

Galaxy LIMS for Next Generation Sequencing

Jelle Scholtalbers (jelle.scholtalbers@tron-mainz.de)



October 3, 2012 Bern, Switzerland **TRON** is a non-profit, transdisciplinary research institute for the treatment of cancer and other diseases of the immune system.

TRON researches the development of novel platforms for individualized therapies and biomarkers and translates basic research into drug applications.

TRON partners with academic institutions, biotech companies and the pharmaceutical industry with a focus on the quick translation of research into diagnostic products, pharmaceuticals and therapies.

TRON is located in Mainz, Germany, at the campus of the University Medical Center of the Johannes Gutenberg University.





Patient benefit using NGS





- Handle small lab with one HiSeq 2000
 - Many samples per flow cell
- Ease of use
- Tracked and accessible sample information
 - Library preparations
 - QC data / Bioanalyzer pictures
- Flow cell design



- Automatic data preprocessing
 - CASAVA
 - Data organization
- Automatic workflow execution
- Request owner can analyze own data



- Broad analysis platform
- Create and 'publish' workflows
- Basic Sample-Tracking available
 - Request & Sample tracking
 - User Roles
 - Sequencer Connection*
- Python code & I'm biased



Galaxy – Basic Sample Tracking

Galaxy

Screencasts

1. Create forms

5. Transfer datasets

()

Accessible Page | STS Screencasts

Galaxy Sample Tracking System

2. Create sequencer configuration

000 **Galaxy** Administration ▲ ► ▲ ▲ ▲ ▲ ▲ ▲ ▲ ★ ● http://galaxy 8080/admin C Q+ Grogie Galaxy Analyze Data Workflow Shared Data Lab Visualization Admin Help User Administration Sequencing Request "MyRequest1" Security Manage users · Manage groups Request Information Manage roles. 3. Create sequencing request as an admin Data Sample Information · Manage data librartes 4. Admin tasks: assign bar code, change sample state, reject request Name Barcode State Data Library Folder Server · Reload a tool's configuration 8 Sample_1 Unsubmitted DataLibrary001 DataLibrary001 * · Profile memory usage 0 Sample_1_2 DataLibrary001 н Una, Omitted DataLibrary001 Manage jobs 6. Create sequencing request as a regular user Forms 0 Sample_1_3 DataLibrary001 DataLibrary001 н Unsult sted · Manage forms Sample Tracking Sample details Sequencer configurations · Sequencing requests Hun parameters · Find samples Copy 1 sample(s) from sample None 2 (Add New) (tdit samples) E Import samples 0 01:42/01:55

Workflow

Shared Data

Visualization

Help

User

Analyze Data

Galaxy - NGLIMS

💳 Galaxy	Analyze Data	Workflow	Data Libraries	Lab	Visualization	Admin	Help	Use
Samples • Define samples and services • Submit samples as a project • View projects Sequencing • Queues • Runs	Samples TJa3 Sample Copy	4		Lane 1 TJa1 S Lane 2 PhiX c Lane 3 BCa1 1 Lane 4 Lane 5 Lane 6 Lane 7 Lane 8	Sample 2 ontrol Fest sample 1			

http://wiki.g2.bx.psu.edu/Admin/Sample%20Tracking/Next%20Gen



- Small menu changes for better navigation
- Sample submission form
- Borrow flow cell design idea from Brad Chapman
- Update wiki with request and flow cell information
- Scripts and API to manage data



NGS Process



10

Implementation



Customer

💳 Galaxy / TF	RON Analyze Data	Workflow	Shared Data -	Lab - Visualization -	Help - User -	Using 0 bytes
Sequencing Re Advanced Search	quests ्			Sequencing Requests Find Samples Help		Create new request
Name	Barcode	Descriptio	<u>n</u> Samples	Туре	Last Updated ↑	State
■ MyExp301 *	430042430		Z	NGS Request	Sep 04, 2012	In Progress
MyExp300 -	311233211		<u>6</u>	NGS Request	Sep 04, 2012	In Progress
🗐 <u>89888x</u> -	134342123		2	NGS Request	Aug 15, 2012	In Progress
For 0 selected item	s: Delete Undel	lete				



Customer

Calaxy / TRON Analyze Data Workflow Shared Data Lab Visualization Help User Return to request overview View request "MyExp304" Edit request Add samples View samples Associate Data View history Add samples to sequencing request "MyExp304" Edit request Add samples View samples Associate Data View history Add samples to sequencing request "MyExp304" State Data Library Folder Sample_1 Sequencing Projects MyExp304

(required)

For each sample, select the data library and folder in which you would like the run datasets deposited. To automatically run a workflow on run datastets, select a history first ar

Sample information

Name	Sequencing Type	Read Length	Organism	Sample Origin FFPE		Extracted
Sample 1	RNA: mRNA Sequencing 🔹	50bp 👻	Human Mar. 2006 (NCBI36/hg18) (hg18) 🛛 🔻	Tissue 🔻	No 🔻	RNA
Sample_1	(required)	(required)	(optional)	(optional)	(required)	(required)

Library information

Multiplexing?	Describe
	1: Sample_1.1, Sample_2.1, Sample_3.1 2: Sample_1.2, Sample_2.2, Sample_3.2

Name	Library Prepared?	Illumina Index	
Sample 1.1		AT	
Sample_1.1		1: ATCACG	
		2: CGATGT	
Sample_2.1		6: GCCAAT	
		7: CAGATC	
Sample_3.1		9: GATCAG	
		4 ATOLOG	

Ξ	🗧 Galaxy / TRON	Analyze Data	Workflow Shared	Data• Lab•	Visualization •	Help∓ User		Using 0 bytes
(Return to request overview	View request "MyE:	xp304" Edit reques	t Add samples	Edit samples	Associate Dat	a View history	Submit request
	The selected dataset is be away from Galaxy and using the selected dataset is be away from Galaxy and using the selected dataset is be away from Galaxy and using the selected dataset is be	ing uploaded to the ng the browser's 'sto	library 'Sequencing F p' or 'reload' buttons	Projects' into fold	er 'MyExp304'.	Please wait a co	uple of minutes b	efore navigating

Vpload files

File:	Message:		Genome:	
	Customer QC related to requ	est .::	Click to Search or Sel	ect 🔻
Upload to library	This information will be displa in the data library browser	yed in the "Message" column for this	5 dataset	
Associated data:	sets			
Name	Message	Description Uploaded By	File type File type	File Size +
Gels.pdf -	Customer QC related to request	x@tron-mainz.de	auto 2012-10-01	319.0 Kb

() TIP: Several compression options are available for downloading multiple library datasets simultaneously:

· gzip: Recommended for fast network connections

· bzip2: Recommended for slower network connections (smaller size but takes longer to compress)

· zip: Not recommended but is provided as an option for those who cannot open the above formats

Galaxy LIMS - Customer



Lab Admin – Manage requests

💳 Galaxy / TRON	Analyze Data Wor	kflow Shared	Data - Lab -	Visualizati	on - Lab Admir	r▼ Help≖ User	1 5	Using 0 bytes
Lab Administration	Sequencing Ro	equests			Browse F	lowcells	Create new req	uest Create flowcell
Data <u>Manage data libraries</u> Form Definitions	Advanced Search	Q			Sequenc Manage	ing Requests data libraries		
Manage form definitions	<u>Name</u>	<u>Barcode</u>	Description	Samples	Туре	Last Updated †	State	User
Sample Tracking Manage sequencers and 	MyFxn304	304132002	Bern	<u>6</u>	NGS Request	28 minutes ago	In Progress	x@tron-mainz.de
external services	Add Samples Edit Samples	30042430		Z	NGS Request	Sep 04, 2012	In Progress	x@tron-mainz.de
<u>Manage request types</u> <u>Sequencing requests</u>	View Data Myl Update state OC details 0	11233211		<u>6</u>	NGS Request	Sep 04, 2012	In Progress	x@tron-mainz.de
 <u>MyExp304</u> <u>MyExp301</u> 	Reject Delete	44042241		2	NGS Request	Aug 16, 2012	In Progress	h@tron-mainz.de
 MyExp300 x234 	View History Get submission f	orm 21120230		<u>4</u>	NGS Request	Aug 16, 2012	Rejected	h@tron-mainz.de
• <u>X34</u>	🔲 <u>89888x</u> -	134342123		2	NGS Request	Aug 15, 2012	In Progress	x@tron-mainz.de
• <u>89888x</u> 	For 0 selected ite	ms: Delete	Undelete					

- Flow cells
 - Create flowcell
- Find samples

Lab Admin – Create flow cell

 \bigotimes

– Galaxy / TRON		Analyze Data Workflow	v Shared Data - Lab	Visualization La	b Admin≖ Help≖ I
Lab Administration	Find samples to put on a Fl	lowCell			
Data Manage data libraries Form Definitions Manage form definitions Sample Tracking Manage sequencers and external services	Filter on: read length ▼ -> All ▼ Do Select one Fi 50bp 100bp Th All single end	shed state: 🔄			
<u>Manage request types</u>	The 2x50bp (Paired End) 2x100bp (Paired End)	y the date the requests where o	reated.		
Sequencing requests	All paired end			Read Length	♦ Sample ♦
 MyExp304 MyExp301 	Sequencing request: MyEx	p304 Type: NGS Request Sta	te: In Progress	50bp	Sample_3.2
► <u>MyExp300</u>	Sequencing request: MyEx	p304 Type: NGS Request Sta	te: In Progress	50bp	Sample_2.2
 x234 x34 	Sequencing request: MyEx	50bp	Sample_1.2		
• <u>89888x</u>	Sequencing request: MyEx	p304 Type: NGS Request Sta	te: In Progress	50bp	Sample_3.1
···· ▼ <u>Flow cells</u>	Sequencing request: MyEx	p304 Type: NGS Request Sta	te: In Progress	50bp	Sample_2.1
<u>Create flowcell</u>	Sequencing request: MyEx	p304 Type: NGS Request Sta	te: In Progress	50bp	Sample_1.1
 232y COC00CXX 	Sequencing request: 8988	8x Type: NGS Request State:	In Progress	50bp	Sample_2
• <u>CX1008RCV</u>	Sequencing request: 8988	8x Type: NGS Request State:	In Progress	50bp	Sample_1
<u>Find samples</u>	Sequencing request: MyEx	p301 Type: NGS Request Sta	te: In Progress	50bp	S4
	Sequencing request: MyEx	p301 Type: NGS Request Sta	te: In Progress	50bp	53
	Sequencing request: MyEx	p301 Type: NGS Request Sta	te: In Progress	50bp	S2
	Sequencing request: MyEx	p301 Type: NGS Request Sta	te: In Progress	50bp	S1
	Sequencing request: MyEx	p300 Type: NGS Request Sta	te: In Progress	50bp	X1.3
	For selected samples	reate flowcell Show suggest	ed flowcell		

Lab Admin – Create flow cell

 \bigotimes

Galaxy / TRON	Analyze Data Workflow	Shared Data 👻 La	b • Visualization • Lab Admin • Help • User •		Using 0 byte
Lab Administration			Flowcell		
Data	TRON Flowcell -		Name: C1C00CXX Description: 50bp		
Manage data libraries	Samples	Control	Lane 1	Indexes	Dilutions
Form Definitions	X1.3 (X1.3)	PhiX	Add control here		Actual (Advised)
Manage form definitions	X1.3 (X1.3)	Add to all	○ S1.1 (Sample_1.1)	ATCACG	(NA)
Sample Tracking	X1.3 (X1.3)		S2.1 (Sample_2.1)	CGATGT	(NA)
 Manage sequencers and external services 			○ S3.1 (Sample_3.1)	TTAGGC	(NA)
<u>external services</u>				ATCACG	(NA)
<u>Manage request types</u> Sequencing requests			Lane 2	Indexes	Dilutions
MyExp304			Add control here		Actual (Advised)
MyExp304			52.2 (Sample 2.2)	CGATGT	(NA)
• <u>MyExp300</u>			53.2 (Sample 3.2)	TTAGGC	(NA)
► x234			Lane 3	Indexes	Dilutions
► <u>X34</u>			Add control here		Actual (Advised)
▶ <u>89888x</u>			MS1 (S1)	CTTGTA	(5013.38)
			MS2 (S2)	GGCTAC	(3998.0)
			00 MS3 (S3)	TAGCTT	(5105.69)
<u>Flow cells</u> Croote flowcoll			M54 (S4)	GATCAG	(5850.31)
- <u>Create nowcen</u>			Lane 4	Indexes	Dilutions
▶ <u>232y</u>			Add control here		Actual (Advised)
► <u>COC00CXX</u>			Add sample here		
• <u>CX1008RCV</u>			Lane 5	Indexes	Dilutions
			Add control here		Actual (Advised)
<u>Find samples</u>			Add sample here		

Lab Admin – Create flow cell

Lab Administration	Defined flowe	cells	
Data Manage data libraries Form Definitions Manage form definitions 	Advanced Search Flowcell created		
Sample Tracking Manage sequencers and external services Manage request types Sequencing requests MyExp304	Name Edit Submit Finished	sample_sheet_flowcell_C1C00CXX.csv [Schreibgeschützt] - Microsoft ExcelDateiStartEinfügenSeitenlayoutFormeinDatenÜberprüfenAnsichtCalibri* 11AA \equiv <th< th=""><th>≓⇔ Einfüger I∛ Löschen Format Zellen</th></th<>	≓⇔ Einfüger I∛ Löschen Format Zellen
 MyExp301 MyExp300 x234 X34 89888x 	Add data Sampleshee Samples F Revert Delete	A1 fx FCID A1 fx FCID A B C D E F G H I FCID Lane SampleID SampleRef Index Description Control Recipe Ope 2 C1C00CXX 1 Sample_1.1 hg18 ATCACG Bern;UsedKi N Iims 3 C1C00CXX 1 Sample_2.1 hg18 CGATGT Bern;UsedKi N Iims 4 C1C00CXX 1 Sample_1.2 hg18 ATCACG Bern;UsedKi N Iims 5 C1C00CXX 1 Sample_1.2 hg18 ATCACG Bern;UsedKi N Iims 6 C1C00CXX 2 Sample_2.2 hg18 CGATGT Bern;UsedKi N Iims 7 C1C00CXX 2 Sample_3.2 hg18 TTAGGC Bern;UsedKi N Iims 7 C1C00CXX 2 Sample_3.2 hg18 TTAGGC Bern;UsedKi N Iims	I S S S S S S S S S S S S S
		8C1C00CXX3 S1mm9CTTGTA;UsedKits:NImms9C1C00CXX3 S2mm9GGCTAC;UsedKits:NImms10C1C00CXX3 S3mm9TAGCTT;UsedKits:NImms11C1C00CXX3 S4mm9GATCAG;UsedKits:NImms12C1C00CXX4 X1.3GCCAAT;UsedKits:NImms13C1C00CXX5 X1.3GCCAAT;UsedKits:NImms14C1C00CXX6 X1.3GCCAAT;UsedKits:NImms15IIIIIII	S S S S S S S S S
<u>Find samples</u>		16 17 18 19 20 I(← → →) sample_sheet_flowcell_C1C00CXX / 22 / 1	

Customer – View Data

🗧 Galaxy / TRO	Nze Data 🛛 Wo	orkflow Shared	Data≁ Lab≁	Visualization - I	Help Vser -		
Sequencing Requ	ests			Crea	ate new request		
Advanced Search							
Name Barco	ode <u>Descri</u>	<u>ption</u> Samples	Туре	Last Updated	State		
View Request	<u>02</u> Bern	<u>6</u>	NGS Request	~ 1 hour ago	In Progress		
Get submission form	n <u>30</u>	Z	NGS Request	Sep 04, 2012	In Progress		
- Galaxy / TRON	Analyze Data	Workflow Share	ed Data 👻 🛛 Lab	 Visualization - 	Help - User -	U	Jsing 0 bytes
Data Library "Sequ	encing Pro	ojects"		Return	Go to request	Add datasets	Add folder
Name	Message			Uploaded by	Data type	Date uploaded	File size
🔲 🏾 🔁 x@tron-mainz.de 👻	Your sequer	icing requests will	be placed here				
🔲 🕨 📴 MyExp300 👻	Containing a	all data related to	this request				
🔲 🕨 📴 MyExp301 👻	Containing a	all data related to	this request				
🔲 🏾 📄 MyExp304 💌	Containing a	all data related to	this request				
Gels.pdf 👻				x@tron-mainz	.de pdf	2012-10-01	319.0 Kb
For selected datasets:	Import to currer	nt history 💌 🖸	io				
8						© 2012 TRC)N gGmbH

Implementation



API – Flow cell processing



Customer – View Data

💳 Galaxy / TRON	ze Data Workflow	Shared	Data → Lab →	Visualization -	Help≁ User≁		
Sequencing Reque	sts			Crea	ate new request		
Advanced Search							
<u>Name</u> <u>Barcod</u>	e <u>Description</u>	Samples	Туре	Last Updated †	State		
View Request	- <u>02</u> Bern	<u>6</u>	NGS Request	~ 1 hour ago	In Progress		
🔲 👔 Get submission form	30	<u>7</u>	NGS Request	Sep 04, 2012	In Progress		
- Galaxy / TRON	alyze Data Workfl	ow Share	ed Data 👻 Lab	 Visualization - 	Help - User -		Jsina 0 bytes
Data Library "Seque	ncing Project	ts″		Return	Go to request	Add datasets	Add folder
Data Library "Seque	ncing Project	ts″		Return Uploaded by	Go to request Data type	Add datasets Date uploaded	Add folder
Data Library "Seque	ncing Project Message	ts″		Return Uploaded by	Go to request Data type	Add datasets Date uploaded	Add folder
Data Library "Seque Name x@tron-mainz.de •	Message Your sequencing re	t s'' equests will	l be placed here	Return Uploaded by	Go to request Data type	Add datasets Date uploaded	Add folder File size
Data Library "Seque Name x@tron-mainz.de MyExp300	Message Your sequencing re Containing all data	t s'' equests will a related to	l be placed here this request	Return Uploaded by	Go to request Data type	Add datasets Date uploaded	Add folder File size
Data Library "Seque Name X@tron-mainz.de MyExp300 MyExp301 MyExp301	Message Your sequencing re Containing all data	t s'' equests will a related to a related to	l be placed here this request this request	Return Uploaded by	Go to request Data type	Add datasets Date uploaded	Add folder File size
Data Library "Seque Name Name MyExp300 MyExp301 MyExp304	Message Your sequencing re Containing all data Containing all data	equests will a related to a related to a related to	be placed here this request this request this request	Return Uploaded by	Go to request Data type	Add datasets Date uploaded	Add folder File size
Data Library "Seque Name Name MyExp300 MyExp301 MyExp301 Gels.pdf	Message Your sequencing re Containing all data Containing all data	equests will related to related to related to	l be placed here this request this request this request	Return Uploaded by x@tron-mainz	Go to request Data type	Add datasets Date uploaded 2012-10-01	Add folder File size

Wiki - Confluence

Bro

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

)ashboard

Feed Builder

People Directory

Spaces:	Favourite	lobal	
🚯 <u>Bioma</u>	rkers - expressio	1	🔞 📑 🔅
🌍 <u>Bioma</u>	rkers - nucleotide	<u>s</u>	🔞 📑 🖈

Recently Updated



Wiki - Confluence

Added by Bot by Jelle, last edited by Bot by Jelle on May 10, 2011 (view change)

Galaxy Sample Tracking Notification

User: jelle.scholtalbers@tron-mainz.de

Sequencing request: Demo Sequencing request id: 93 Sequencing description: Galaxy Sequencing Request Basic info after completion Submit time: 2011-05-10 13:15:36.583012 Create time: 2011-05-10 13:15:36.583012

Sequencer configuration: TESTING-REQUEST Sequencing request state: In Progress

Comments (2) Hide Comments | Collapse All | Add Comment

	Bot by Jelle says:	10 minutes ago
0	State = New	
	Edit Remove Reply	co
	Bot by Jelle says:	9 minutes ago
0	State = Complete	
	Edit Remove Reply	co

Add Comment

Powered by a free Atlassian Confluence Community License granted to TRON - Translationale Onkologie. Evaluate Confluence today

Powered by Atlassian Confluence 3.4.6, the Enterprise Wiki | Report a bug | Atlassian News

- Javascript for user input checks and flow cell design
- Lims_admin_users and separate interface
- Flowcell Form
 - Describes lanes, controls, indexes
 - Data is stored in the form_values
- RabbitMQ AMQP
 - Used for sending XML's with form data
 - Parsers for Confluence and (Semantic) MediaWiki
- Maybe of interest to all: "welcome.mako" and "<(zcat x.fastq.gz)"



- Flowcell ~ Request
 - request_type = "
- Added: Column 'illumina_index' to Sample table
- Added: Barcode table (pre-filled...)
- Added: 'Barcode', 'Folder_id', 'Library_id' columns to Request table
- Added: 'Flowcell_form_id' column to Request_type



- Paper submitted
- Clean-up code and remove hardcoded 'TRON'
- (Loosely) Integrate data with other systems e.g. CaTissue, Semantic MediaWiki (holding gene, cell line, .. information)
- More documentation
- Include YOUR comments ;)



Acknowledgements

TRON

Director Ugur Sahin

NGS Lab

Jos de Graaf Valesca Boisguérin Christian Albrecht Daniel Kunz Katja Manninen Christin Borgwardt Bioinformatics John Castle Patrick Sorn Jasmin Rößler Christian Bender Arbel Tadmor Rene Staritzbichler Martin Loewer Thomas Rössler Thorsten Litzenberger Marius Byl Sebastian Boegel Ludmila Schemarow

