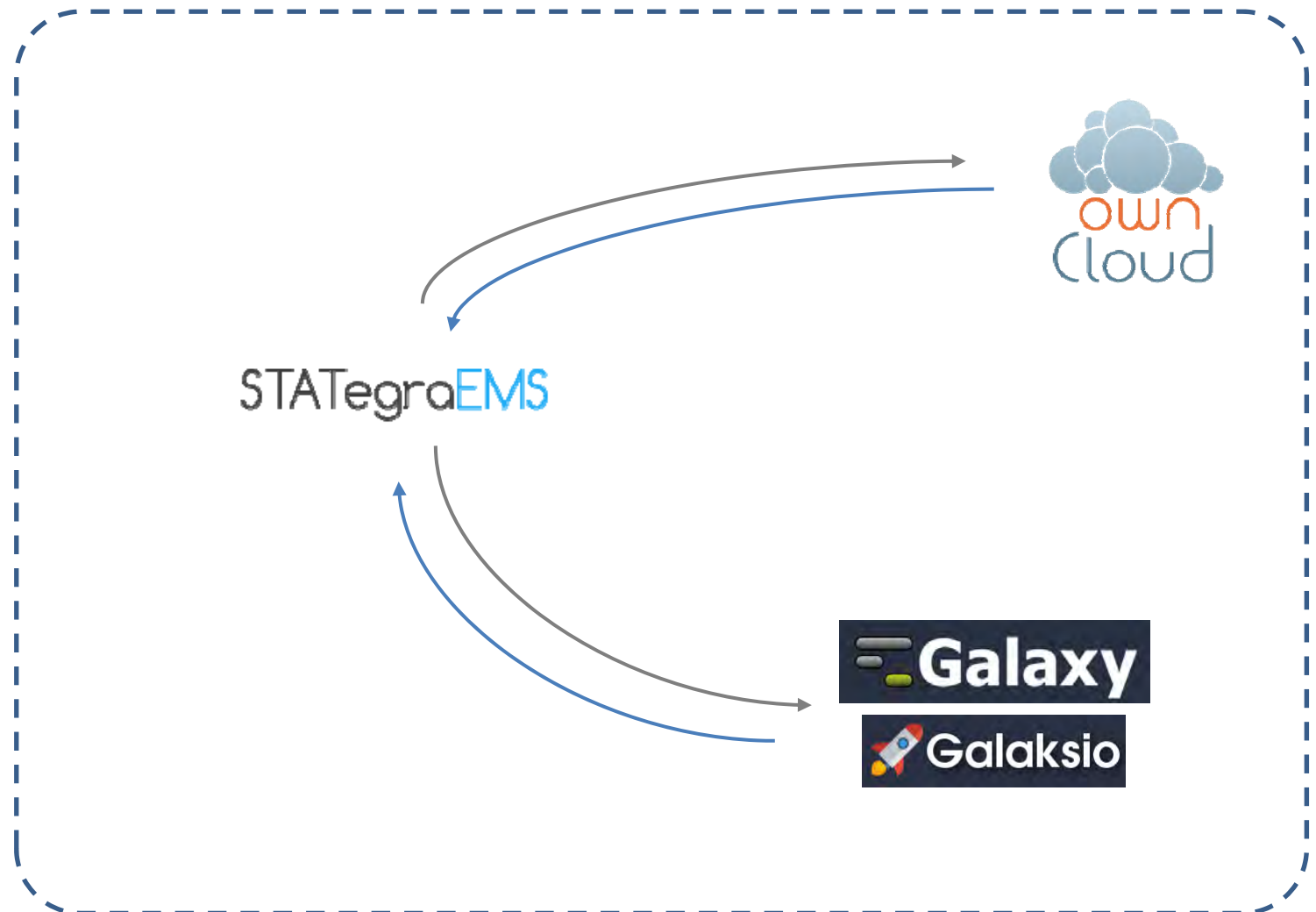
Overview

- **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.

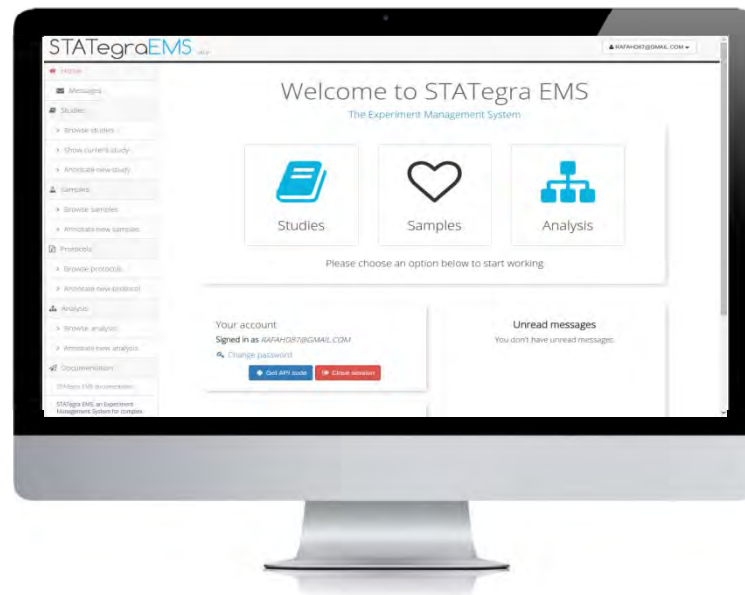
Overview



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

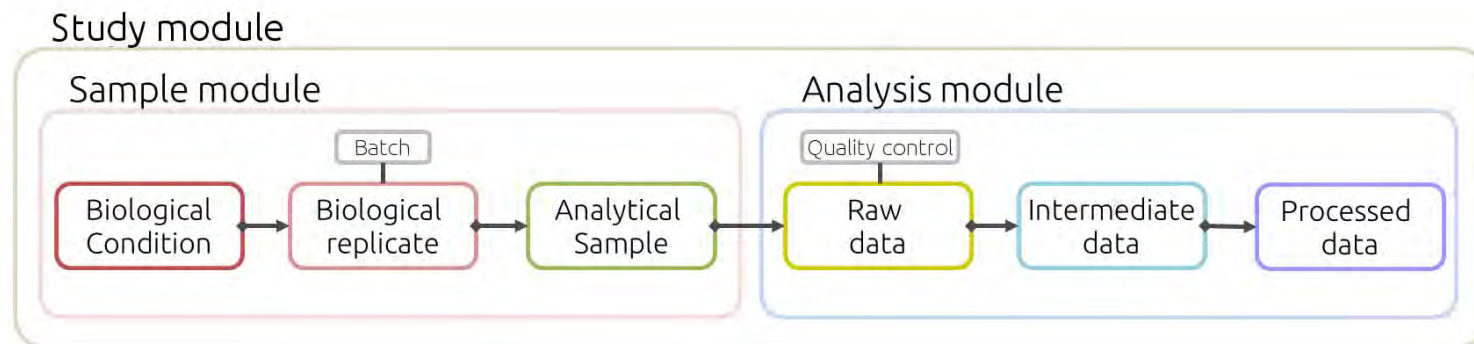
Features

- 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

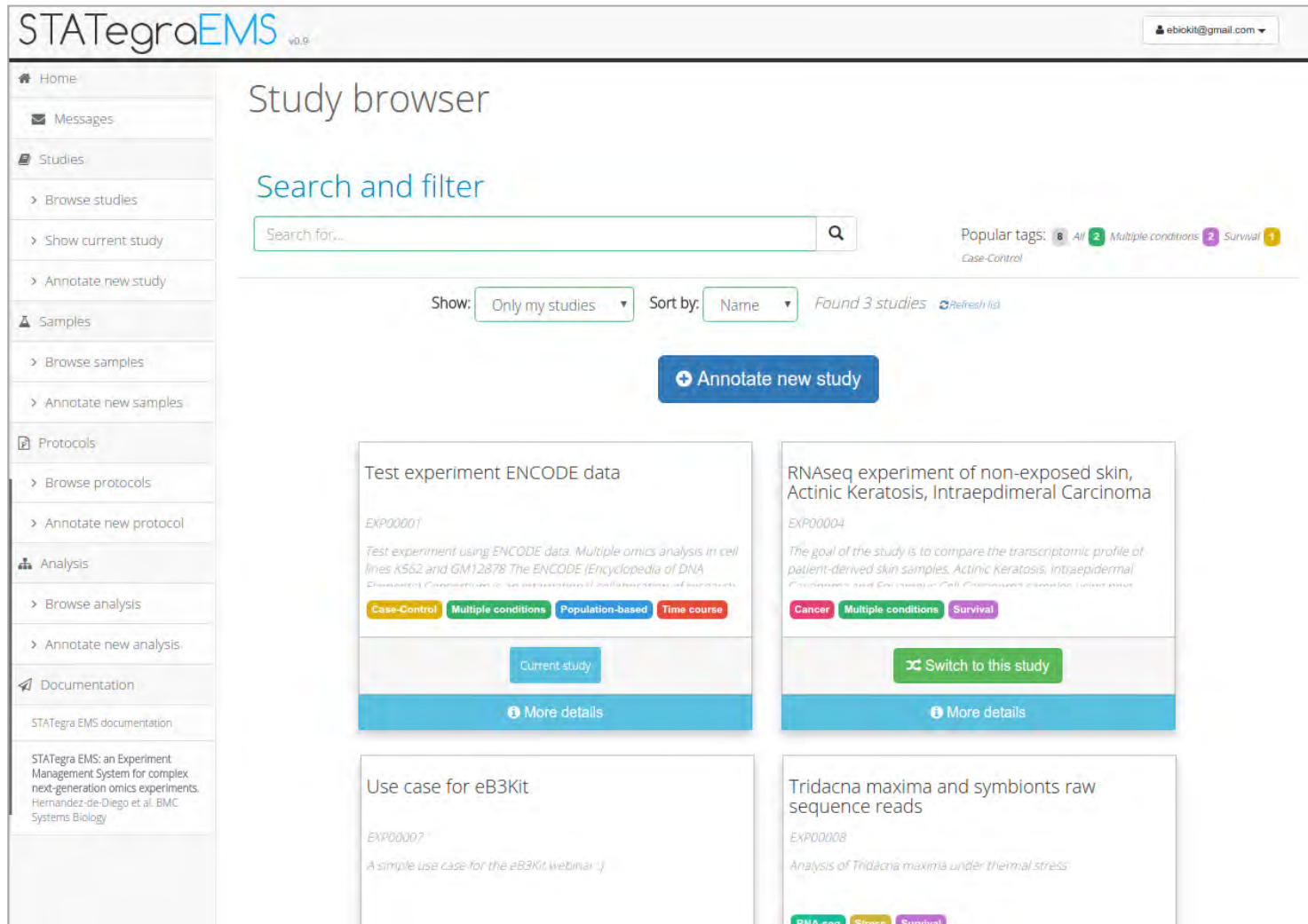
Some quick concepts

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



The STATegra EMS

Study annotation (i)



The screenshot displays the STATegraEMS v0.9 web interface. On the left is a sidebar menu with sections: Home, Messages, Studies (with sub-items: Browse studies, Show current study, Annotate new study), Samples (with sub-items: Browse samples, Annotate new samples), Protocols (with sub-items: Browse protocols, Annotate new protocol), Analysis (with sub-items: Browse analysis, Annotate new analysis), and Documentation (with sub-item: STATegra EMS documentation). The main content area is titled 'Study browser' and includes a 'Search and filter' section with a search bar and a 'Popular tags' section showing 'Case-Control' (8), 'All' (2), 'Multiple conditions' (2), and 'Survival' (1). Below this are filters for 'Show: Only my studies' and 'Sort by: Name', indicating 'Found 3 studies'. A blue button 'Annotate new study' is present. The interface shows four study cards: 1. 'Test experiment ENCODE data' (EXP00001) with tags 'Case-Control', 'Multiple conditions', 'Population-based', and 'Time course', and a 'Current study' button. 2. 'RNAseq experiment of non-exposed skin, Actinic Keratosis, Intraepidermal Carcinoma' (EXP00004) with tags 'Cancer', 'Multiple conditions', and 'Survival', and a 'Switch to this study' button. 3. 'Use case for eB3Kit' (EXP00007) with a description 'A simple use case for the eB3Kit webinar :)' and no tags. 4. 'Tridacna maxima and symbionts raw sequence reads' (EXP00008) with tags 'RNA-seq', 'Stress', and 'Survival', and a description 'Analysis of Tridacna maxima under thermal stress'.

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

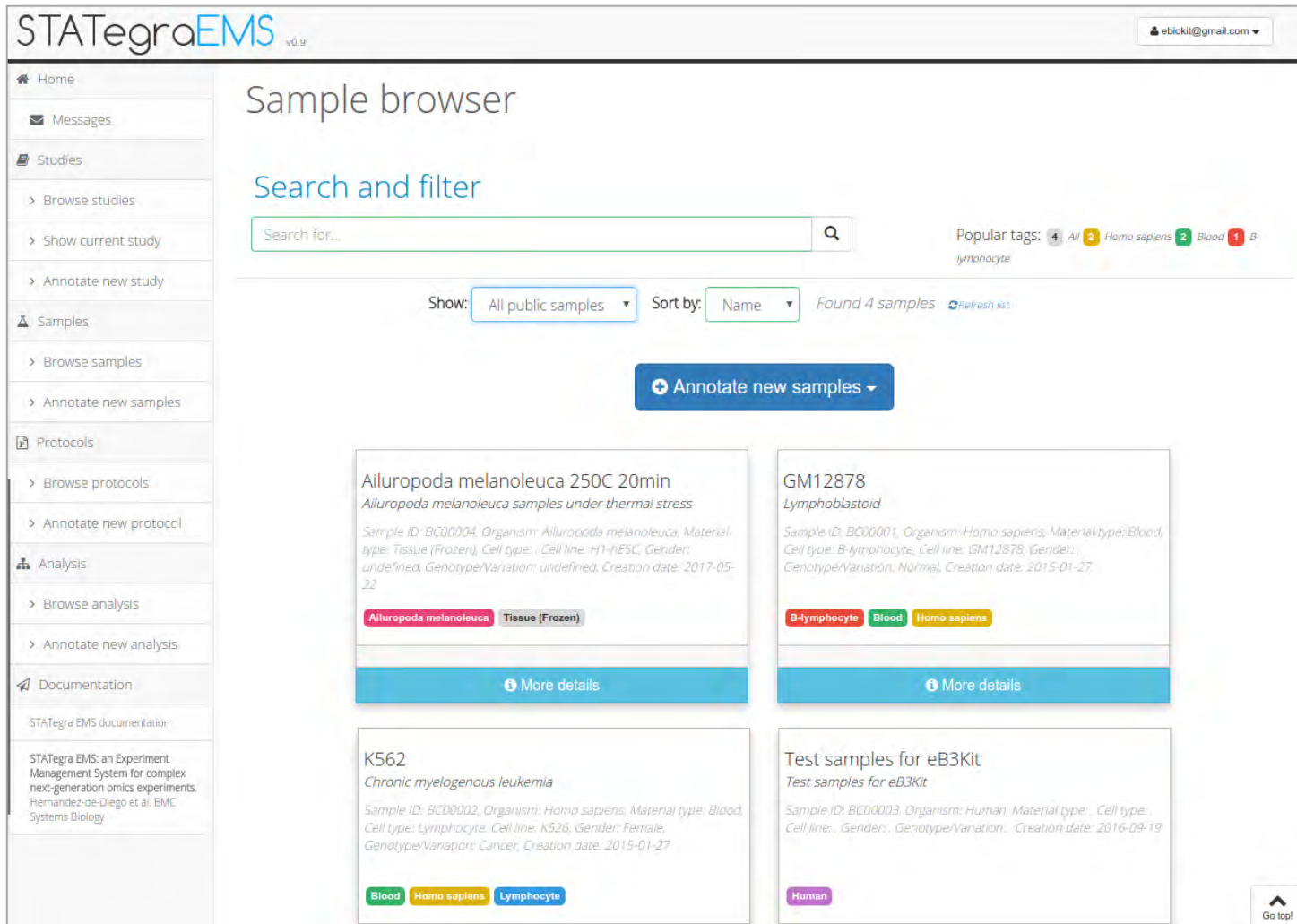
Support for external and local storage systems (e.g. FTP)

Experimental design

(type of study, planned measurements, etc.)

The STATegra EMS

Samples annotation (i)



STATegraEMS v0.9 ebiokit@gmail.com

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

Sample browser

Search and filter

Search for...

Popular tags: 4 All 3 Homo sapiens 2 Blood 1 B-lymphocyte

Show: All public samples Sort by: Name Found 4 samples Refresh list

Annotate new samples

Ailuropoda melanoleuca 250C 20min
Ailuropoda melanoleuca samples under thermal stress

Sample ID: BC00004, Organism: Ailuropoda melanoleuca, Material type: Tissue (Frozen), Cell type: H1-hESC, Gender: undefined, Genotype/Variation: undefined, Creation date: 2017-05-22

Ailuropoda melanoleuca Tissue (Frozen)

More details

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

B-lymphocyte Blood Homo sapiens

More details

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

Blood Homo sapiens Lymphocyte

More details

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

Human

More details

Go top

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

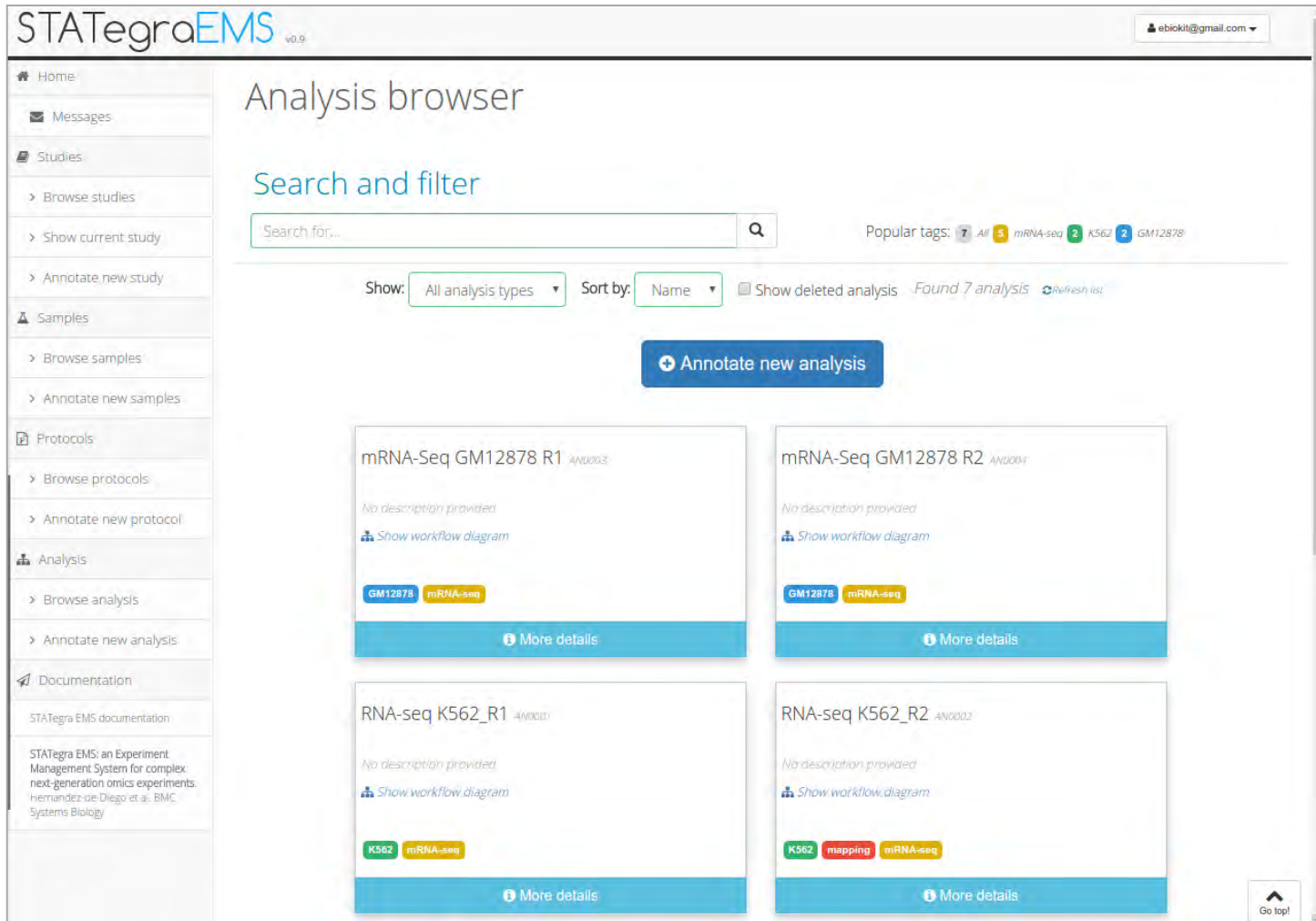
Sample ID BR00001.002

Sample name GM12878_R2

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

The STATegra EMS

Analysis annotation (i)



The screenshot displays the STATegraEMS v0.9 web application interface. On the left is a sidebar menu with categories: Home, Messages, Studies (with sub-items: Browse studies, Show current study, Annotate new study), Samples (with sub-items: Browse samples, Annotate new samples), Protocols (with sub-items: Browse protocols, Annotate new protocol), Analysis (with sub-items: Browse analysis, Annotate new analysis), and Documentation (with sub-item: STATegra EMS documentation). The main content area is titled 'Analysis browser' and includes a 'Search and filter' section with a search bar and popular tags (7 All, 5 mRNA-seq, 2 K562, 2 GM12878). Below the search bar are filters for 'Show: All analysis types' and 'Sort by: Name', along with a checkbox for 'Show deleted analysis' and a message 'Found 7 analysis'. A blue button labeled 'Annotate new analysis' is positioned above a grid of four analysis cards. The cards are: 'mRNA-Seq GM12878 R1' (AN0003), 'mRNA-Seq GM12878 R2' (AN0004), 'RNA-seq K562_R1' (AN0001), and 'RNA-seq K562_R2' (AN0002). Each card shows 'No description provided', a 'Show workflow diagram' link, and tags (e.g., GM12878, mRNA-seq, K562, mapping, mRNA-seq). Each card also has a 'More details' link at the bottom. The bottom right corner features a 'Go top!' button.

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

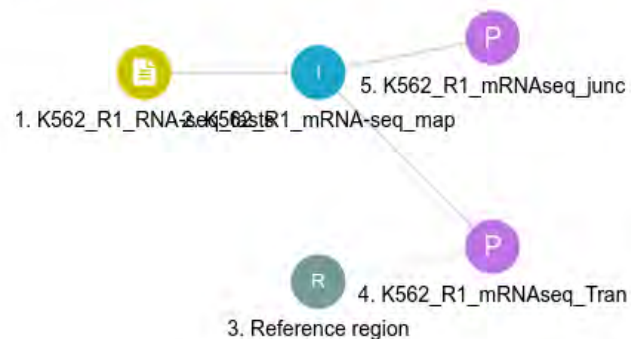
Analysis tags

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/wgEncodeCaltech	ⓘ
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*}

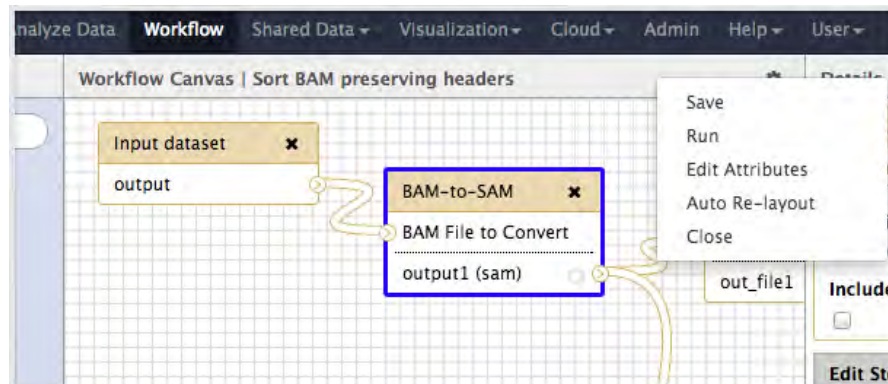
From High-Throughput Omics and Data Integration Workshop
Barcelona, Spain. 13-15 February 2013

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.



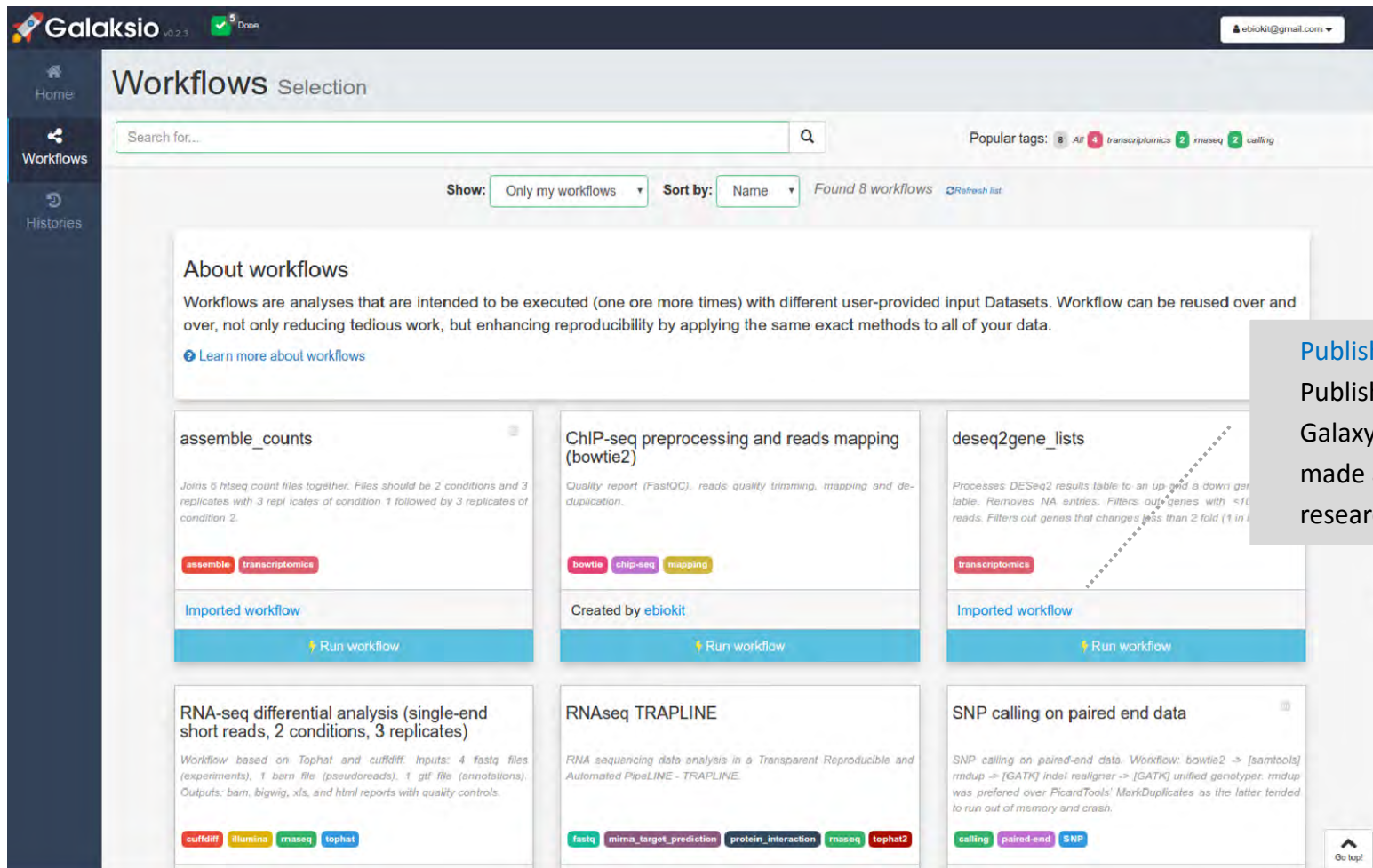
Galaksio

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 v0.2.3 interface. The top navigation bar includes 'Home', 'Workflows', and 'Histories'. The 'Workflows' section is active, displaying a 'Workflows Selection' page. A search bar is present, along with filters for 'Show: Only my workflows' and 'Sort by: Name'. A 'Popular tags' section lists 'All', 'transcriptomics', 'rnaseq', and 'calling'. Below this, a grid of workflow cards is shown. Each card includes a title, a brief description, a list of tags, the creator's name, and a 'Run workflow' button. The workflows displayed 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 web interface. At the top, the Galaksio logo and version (0.2.3) are visible, along with a 'Done' button and a user profile dropdown. The left sidebar contains navigation 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 shows the author 'ebiokit' and a 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, there are buttons for 'Cancel and go back' and 'Next step'.

Running ChIP-seq preprocessing and reads mapping (bowtie2)

1 Workflow overview 2 History selection 3 Input selection 4 Custom parameters 5 Run workflow 6 Results

Details
Author: ebiokit
Description: Quality report (FastQC), reads quality trimming, mapping and de-duplication.

Workflow overview

1. Input dataset 2. FASTQ Groomer 3. FastQC 4. FASTQ Quality Trimmer 5. Bowtie2

< Cancel and go back > 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.