
A Galactic Experience and Extending Galaxy Sample Tracking

Jelle Scholtalbers

**Translational Oncology and Immunology (TRON)
Mainz, Germany**

Computational Medicine & Medical Genomics

Jelle Scholtalbers (jelle.scholtalbers@tron-mainz.de), John Castle, Michael Koslowski, Martin Löwer, Michael Rauer, Thomas Rösler, Ludmila Schemarow, Marius Byl, Thomas Bukur, Jos de Graaf, Valesca Boisguérin, Beate Stradmann-Bellinghausen, Ugur Sahin



Why Galaxy?

- Introducing lab technicians to bioinformatics
 - Needed an interface to command line scripts
- Workflows/Pipelines
 - Local tools
 - Keep data in-house
- Keeping track
 - Something better than `linux history`
 -

Tools and uses

- Simple converters & Text manipulation
 - `sort myUnsortedIDs.txt | uniq -d`
 - Search & Replace
 -
- Alignments
 - NCBI Blast
 - Basic until wrappers in repository became available
 - TimeLogic® DeCypher cmd tools
 - Nicer queuing, retrieve references using fastacmd, database creation
 -
- Send GFF to GMOD GBrowse

General comments

- Would like to have
 - Text editor
 - Expanded conditional statements
 - More advanced workflow functions
 - Loops
 - `Foreach 'split fasta' do 'alignments'`
 - Conditional
 -
- Most thankful for
 - History -> Workflow
 - **E-mail support!**
 - **galaxy-user & galaxy-dev@lists.bx.psu.edu**

Building Bridges

“ There does not exist a category of science to which one can give the name ‘*applied science*.’ There are science and the applications of science, bound together as the fruit of the tree which bears it.”

Louis Pasteur



Strengths:

- Innovation
- Basic research
- Applied research
- Clinical research
- Patient access

Needs:

- Translation of new concepts
- Patent exploitation
- Access to
 - Innovative medicines
 - Innovative diagnostics
 - Technology



Range of tasks

- Innovation
- Innovation maturation
- Implementation of ideas to Proof-of-Concept (PoC)
- Value add
- Link research and clinics
- Link academia and pharma
- Attract top talent / researchers
- Spin-offs

Biotechnology & Pharma

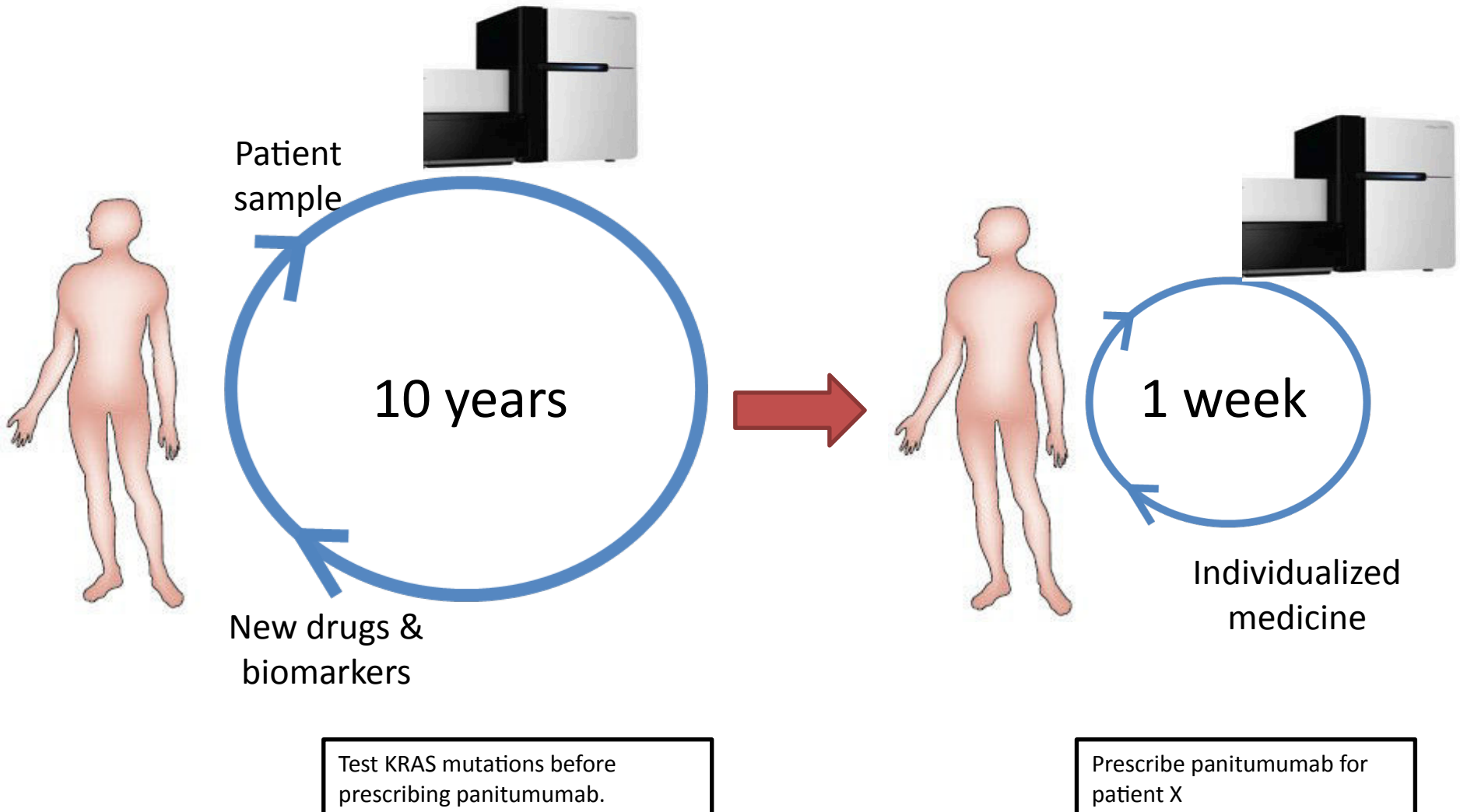
Strengths:

- Adequate budgets
- Process control
- Product focused
- Value add / markets

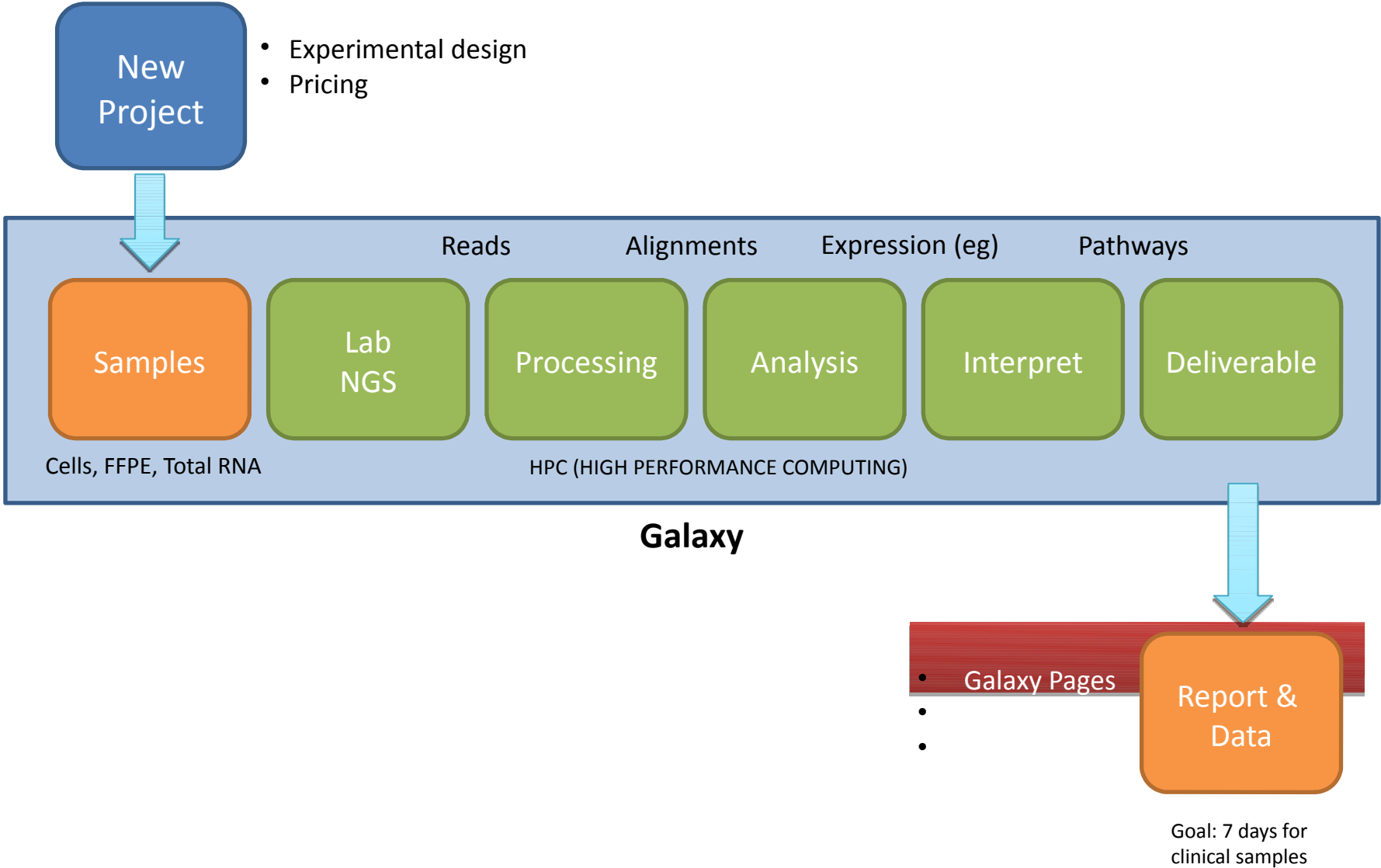
Needs

- Integration in excellent academic environment
- Access to mature innovations
- Patient-oriented research

Patient benefit using NGS



NGS Process

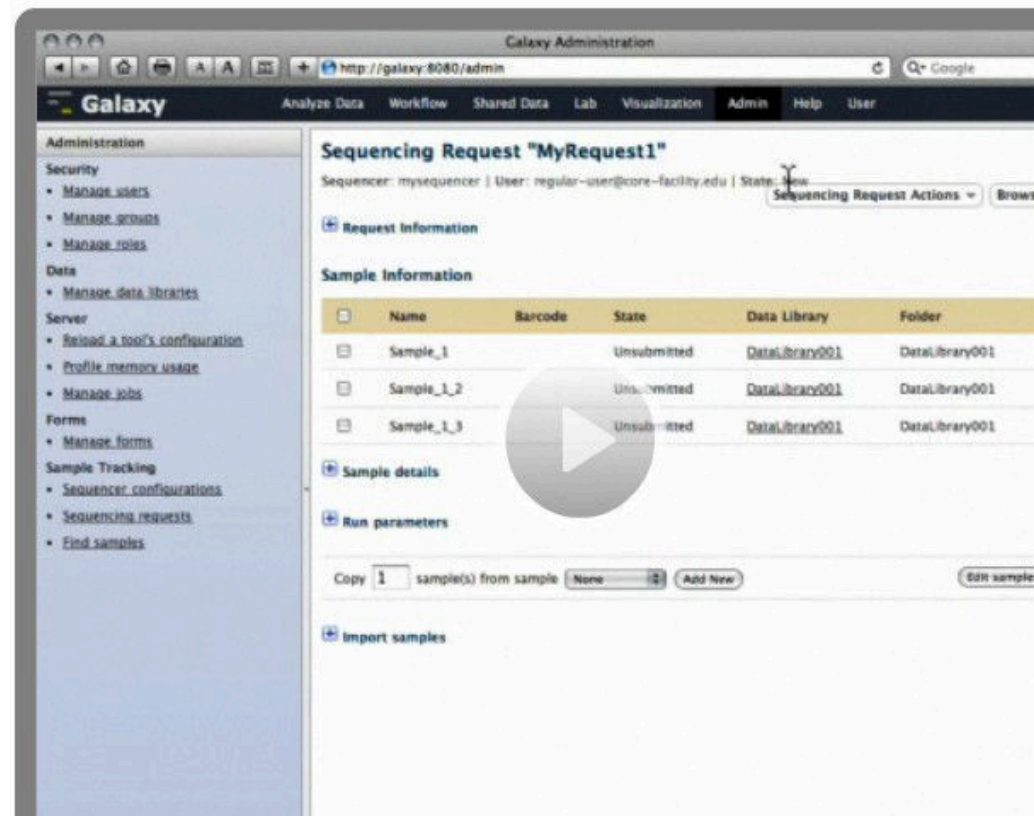


Galaxy Sample Tracking

Galaxy Sample Tracking System

Screenscasts

1. [Create forms](#)
2. [Create sequencer configuration](#)
3. [Create sequencing request as an admin](#)
4. [Admin tasks](#): assign bar code, change sample state, reject request
5. [Transfer datasets](#)
6. [Create sequencing request as a regular user](#)



Wishes

- Better layout
 - Easier navigation
- Flowcell design
 - Illumina HiSeq2000
- Track sample information
 - Library preparations
 - QC data/Bioanalyser pictures

Massachusetts General Hospital

The screenshot displays a web-based interface for next-generation sequencing (NGS) details. On the left, a sidebar menu lists navigation options: 'Samples' (with sub-items: 'Define samples and services', 'Submit samples as a project', 'View projects', 'Sequencing results'), 'Sequencing' (with sub-items: 'Queues', 'Runs', 'Plots'), and 'Plots'. The main content area is titled 'Next gen sequencing details' and contains several input fields and a table.

Run folder: A text input field containing '110107_'. Below it, the text '(Required)' is displayed.

Run notes: A large text area for notes. Below it, the text '(Required)' is displayed.

Samples: A table listing samples to be sequenced.

Samples
BCa14 Test sample 4 (Illumina 2-4)
BCa6 Testing
PhiX control

Copy: A button labeled 'Copy' is located below the samples table.

Lane 1 through Lane 7: A column of seven empty text input fields, each labeled 'Lane 1' through 'Lane 7' respectively, for assigning samples to sequencing flow cells.

Insert mode: A button labeled 'Insert mode' is located at the bottom right of the interface.

A text box at the bottom of the interface states: "A drag and drop interface is used to assign samples to sequencing flow cells."

Great, but..

- Define 1 sample over 3 pages
 - Expected experiments will have >300!
 -
- Users define the multiplexing barcodes/indexes
 - Many will not know what to do...
 - Limits flowcell design
- -

Custom

- Keep it simple
-
- Borrow flowcell design idea from Brad Chapman
-
- Make small interface modifications
-
- Send requests and flowcell information to a wiki

Interface - notifications

Create a new sequencing request

Create a new sequencing request

Select a request type configuration:

TESTING-REQUEST

Contact the lab manager if you are not sure about the request type configuration.

Name of the Experiment

Demo

(Required)

Description

Conference

(Optional)

LabGroup

TRON

Your Laboratory/Institute/Company (Required)

Billing Account Number

....

Which account should be used for billing. Please contact us for more information (Required)

Accompanying datasets

Demo

Please provide any quality data/pictures that go with this experiment by uploading them into your history and attach them here.
(Optional)

Send to:

☒ jelle.scholtalbers@tron-mainz.de (sequencing request owner)

Additional email addresses:

Enter one email address per line

Select sample states to send email notification:

☒ New ☒ Arrived ☒ Library prepared ☒ Flowcell prepared ☒ Sequencing Finished ☒ Raw data available ☒ All data available

Email notification will be sent when all the samples of this sequencing request are in the selected states.

Save Add samples

Interface - add samples

Return to request overview View request "Demo" Edit request Add samples View history

Add Samples to Sequencing

Name
Sample_1 (required)
Sample_1_2 (required)
Sample_1_3 (required)
Sample_1_4 (required)
Sample_1_5 (required)
Sample_1_6 (required)

For each sample, select the layout:

Layout1

Name	Sequencing
Sample_1	DNA: Genomic DNA Sequencing (required)
Sample_1_2	DNA: Genomic DNA Sequencing (required)
Sample_1_3	DNA: Genomic DNA Sequencing (required)
Sample_1_4	DNA: Genomic DNA Sequencing (required)
Sample_1_5	DNA: Genomic DNA Sequencing (required)
Sample_1_6	DNA: Genomic DNA Sequencing (required)

Copy 1 samples from sample Sample_1

Select the sample from which the new sample should be copied or leave selection as None to add a new "generic" sample.

Add sample Save Cancel

Click the Add sample button for each new sample and click the Save button when you have finished adding samples.

History	Workflow
Demo	TRON - SNP discovery
Demo	TRON - SNP discovery
Demo	TRON - SNP discovery
Demo	TRON - SNP discovery
Demo	TRON - SNP discovery
Demo	TRON - SNP discovery

a workflow on run datasets, select a history first and then the desired workflow.

/ul	Library Prepared	Basepairs	Concentration ng/ul	Sample QC	Sample Type	Read length
☒				Select one (optional)	normal (optional)	50bp (optional)
☒	50 (required)			Select one (optional)	normal (optional)	50bp (optional)
☒				Demo (optional)	normal (optional)	50bp (optional)
☐	Was the library prepared? Uncheck to let us create the library.			Select one (optional)	FFPE/in block (optional)	2x100bp (Paired End) (optional)
☐				Select one (optional)	FFPE/in block (optional)	2x100bp (Paired End) (optional)
☐				Select one (optional)	FFPE/in block (optional)	2x100bp (Paired End) (optional)

Interface - Admin

1 samples to put on a flowcell

1 samples using:

form value

Select a sample attribute for searching. To search for a sample with a dataset name, select the dataset name above. This will return all the samples that are associated with a dataset with that name.

not limit on sequencing finished state: ☐

organism=hg19

Wildcard search (%) can be used as placeholder for any sequence of characters or words. For example, to search for samples starting with 'mysample' use 'mysample%' as the search string.

When 'form value' search type is selected, then enter the search string in 'field label=value' format. For example, when searching for all samples whose 'Volume' field is 1.3mL, then the search string should be 'Volume=1.3mL' (without quotes). Possible fields:

There are 3 samples that matched the given criteria.

Search results are sorted by the date the requests were created.

Request	Sample	Barcode	State
Sequencing request: Demo Type: TESTING-REQUEST State: In Progress	Sample_1_3	2342345	Arrived
Sequencing request: Demo Type: TESTING-REQUEST State: In Progress	Sample_1_2	2343444	Arrived
Sequencing request: Demo Type: TESTING-REQUEST State: In Progress	Sample_1	2342342	Arrived

Indexes

Control

PhiX
Add to all

Name:

Description:

Lane 1

Indexes

PhiX (X)	
Sample_1 ()	ATCACG (X)
Sample_1_2 ()	TTAGGC (X)
Sample_1_3 ()	TGACCA (X)

Lane 2

Indexes

PhiX (X)	
Sample_1_4 ()	ACAGTG (X)
Sample_1_7 ()	TAGCTT (X)
Sample_1_5 ()	CAGATC (X)

Lane 3

Indexes

PhiX (X)	
Sample_1_6 ()	Add index
Sample_1_9 ()	

Lane 4

Indexes

PhiX (X)	
Add sample here	

Lane 5

Indexes

PhiX (X)	
Add sample here	

Lane 6

Indexes

PhiX (X)	
Add sample here	

Lane 7

Indexes

PhiX (X)	
Add sample here	

Lane 8

Indexes

--	--

25.05.2011

Name	Description	Samples	Last Updated ↑	State	User
Edit	Presentation	8	less than a minute ago	New	jelle.s
Submit		5	3 minutes ago	New	jelle.s
Finished					
Samplesheet	First	2	2 days ago	New	jelle.s
Completed					

	A	B	C	D	E	F	G	H	I
1	FCID	Lane	SampleID	SampleRe	Index	Description	Control	Recipe	Operator
2	FXID	1	2342342	hg19	ATCACG	Conference	PhiX	NA	jelle.scholt
3	FXID	1	2343444	hg19	CGATGT	Conference	PhiX	NA	jelle.scholt
4	FXID	2	2342345	hg19	TGACCA	Conference	PhiX	NA	jelle.scholt
5	FXID	2	2423422	mmtv	CAGATC	Conference	PhiX	NA	jelle.scholt
6	FXID	2	2423526	mmtv	ACTTGA	Conference	PhiX	NA	jelle.scholt
7	FXID	3	Sample_1	mm9	None	Test	NA	NA	jelle.scholt
8	FXID	3	Sample 1	mm9	None	Test	NA	NA	jelle.scholt

Wiki – Confluence

Dashboard



Dashboard

Welcome to the TRON Wiki

Get started by browsing through the pages or create a new one.
To create a new project please contact the administrators.

[Add Space](#)

[Feed Builder](#)

[People Directory](#)

Search

Recently Updated



Bot by Jelle

[Re: = GALAXY-REQUEST = 86 = Test permission settings](#) commented 5 minutes ago
State = In Progress

[= GALAXY-REQUEST = 86 = Test permission settings](#) updated 5 minutes ago (view change)

[Re: = GALAXY-REQUEST = 93 = Demo](#) commented 5 minutes ago
State = In Progress

[= GALAXY-REQUEST = 93 = Demo](#) updated 5 minutes ago (view change)

[Re: = GALAXY-FLOWCELL = 94 = FXID](#) commented 5 minutes ago

xy Sample Tracking Notification

=====

r: jelle.scholtalbers@tron-mainz.de

Sequencing request: Demo

Sequencing request id: 93

Sequencing description: Galaxy Sequencing Request Basic info after completion

Start time: 2011-05-10 13:15:36.583012

End time: 2011-05-10 13:15:36.583012

Sequencer configuration: TESTING-REQUEST

Sequencing request state: In Progress

Number of samples: 6

Samples in state:

Sample Name	Data Library	Folder	History	Barcode	Workflow	SequencingApplication	SourceOrganism	Sample State
Sample_1	Sequencing Projects	Public	Demo	2342342	TRON - SNP discovery	DNA: Genomic DNA Sequencing	hg19	Not started
Sample_1_2	Sequencing Projects	Public	Demo	2343444	TRON - SNP discovery	DNA: Genomic DNA Sequencing	hg19	Not started
Sample_1_3	Sequencing Projects	Public	Demo	2342345	TRON - SNP discovery	DNA: Genomic DNA Sequencing	hg19	Not started
Sample_1_4	Sequencing Projects	Public	Demo	2423422	TRON - SNP discovery	DNA: Genomic DNA Sequencing	mmtv	Not started
Sample_1_5	Sequencing Projects	Public	Demo	2423526	TRON - SNP discovery	DNA: Genomic DNA Sequencing	mmtv	Not started
Sample_1_6	Sequencing Projects	Public	Demo	5675673	TRON - SNP discovery	DNA: Genomic DNA Sequencing	mmtv	Not started

GALAXY-FLOWCELL = 94 = FXID

[Edit](#) [Add ▾](#)

led by [Bot by Jelle](#), last edited by [Bot by Jelle](#) on May 10, 2011 ([view change](#))

Galaxy Flowcell Notification

=====

er: jelle.scholtalbers@tron-mainz.de

me: FXID

escription: Presentation

bmit time: None

reate time: None

ate: Complete

ane(control)	Sample (ID)	Multiplex Barcode	State	Concentration (Advised)	Request (ID)
--------------	-------------	----------------------	-------	----------------------------	--------------

State = NEW

[Edit](#) | [Remove](#) | [Reply](#)

[Bot by Jelle](#) says:

9 minutes ago

State = Complete

[Remove](#) | [Reply](#)

[Comment](#)

Technical

- Flowcell ~ Request
 - `request_type = ''`
 -
- Flowcell Form
 - Describes lanes, controls, indexes
 - Data is stored in the `form_values`
-
- 1 column `'illumina_index'` added to Sample table
-
- RabbitMQ – AMQP
 - Used for sending Wiki updates
 - Both Confluence and (Semantic) MediaWiki supported

Security

- [Manage users](#)
- [Manage groups](#)
- [Manage roles](#)

Data

- [Manage data libraries](#)

Server

- [Reload a tool's configuration](#)
- [Profile memory usage](#)
- [Manage jobs](#)

Form Definitions

- [Manage form definitions](#)

Sample Tracking

- [Manage sequencers and external services](#)
- [Manage request types](#)
- [Sequencing requests](#)
- [Flowcells](#)
- [Create flowcell](#)
- [Find samples](#)

Form definition "Flowcell" (Flowcell Form)**Lane 1**TextField: *Drag and Drop your sample. - (optional)***Lane 2**TextField: *Drag and Drop your sample. - (optional)***Lane 3**TextField: *Drag and Drop your sample. - (optional)***Lane 4**TextField: *Drag and Drop your sample. - (optional)***Lane 5**TextField: *Drag and Drop your sample. - (optional)***Lane 6**TextField: *Drag and Drop your sample. - (optional)***Lane 7**TextField: *Drag and Drop your sample. - (optional)***Lane 8**TextField: *Drag and Drop your sample. - (optional)***Illumina Indexes**SelectField: *Drag them to your sample on a lane - (optional)***Options:**

ATCACG CGATGT TTAGGC TGACCA ACAGTG GCCAAT CAGATC ACTTGA
 GATCAG TAGCTT GGCTAC CTTGTA

ControlsSelectField: *- (optional)***Options:**

PhiX

Outlook

- Setting up the actual data transfer and analyses
 - Waiting for new Illumina preprocessing software
- Enable semantics on Confluence
 - zAgile Wikidsmart
 -
- Send data to Ingenuity

Send to Ingenuity

Galaxy / TRON

Analyze Data Workflow Shared Data Lab Visualization Help Use

Tools Options

- Get Data
- Send Data
- ENCODE Tools
- Lift-Over
- Text Manipulation
- Filter and Sort
- Join, Subtract and Group
- Convert Formats
- Extract Features
- Fetch Sequences
- Fetch Alignments
- Get Genomic Scores
- Operate on Genomic Intervals
- Statistics
- Wavelet Analysis
- Graph/Display Data
- Regional Variation
- Multiple regression
- Multivariate Analysis
- Evolution
- Metagenomic analyses
- FASTA manipulation
- Send to Ingenuity
- NGS: QC and manipulation
- NGS: Mapping
- NGS: Indel Analysis
- NGS: RNA Analysis
- NGS: SAM Tools
- NGS: Peak Calling

Welcome to Galaxy

Hosted by the Institute for Translational Oncology

 **Come work with us**

Visit our job offers:

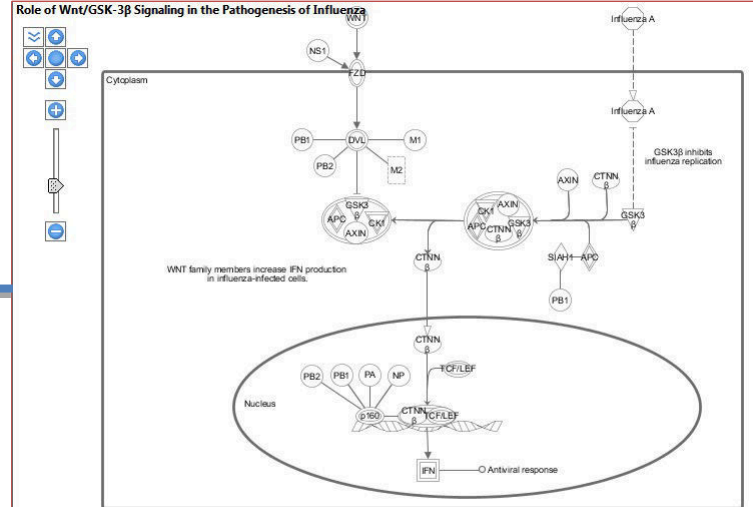
- 1 Manager Intellectual Property (m/w)
- 2 technische/n Assistentin/en – serologische Assays
- 3 Technische Assistenz (w/m)
- 4 Genomics scientist or PhD student
- 5 Bioinformatician
- 6 Computational biologist / Bioinformatician
- 7 NGS bioinformatician



←

INGENUITY
SYSTEMS

Role of Wnt/GSK-3 β Signaling in the Pathogenesis of Influenza



Acknowledgements



Massachusetts General Hospital

- Brad Chapman



- | | |
|----------------------|---------------------|
| • Ugur Sahin | • John Castle |
| • Jos de Graaf | • Ludmila Schemarow |
| • Valesca Boisguérin | • Marius Byl |
| • Martin Löwer | • Thomas Bukur |
| • Thomas Rösler | • Michael Rauer |