Jenkins as a Scientific Data and Image Processing Platform

Ioannis K. Moutsatsos, Ph.D., M.SE.
Novartis Institutes for Biomedical Research
www.novartis.com

June 18, 2014

#jenkinsconf

Life Sciences are Computational Sciences

- Modern life sciences (biomedical research, systems biology) are heavily dependent on
  - Data Management
  - Computational Analysis
  - Computational Modeling
- Modern laboratory technologies and instrumentation generate data that are
  - Big
  - Heterogeneous
  - Complex

Computational Challenges & Opportunities

Scientists

- Face daily challenges by continuing increases in computational complexity
- Focused on the biology and not the compute problem
- Have varying and rapidly changing requirements

Life Sciences Research

- Benefits from computational systems that are
  - Easy to use
  - Fast to implement
  - Flexible
  - Support
    - Collaboration
    - Transparency
    - Automation
    - Reproducible Research
    - Open standards

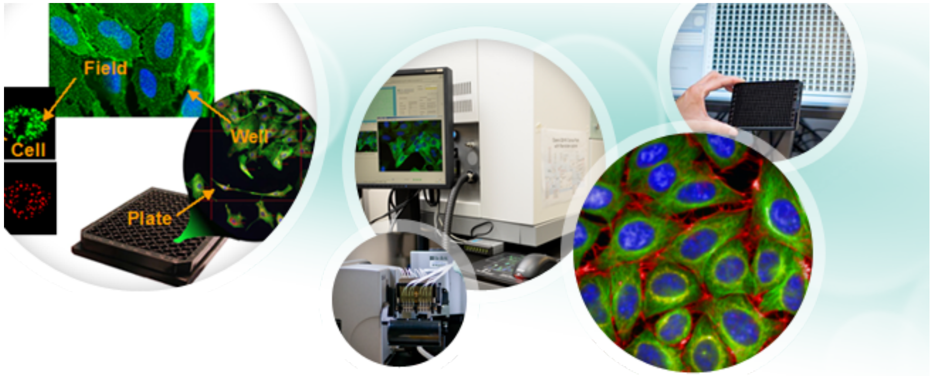
Talk Outline

- A life sciences computational challenge
  - High Content Image Analysis
    - What is it?
- Jenkins-CI as a scientific data/image processing platform
  - Functionality with standard plugins
  - How Jenkins-CI provided a HP image analysis platform for lab scientists
- Jenkins as a data analytics platform
  - Domain specific analysis and visualization plugins
  - The Jenkins pros and cons
    - What are we missing?
- Where do we want to take Jenkins?

High Content Screening: HCS

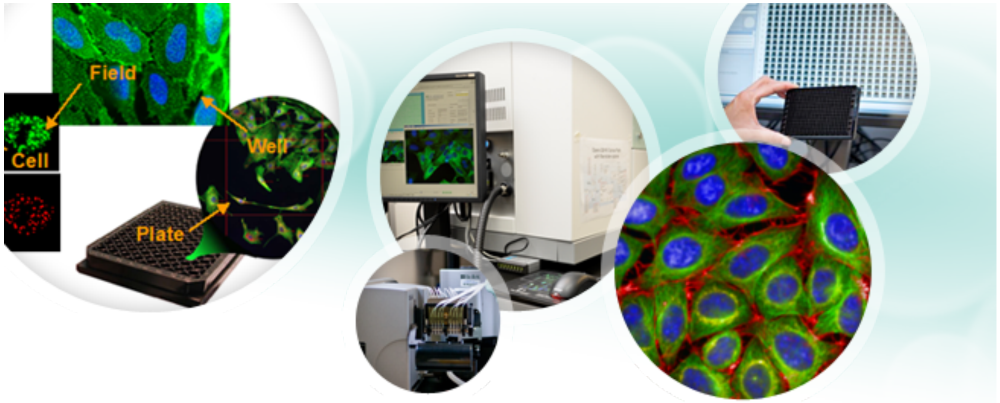
High throughput automated fluorescent microscopy for drug discovery

- Wet Lab Workflow
  - Cells grown on high density arrays
  - Cells treated with large number of chemical or biological factors
  - Cells stained with fluorescent antibodies
- Data Acquisition
  - Stained cells are imaged in high throughput mode using a computerized microscope
- Computational Workflow
  - Cell images processed to extract phenotypic measurements
  - Measurements analyzed to understand factor effects

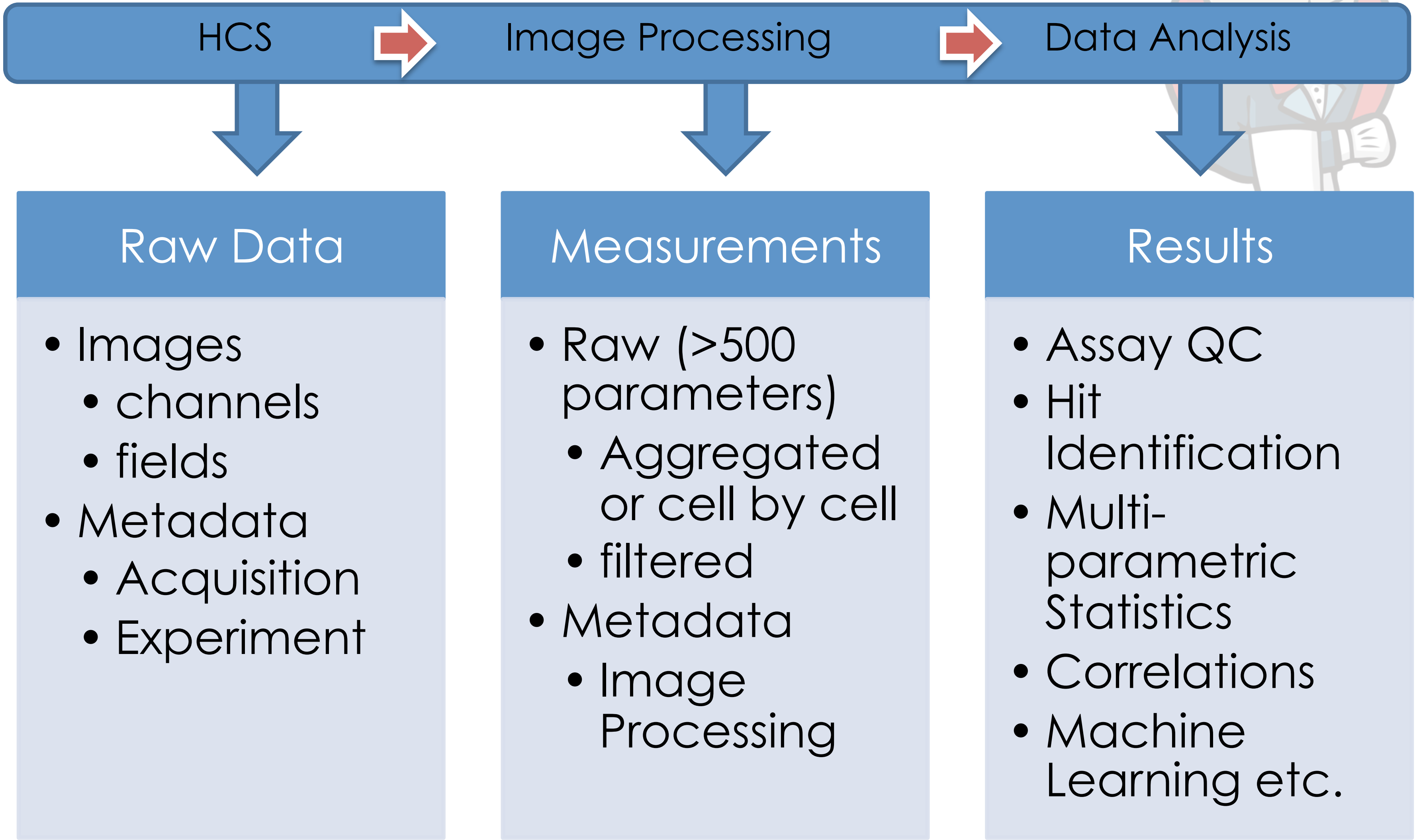


High Content Screening

- Novartis
  - High Throughput Biology (my group)
    - Data from 2010-2013
- Captured
  - 83 Terabytes of high content image data
    - 17.5 million wells
    - 27 million images
    - ~540 days of imaging time
    - ~1.5 years of computing time



HCS: Workflow and Data Stream



HCS → Image Processing → Data Analysis

Raw Data

- Images
  - channels
  - fields
- Metadata
  - Acquisition
  - Experiment

Measurements

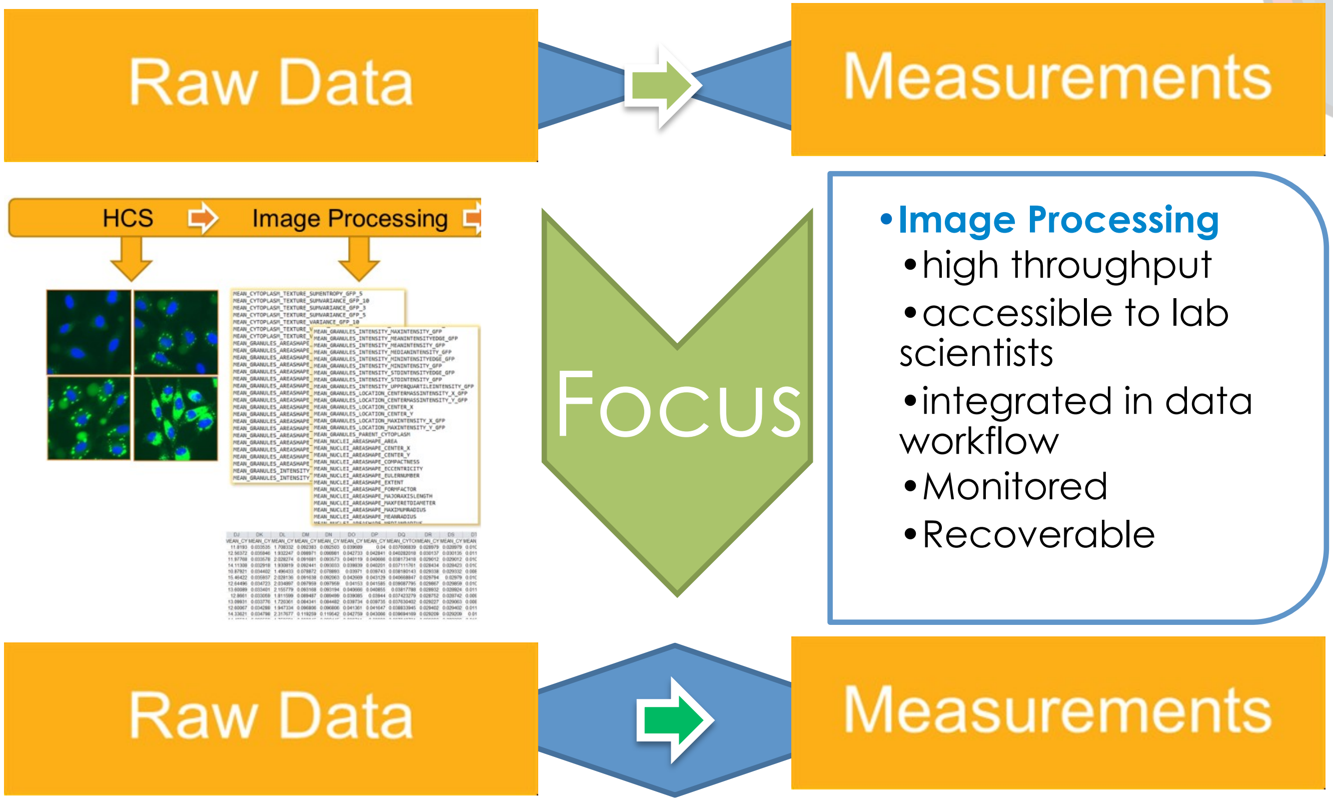
- Raw (>500 parameters)
  - Aggregated or cell by cell
  - filtered
- Metadata
  - Image Processing

Results

- Assay QC
- Hit Identification
- Multi-parametric Statistics
- Correlations
- Machine Learning etc.

HCS-High Performance Image Analysis

Intial Focus: Remove Image Processing Bottleneck



- **Image Processing**
  - high throughput
  - accessible to lab scientists
  - integrated in data workflow
  - Monitored
  - Recoverable

HCS: Image Measurements and Analytics

Easily Accessible, High Performance Image Analytics

- Vision
  - Image and data analysis using high performance (HP) image processing tools
    - Accessible, scalable, affordable, flexible and well-supported
- Strategy
  - Evaluate and adopt open-source, community supported tools
    - CellProfiler, ImageJ, Jenkins-CI
  - Utilize NIBR-IT systems and resources
    - Linux Compute Engine (cluster) / Network Attached Storage
    - Development expertise (UI, data management and web-services)
  - Increase usability of NIBR-IT systems and resources
  - Engage and provide timely and practical functionality to both expert and casual imaging platform users
- Tactics
  - Develop functional prototypes (Jenkins-CellProfiler, Test Mosaic, R-Analytics)
  - Collaborate to develop new image/data analysis systems
  - Explore imaging tools and data space. Define HP image processing requirements.
  - Provide training, support and engage in community building

CellProfiler
cell image analysis software

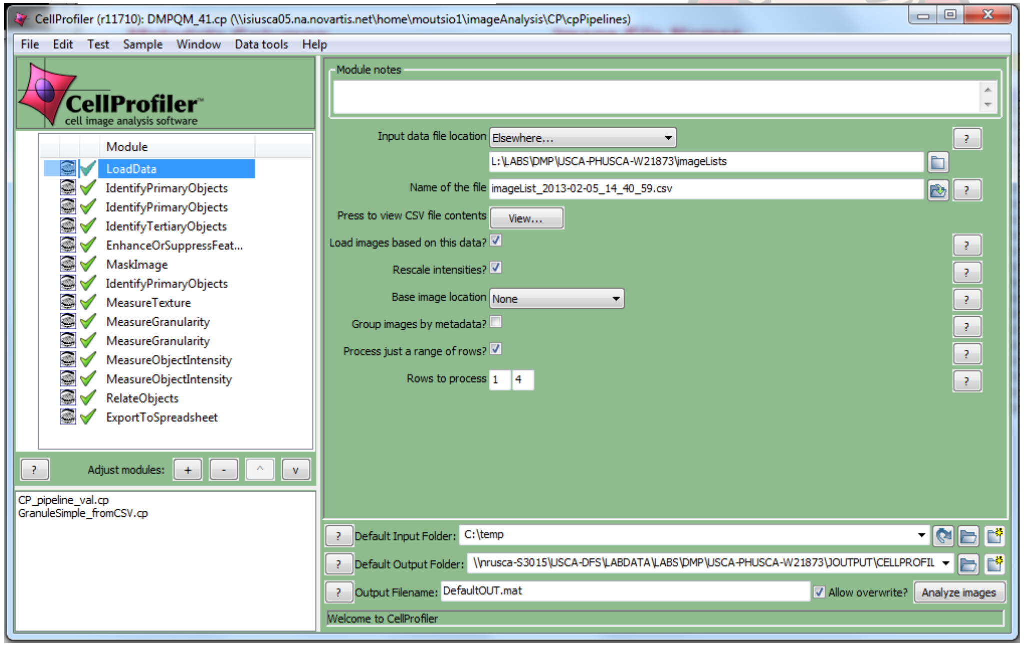
Jenkins

Linux

CellProfiler

- Open Source Image Processing
- Platform independent
- Desktop client for defining an arbitrarily complex image processing pipeline
- Pipeline can be used by the command line CellProfiler executable
    - Suitable for high throughput analysis
    - Suitable for deployment on a Linux grid engine
    - Can process large image sets (300K + images)

CellProfiler
cell image analysis software



CellProfiler (r11710): DMPQM_41.cp (\\isiusca05.na.novartis.net\home\moutsio1\imageAnalysis\CP\cpPipelines)

Module
- LoadData
- IdentifyPrimaryObjects
- IdentifyPrimaryObjects
- IdentifyTertiaryObjects
- EnhanceOrSuppressFeat...
- MaskImage
- IdentifyPrimaryObjects
- MeasureTexture
- MeasureGranularity
- MeasureGranularity
- MeasureObjectIntensity
- MeasureObjectIntensity
- RelateObjects
- ExportToSpreadsheet

CP_pipeline_val.cp
GranuleSimple_fromCSV.cp

Module notes

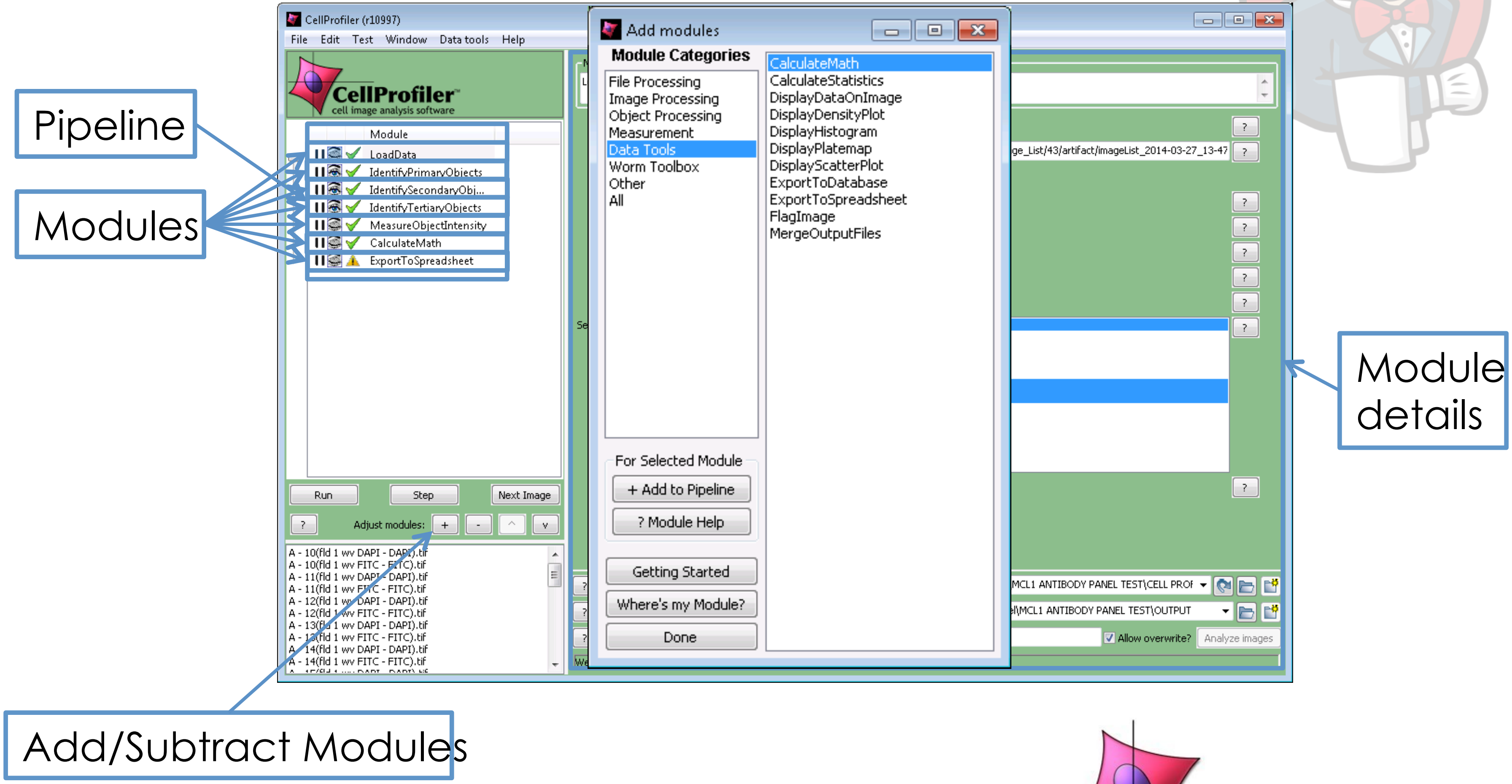
Input data file location: Elsewhere...
L:\LABS\DMP\USCA-PHUSCA-W21873\imageLists
Name of the file: imageList_2013-02-05_14_40_59.csv
Press to view CSV file contents: View...
Load images based on this data? ☑
Rescale intensities? ☑
Base image location: None
Group images by metadata? ☐
Process just a range of rows? ☑
Rows to process: 1 4

Default Input Folder: C:\temp
Default Output Folder: \\nrusca-S3015\USCA-DFS\LABDATA\LABS\DMP\USCA-PHUSCA-W21873\OUTPUT\CELLPROFIL
Output Filename: DefaultOUT.mat ☑ Allow overwrite?
Welcome to CellProfiler

- Developed and supported by the Broad Institute and a sizable scientific user community
- Supports additional imaging tools (ImageJ)

CellProfiler – general anatomy

Nuclear Translocation Assay



LoadData
IdentifyPrimaryObjects
IdentifySecondaryObj...
IdentifyTertiaryObjects
MeasureObjectIntensity
CalculateMath
ExportToSpreadsheet

A - 10(fld 1 wv DAPI - DAPI).tif
A - 10(fld 1 wv FITC - FITC).tif
A - 11(fld 1 wv DAPI - DAPI).tif
A - 11(fld 1 wv FITC - FITC).tif
A - 12(fld 1 wv DAPI - DAPI).tif
A - 12(fld 1 wv FITC - FITC).tif
A - 13(fld 1 wv DAPI - DAPI).tif
A - 13(fld 1 wv FITC - FITC).tif
A - 14(fld 1 wv DAPI - DAPI).tif
A - 14(fld 1 wv FITC - FITC).tif

File Processing
Image Processing
Object Processing
Measurement
Data Tools
Worm Toolbox
Other
All

CalculateMath
CalculateStatistics
DisplayDataOnImage
DisplayDensityPlot
DisplayHistogram
DisplayPlatemap
DisplayScatterPlot
ExportToDatabase
ExportToSpreadsheet
FlagImage
MergeOutputFiles

ge_List/43/artifact/imageList_2014-03-27_13-47

MCL1 ANTIBODY PANEL TEST\CELL PROF

l\MCL1 ANTIBODY PANEL TEST\OUTPUT

Pipeline

Modules

Module details

Add/Subtract Modules

CellProfiler
cell image analysis software

HCS Image/Data Processing

Programming and Prototyping Functional Requirements

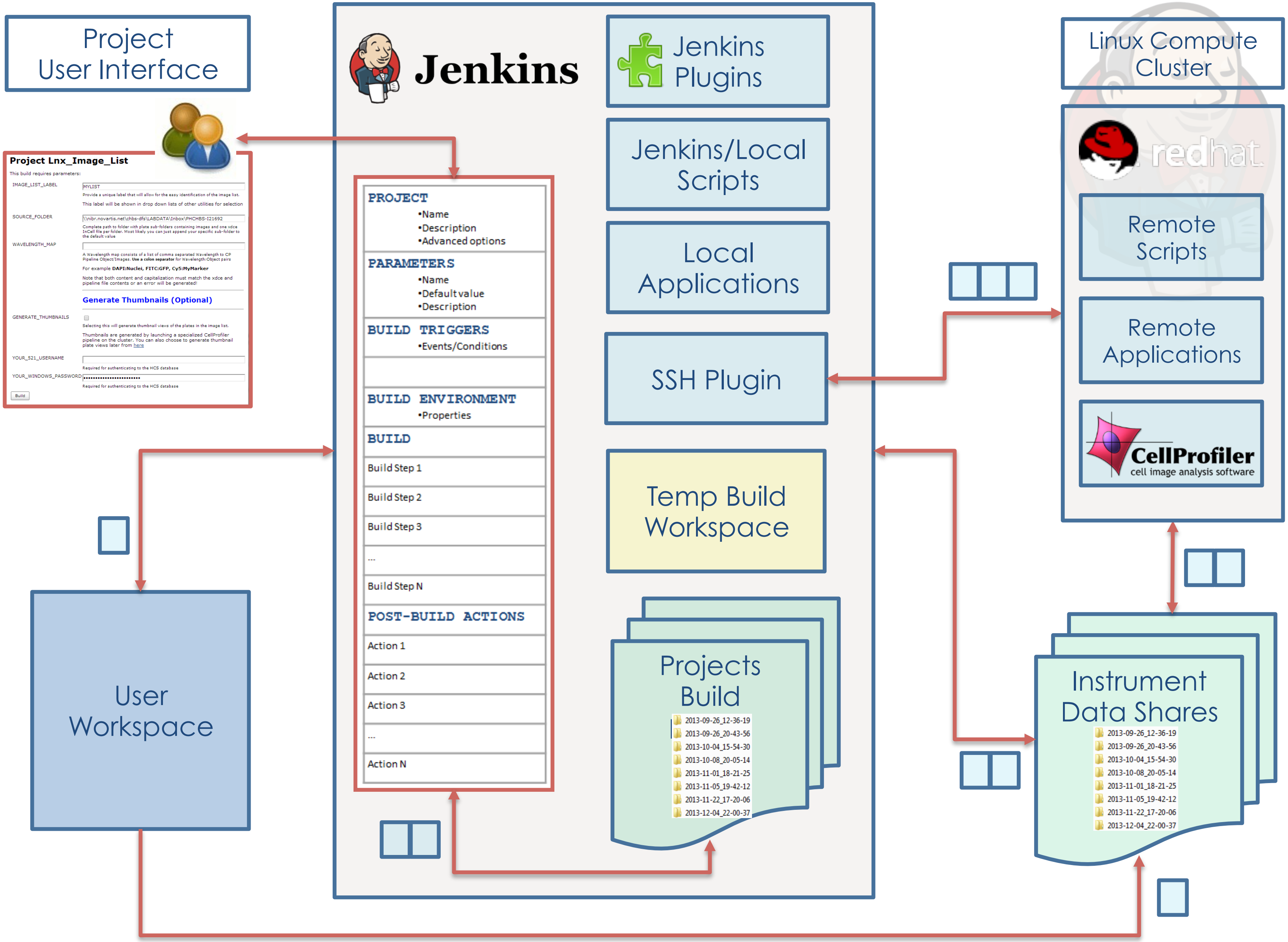
- Scripting
  - Pros
    - Quick prototyping
    - Flexibility
    - Platform independence
  - Cons
    - Unsuitable for end users
      - Requires installation of scripting tools
      - Command Line Interface

- Scripting for end users
  - Requires a user interface
    - Most UI prototypes are either
      - » hard
      - » Pretty but not functional
      - » Or...

- ...not very pretty
  - But quite functional

Why choose Jenkins-CI?

- Why Jenkins-CI?
  - Jenkins allows us to rapidly wrap any command line script or program in a web interface
    - Excellent support for Groovy a Java based, dynamic, modern scripting language
    - Straight forward integration with other languages, tools, OS, frameworks
  - Jenkins has broad community support that provides access to over 800 plugins
    - Plugins allow easy customization of Jenkins for a variety of tasks
  - Jenkins provides basic workflow and web server functionality
    - Which works well in combination with CellProfiler
  - Jenkins is used extensively by the NIBR-IT group to build all kinds of internal software
    - Many software developers know a lot about Jenkins
  - Jenkins is now emerging as a useful Bioinformatics tool
    - The BioUno project



Project User Interface

Project Lnx_Image_List

This build requires parameters:

IMAGE_LIST_LABEL: MYLIST

Provide a unique label that will allow for the easy identification of the image list.

This label will be shown in drop down lists of other utilities for selection

SOURCE_FOLDER: \\nibr.novartis.net\chbs-dfs\LABDATA\Inbox\PHCHBS-I21692

Complete path to folder with plate sub-folders containing images and one xdce InCell file per folder. Most likely you can just append your specific sub-folder to the default value

WAVELENGTH_MAP:

A Wavelength map consists of a list of comma separated Wavelength to CP Pipeline Object/Images. **Use a colon separator** for Wavelength:Object pairs

For example **DAPI:Nuclei, FITC:GFP, Cy5:MyMarker**

Note that both content and capitalization must match the xdce and pipeline file contents or an error will be generated!

Generate Thumbnails (Optional)

GENERATE_THUMBNAILS: □

Selecting this will generate thumbnail views of the plates in the image list.

Thumbnails are generated by launching a specialized CellProfiler pipeline on the cluster. You can also choose to generate thumbnail plate views later from here

YOUR_521_USERNAME:

Required for authenticating to the HCS database

YOUR_WINDOWS_PASSWORD: ••••••••••••••••••••••••

Required for authenticating to the HCS database

Build

User Workspace

Jenkins

PROJECT
- Name
- Description
- Advanced options

PARAMETERS
- Name
- Default value
- Description

BUILD TRIGGERS
- Events/Conditions

BUILD ENVIRONMENT
- Properties

BUILD

Build Step 1

Build Step 2

Build Step 3

...

Build Step N

POST-BUILD ACTIONS

Action 1

Action 2

Action 3

...

Action N

Jenkins Plugins

Jenkins/Local Scripts

Local Applications

SSH Plugin

Temp Build Workspace

Projects Build

2013-09-26_12-36-19
2013-09-26_20-43-56
2013-10-04_15-54-30
2013-10-08_20-05-14
2013-11-01_18-21-25
2013-11-05_19-42-12
2013-11-22_17-20-06
2013-12-04_22-00-37

Linux Compute Cluster

redhat

Remote Scripts

Remote Applications

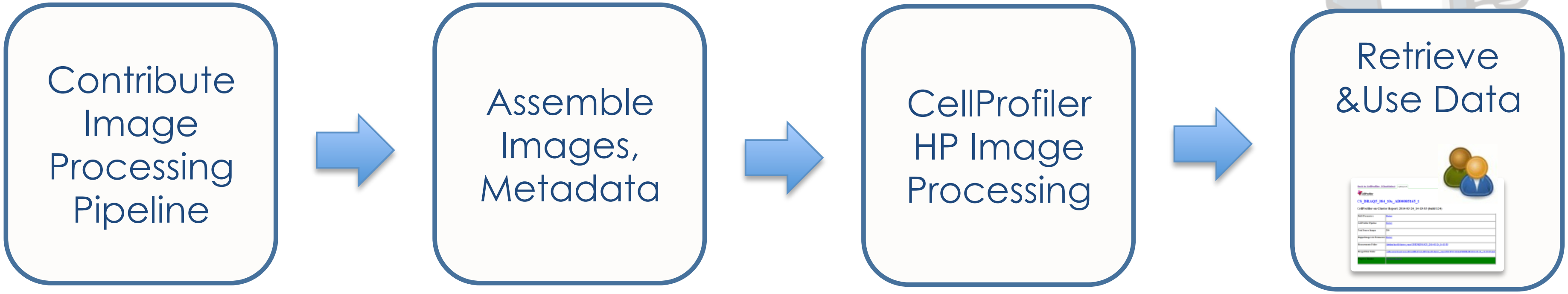
CellProfiler
cell image analysis software

Instrument Data Shares

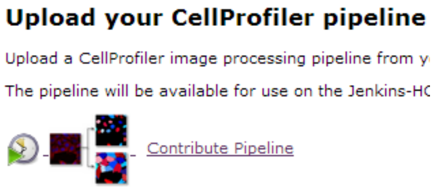
2013-09-26_12-36-19
2013-09-26_20-43-56
2013-10-04_15-54-30
2013-10-08_20-05-14
2013-11-01_18-21-25
2013-11-05_19-42-12
2013-11-22_17-20-06
2013-12-04_22-00-37

Jenkins-CellProfiler

HP Image Processing Workflow: Outline



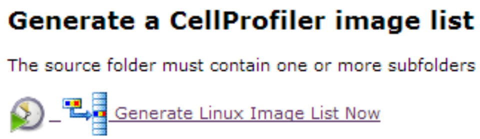
Contribute Image Processing Pipeline → Assemble Images, Metadata → CellProfiler HP Image Processing → Retrieve &Use Data



Upload your CellProfiler pipeline

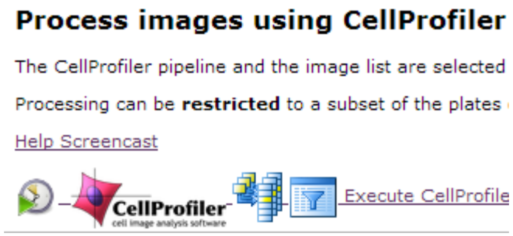
Upload a CellProfiler image processing pipeline from y

The pipeline will be available for use on the Jenkins-HC



Generate a CellProfiler image list

The source folder must contain one or more subfolders



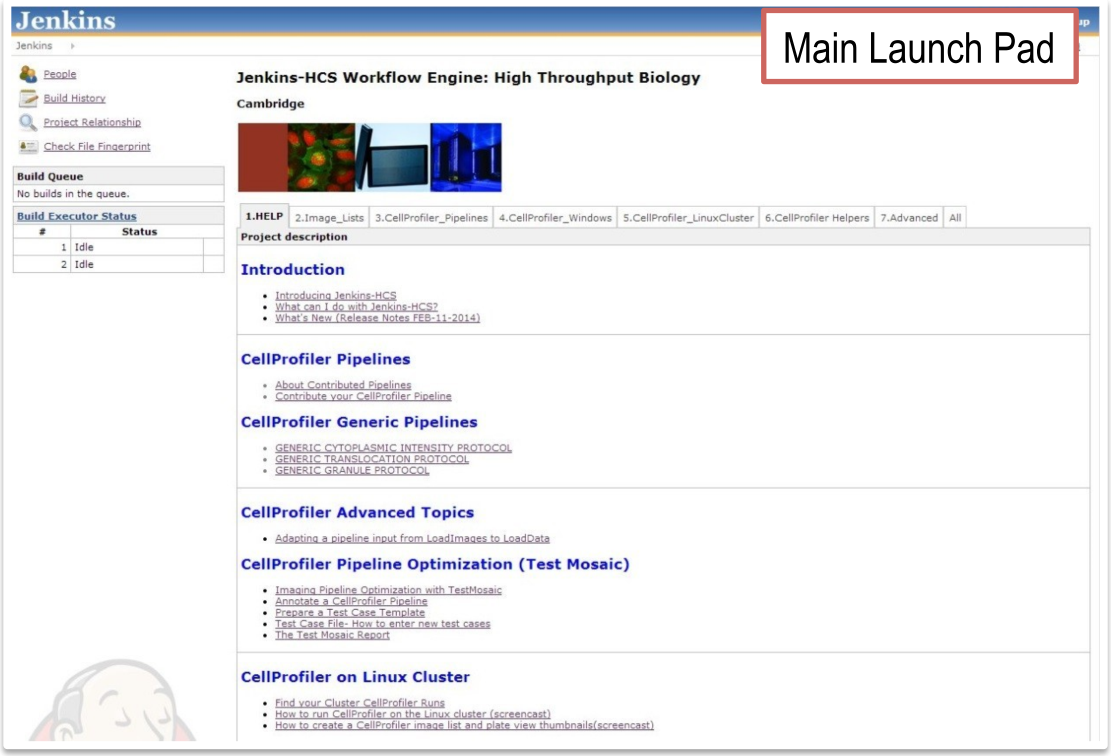
Process images using CellProfiler

The CellProfiler pipeline and the image list are selected

Processing can be **restricted** to a subset of the plates

Jenkins-HCS Workflow Engine

High Level UI Components



Main Launch Pad

Jenkins
Jenkins ›

- People
- Build History
- Project Relationship
- Check File Fingerprint

Build Queue

No builds in the queue.

Build Executor Status

#	Status
1	Idle
2	Idle

Jenkins-HCS Workflow Engine: High Throughput Biology

Cambridge

1.HELP | 2.Image_Lists | 3.CellProfiler_Pipelines | 4.CellProfiler_Windows | 5.CellProfiler_LinuxCluster | 6.CellProfiler Helpers | 7.Advanced | All

Project description

Introduction

- Introducing Jenkins-HCS
- What can I do with Jenkins-HCS?
- What's New (Release Notes FEB-11-2014)

CellProfiler Pipelines

- About Contributed Pipelines
- Contribute your CellProfiler Pipeline

CellProfiler Generic Pipelines

- GENERIC CYTOPLASMIC INTENSITY PROTOCOL
- GENERIC TRANSLOCATION PROTOCOL
- GENERIC GRANULE PROTOCOL

CellProfiler Advanced Topics

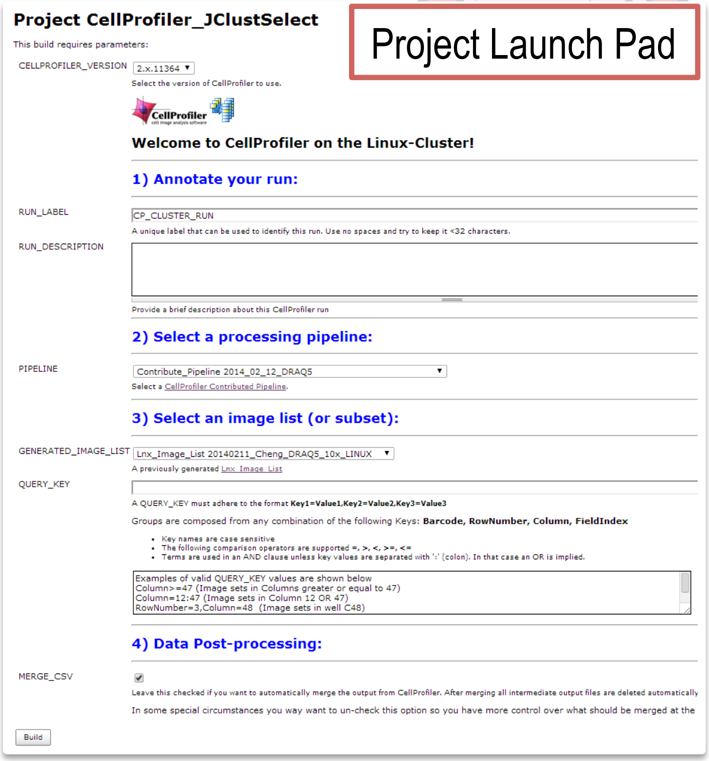
- Adapting a pipeline input from LoadImages to LoadData

CellProfiler Pipeline Optimization (Test Mosaic)

- Imaging Pipeline Optimization with TestMosaic
- Annotate a CellProfiler Pipeline
- Prepare a Test Case Template
- Test Case File- How to enter new test cases
- The Test Mosaic Report

CellProfiler on Linux Cluster

- Find your Cluster CellProfiler Runs
- How to run CellProfiler on the Linux cluster (screencast)
- How to create a CellProfiler image list and plate view thumbnails(screencast)



Project Launch Pad

Project CellProfiler_JClustSelect

This build requires parameters:

CELLPROFILER_VERSION: 2.x.11364 ▾
Select the version of CellProfiler to use.

CellProfiler cell image analysis software

Welcome to CellProfiler on the Linux-Cluster!

1) Annotate your run:

RUN_LABEL: CP_CLUSTER_RUN
A unique label that can be used to identify this run. Use no spaces and try to keep it <32 characters.

RUN_DESCRIPTION:
Provide a brief description about this CellProfiler run

2) Select a processing pipeline:

PIPELINE: Contribute_Pipeline 2014_02_12_DRAQ5 ▾
Select a CellProfiler Contributed Pipeline.

3) Select an image list (or subset):

GENERATED_IMAGE_LIST: Lnx_Image_List 20140211_Cheng_DRAQ5_10x_LINUX ▾
A previously generated Lnx_Image_List

QUERY_KEY:
A QUERY_KEY must adhere to the format **Key1=Value1,Key2=Value2,Key3=Value3**

Groups are composed from any combination of the following Keys: **Barcode, RowNumber, Column, FieldIndex**

- Key names are case sensitive
- The following comparison operators are supported =, >, <, >=, <=
- Terms are used in an AND clause unless key values are separated with ':' (colon). In that case an OR is implied.

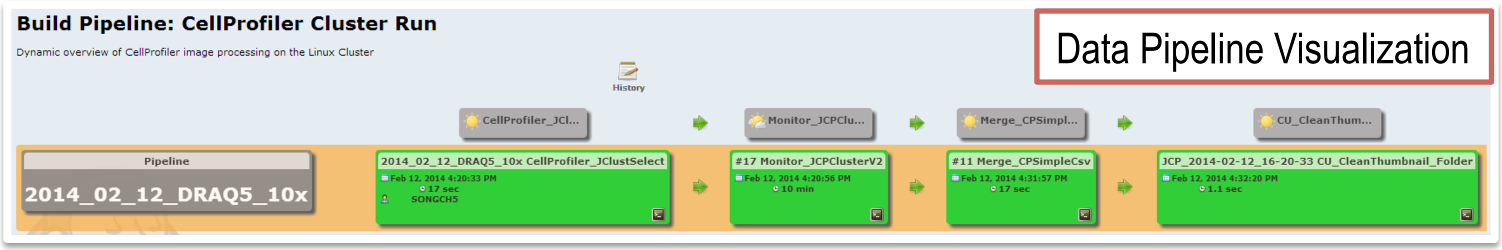
```
Examples of valid QUERY_KEY values are shown below
Column>=47 (Image sets in Columns greater or equal to 47)
Column=12:47 (Image sets in Column 12 OR 47)
RowNumber=3,Column=48  (Image sets in well C48)
```

4) Data Post-processing:

MERGE_CSV: ☑
Leave this checked if you want to automatically merge the output from CellProfiler. After merging all intermediate output files are deleted automatically

In some special circumstances you way want to un-check this option so you have more control over what should be merged at the

Build



Data Pipeline Visualization

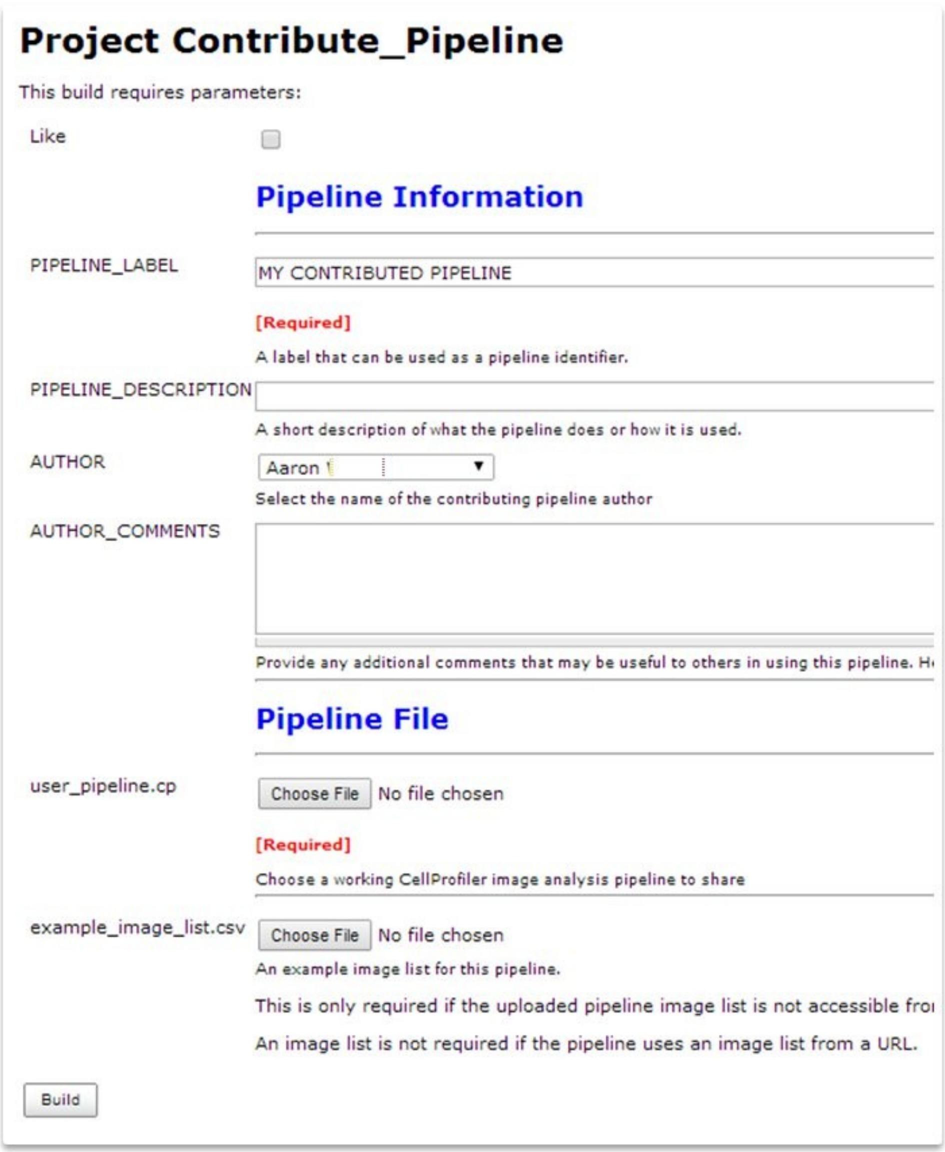
Build Pipeline: CellProfiler Cluster Run

Dynamic overview of CellProfiler image processing on the Linux Cluster

Typical Workflow

Step 1: Contribute a pipeline

- Project: Contribute_Pipeline
  - Upload and annotate a standard CellProfiler image analysis pipeline. Uploaded pipelines are usable in other projects
  - Assumptions
    - The pipeline has been designed and successfully tested on the CellProfiler desktop client
  - Outcome
    - The CellProfiler pipeline file will be uploaded and stored on Jenkins
    - Additional annotation will be extracted and attached to the pipeline



Project Contribute_Pipeline

This build requires parameters:

Like □

Pipeline Information

PIPELINE_LABEL: MY CONTRIBUTED PIPELINE

[Required]

A label that can be used as a pipeline identifier.

PIPELINE_DESCRIPTION:

A short description of what the pipeline does or how it is used.

AUTHOR: Aaron ▼

Select the name of the contributing pipeline author

AUTHOR_COMMENTS:

Provide any additional comments that may be useful to others in using this pipeline. H

Pipeline File

user_pipeline.cp: Choose File No file chosen

[Required]

Choose a working CellProfiler image analysis pipeline to share

example_image_list.csv: Choose File No file chosen

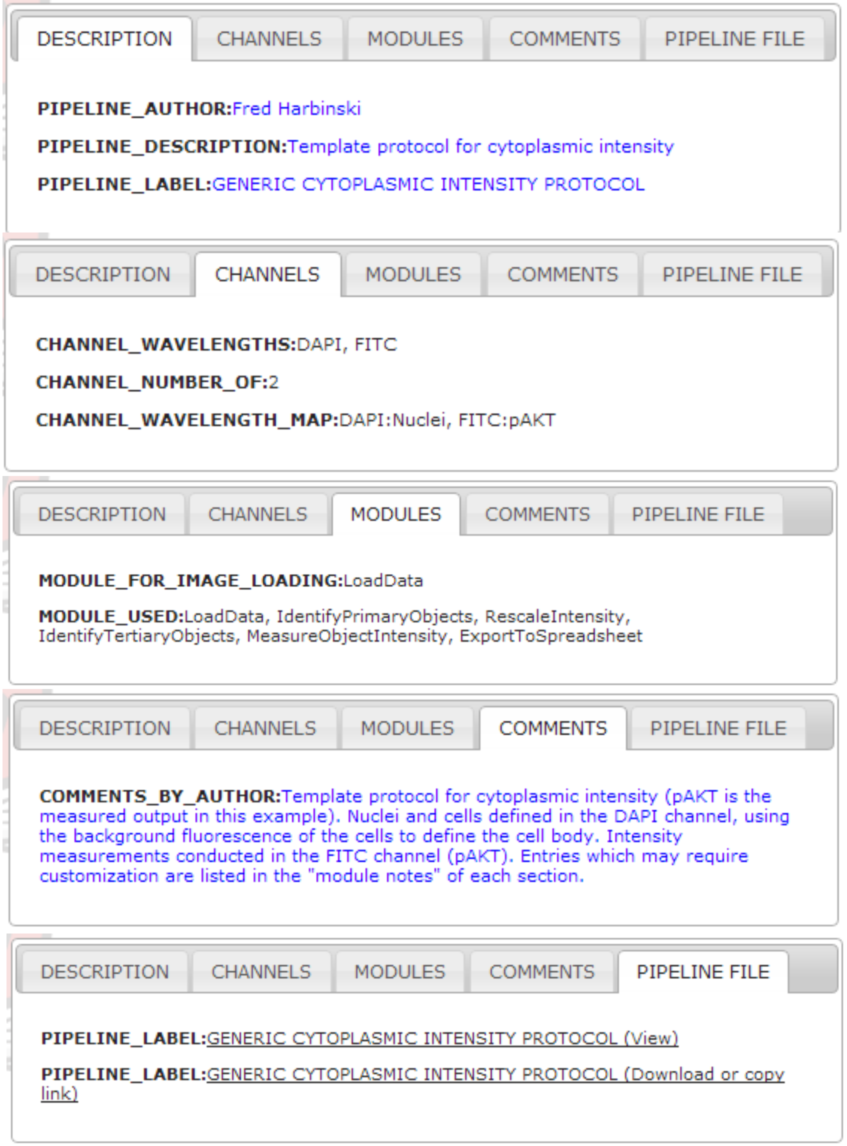
An example image list for this pipeline.

This is only required if the uploaded pipeline image list is not accessible fro

An image list is not required if the pipeline uses an image list from a URL.

Build report from a contributed pipeline

Uses: Summary Display Plugin



PIPELINE_AUTHOR: Fred Harbinski

PIPELINE_DESCRIPTION: Template protocol for cytoplasmic intensity

PIPELINE_LABEL:GENERIC CYTOPLASMIC INTENSITY PROTOCOL

CHANNEL_WAVELENGTHS:DAPI, FITC

CHANNEL_NUMBER_OF:2

CHANNEL_WAVELENGTH_MAP:DAPI:Nuclei, FITC:pAKT

MODULE_FOR_IMAGE_LOADING: LoadData

MODULE_USED: LoadData, IdentifyPrimaryObjects, RescaleIntensity, IdentifyTertiaryObjects, MeasureObjectIntensity, ExportToSpreadsheet

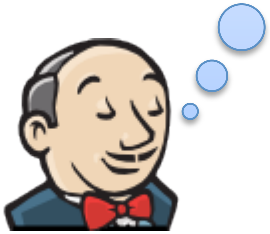
COMMENTS_BY_AUTHOR: Template protocol for cytoplasmic intensity (pAKT is the measured output in this example). Nuclei and cells defined in the DAPI channel, using the background fluorescence of the cells to define the cell body. Intensity measurements conducted in the FITC channel (pAKT). Entries which may require customization are listed in the "module notes" of each section.

PIPELINE_LABEL:GENERIC CYTOPLASMIC INTENSITY PROTOCOL (View)

PIPELINE_LABEL:GENERIC CYTOPLASMIC INTENSITY PROTOCOL (Download or copy link)

- Contributed pipelines are annotated by a combination of user provided and auto-extracted metadata
  - Presented as a tab panel
  - Pipeline can be downloaded and further modified

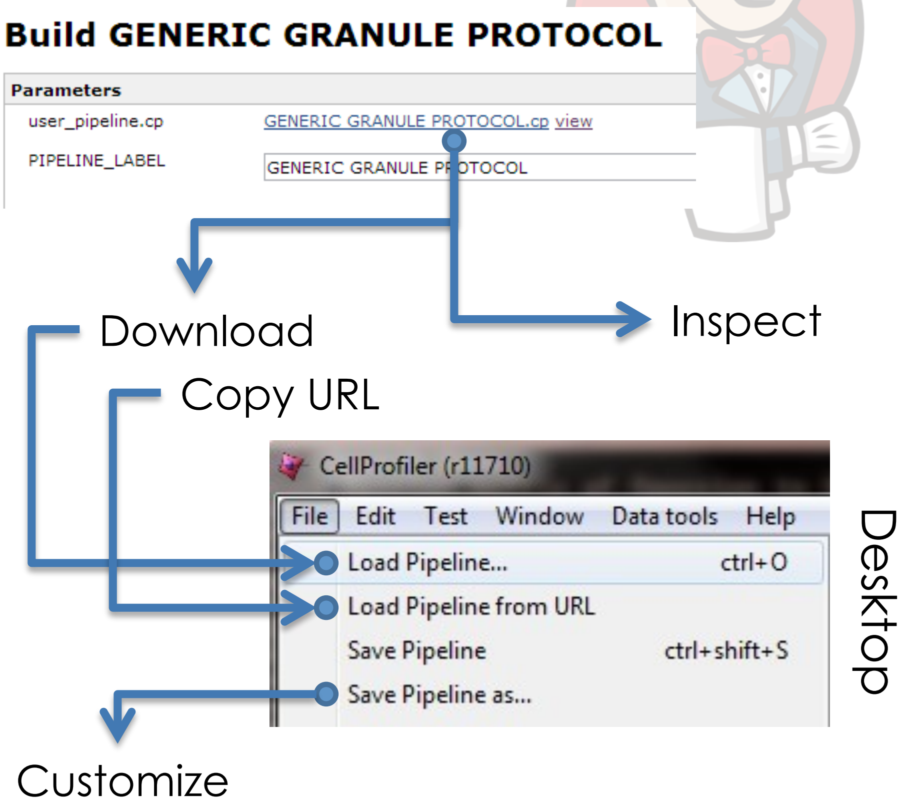
Use PIPELINE FILE tab to download or quickly browse the pipeline



A CellProfiler Pipeline from Jenkins Server

Additional Usage

- CellProfiler pipelines on the Jenkins server can be used as follows:
  - For inspection
  - For re-use
    - On CellProfiler desktop client
    - On Jenkins-CellProfiler
  - For further experimentation
    - Load in desktop client and further customize

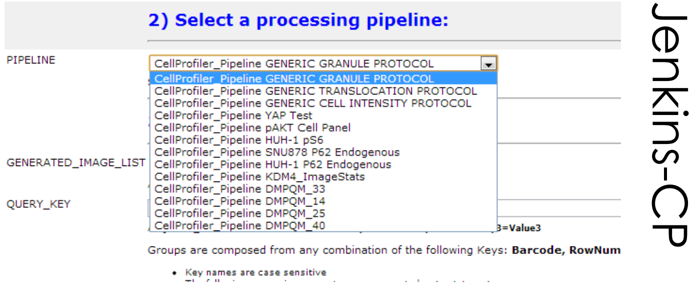


Build GENERIC GRANULE PROTOCOL

Parameters

user_pipeline.cp GENERIC GRANULE PROTOCOL.cp view

PIPELINE_LABEL GENERIC GRANULE PROTOCOL



2) Select a processing pipeline:

PIPELINE: CellProfiler_Pipeline GENERIC GRANULE PROTOCOL

CellProfiler_Pipeline GENERIC GRANULE PROTOCOL
CellProfiler_Pipeline GENERIC TRANSLOCATION PROTOCOL
CellProfiler_Pipeline GENERIC CELL INTENSITY PROTOCOL
CellProfiler_Pipeline YAP Test
CellProfiler_Pipeline pAKT Cell Panel
CellProfiler_Pipeline HUH-1 pS6
CellProfiler_Pipeline SNU878 P62 Endogenous
CellProfiler_Pipeline HUH-1 P62 Endogenous
CellProfiler_Pipeline KDM4_ImageStats
CellProfiler_Pipeline DMPQM_33
CellProfiler_Pipeline DMPQM_14
CellProfiler_Pipeline DMPQM_25
CellProfiler_Pipeline DMPQM_40

GENERATED_IMAGE_LIST

QUERY_KEY

3=Value3

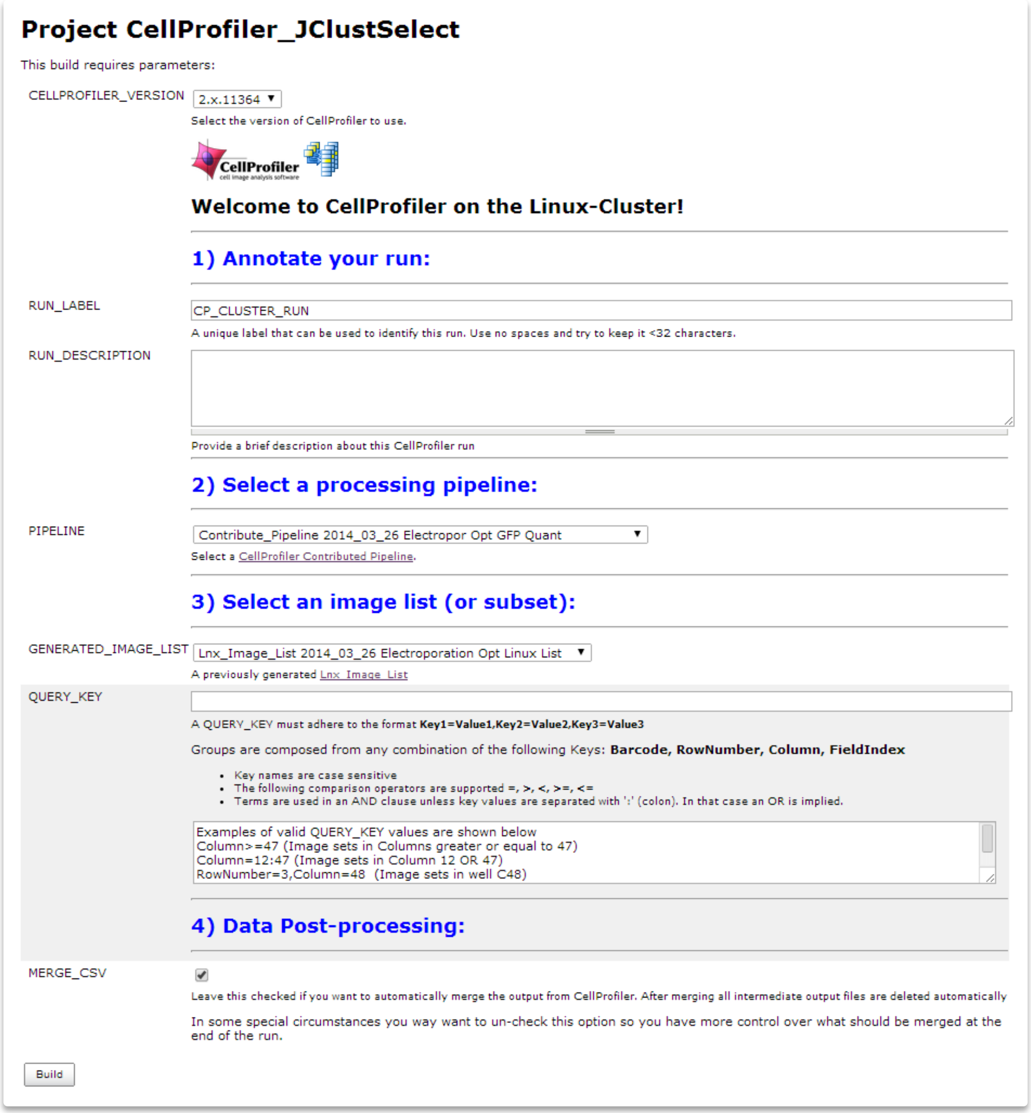
Groups are composed from any combination of the following Keys: **Barcode, RowNum**

- Key names are case sensitive

Typical Workflow

Step 3: Execute CellProfiler on the Linux Cluster

- Project: CellProfiler_JClustSelect
  - Executes a series of image processing steps using the Jenkins-CI CellProfiler
  - Uses the SSH Plugin
  - Typical Assumptions
    - CellProfiler pipeline and a CP formatted image list are stored on the Jenkins server
      - Jenkins build artifacts
  - Outcome
    - Summary report
    - A file containing combined measurements from all the images processed.
      - Results file is in CSV format



Project CellProfiler_JClustSelect

This build requires parameters:

CELLPROFILER_VERSION 2.x.11364 ▼

Select the version of CellProfiler to use.

Welcome to CellProfiler on the Linux-Cluster!

1) Annotate your run:

RUN_LABEL CP_CLUSTER_RUN

A unique label that can be used to identify this run. Use no spaces and try to keep it <32 characters.

RUN_DESCRIPTION

Provide a brief description about this CellProfiler run

2) Select a processing pipeline:

PIPELINE Contribute_Pipeline 2014_03_26 Electropor Opt GFP Quant ▼

Select a CellProfiler Contributed Pipeline.

3) Select an image list (or subset):

GENERATED_IMAGE_LIST Lnx_Image_List 2014_03_26 Electroporation Opt Linux List ▼

A previously generated Lnx_Image_List

QUERY_KEY

A QUERY_KEY must adhere to the format **Key1=Value1,Key2=Value2,Key3=Value3**

Groups are composed from any combination of the following Keys: **Barcode, RowNumber, Column, FieldIndex**

- Key names are case sensitive
- The following comparison operators are supported =, >, <, >=, <=
- Terms are used in an AND clause unless key values are separated with ':' (colon). In that case an OR is implied.

```
Examples of valid QUERY_KEY values are shown below
Column>=47 (Image sets in Columns greater or equal to 47)
Column=12:47 (Image sets in Column 12 OR 47)
RowNumber=3,Column=48  (Image sets in well C48)
```

4) Data Post-processing:

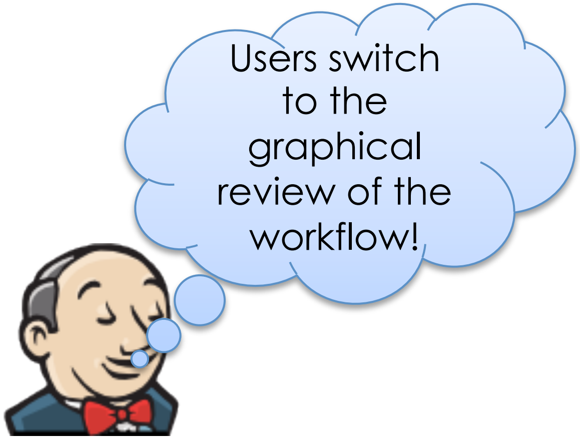
MERGE_CSV ☑

Leave this checked if you want to automatically merge the output from CellProfiler. After merging all intermediate output files are deleted automatically

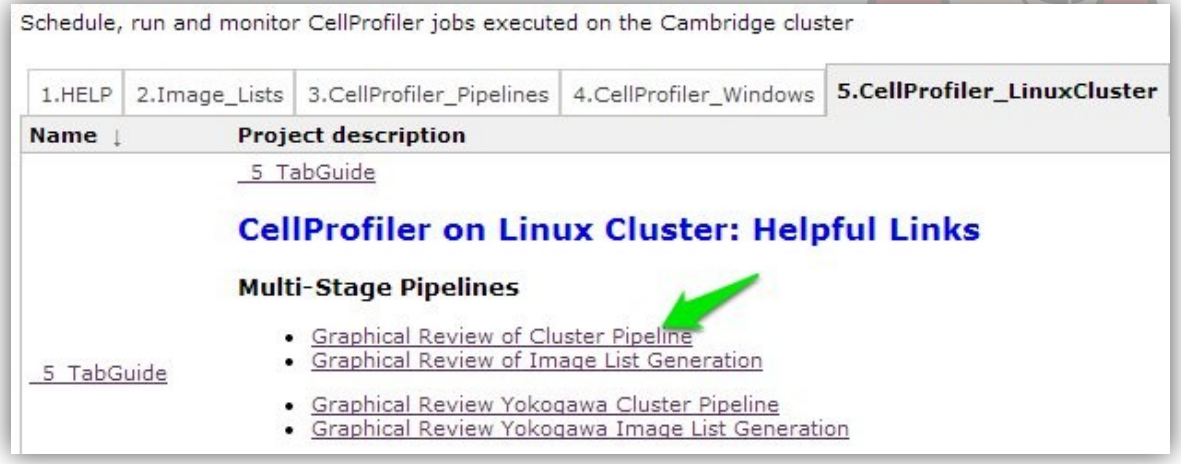
In some special circumstances you way want to un-check this option so you have more control over what should be merged at the end of the run.

Monitoring CellProfiler runs on the cluster

Uses: Build Pipeline Plugin

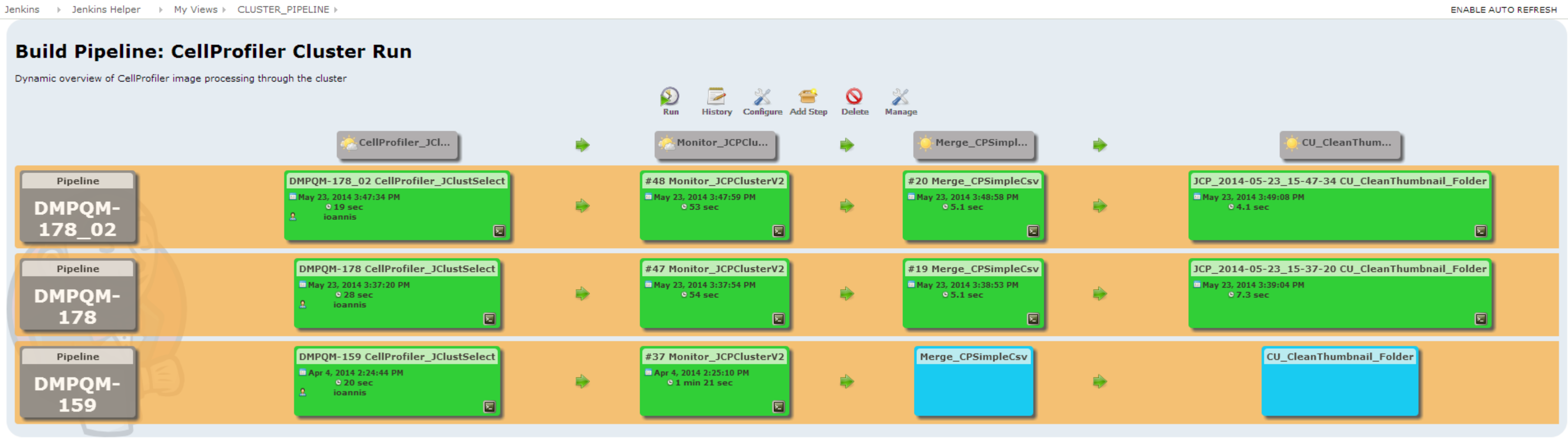


Users switch to the graphical review of the workflow!



Schedule, run and monitor CellProfiler jobs executed on the Cambridge cluster

Name ↓	Project description
5_TabGuide	5_TabGuide **CellProfiler on Linux Cluster: Helpful Links** **Multi-Stage Pipelines** • Graphical Review of Cluster Pipeline • Graphical Review of Image List Generation • Graphical Review Yokogawa Cluster Pipeline • Graphical Review Yokogawa Image List Generation



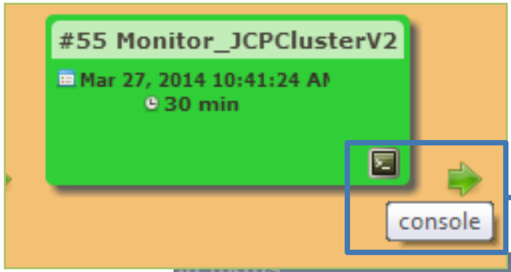
Build Pipeline: CellProfiler Cluster Run

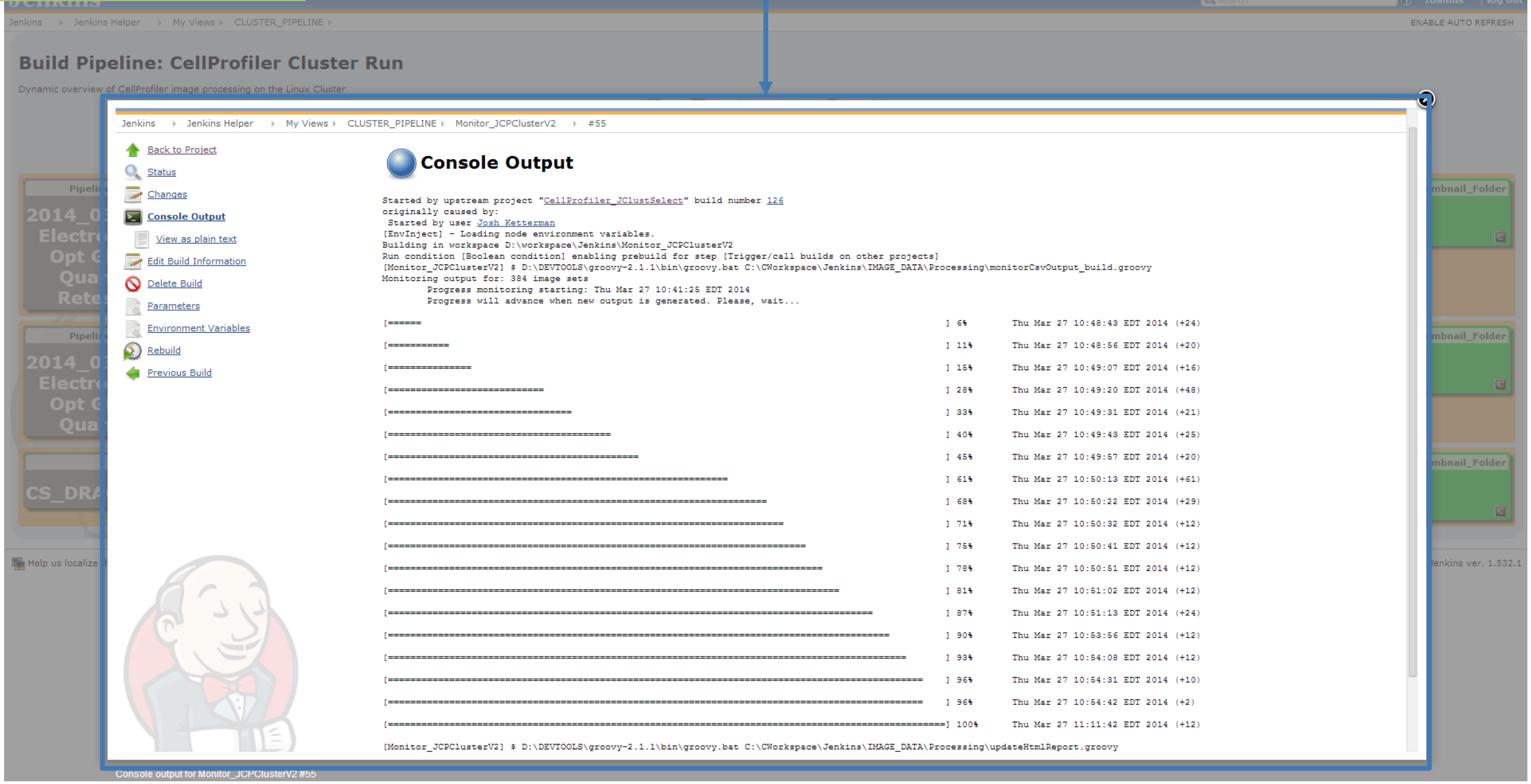
Dynamic overview of CellProfiler image processing through the cluster

	CellProfiler_JCl...	→	Monitor_JCPClu...	→	Merge_CPSimpl...	→	CU_CleanThum...
Pipeline DMPQM-178_02	DMPQM-178_02 CellProfiler_JClustSelect May 23, 2014 3:47:34 PM 19 sec ioannis	→	#48 Monitor_JCPClusterV2 May 23, 2014 3:47:59 PM 53 sec	→	#20 Merge_CPSimpleCsv May 23, 2014 3:48:58 PM 5.1 sec	→	JCP_2014-05-23_15-47-34 CU_CleanThumbnail_Folder May 23, 2014 3:49:08 PM 4.1 sec
Pipeline DMPQM-178	DMPQM-178 CellProfiler_JClustSelect May 23, 2014 3:37:20 PM 28 sec ioannis	→	#47 Monitor_JCPClusterV2 May 23, 2014 3:37:54 PM 54 sec	→	#19 Merge_CPSimpleCsv May 23, 2014 3:38:53 PM 5.1 sec	→	JCP_2014-05-23_15-37-20 CU_CleanThumbnail_Folder May 23, 2014 3:39:04 PM 7.3 sec
Pipeline DMPQM-159	DMPQM-159 CellProfiler_JClustSelect Apr 4, 2014 2:24:44 PM 20 sec ioannis	→	#37 Monitor_JCPClusterV2 Apr 4, 2014 2:25:10 PM 1 min 21 sec	→	Merge_CPSimpleCsv	→	CU_CleanThumbnail_Folder

Monitoring CellProfiler runs on the cluster

Uses: [Build Pipeline Plugin](#) and the Console





Build Pipeline: CellProfiler Cluster Run

Dynamic overview of CellProfiler image processing on the Linux Cluster

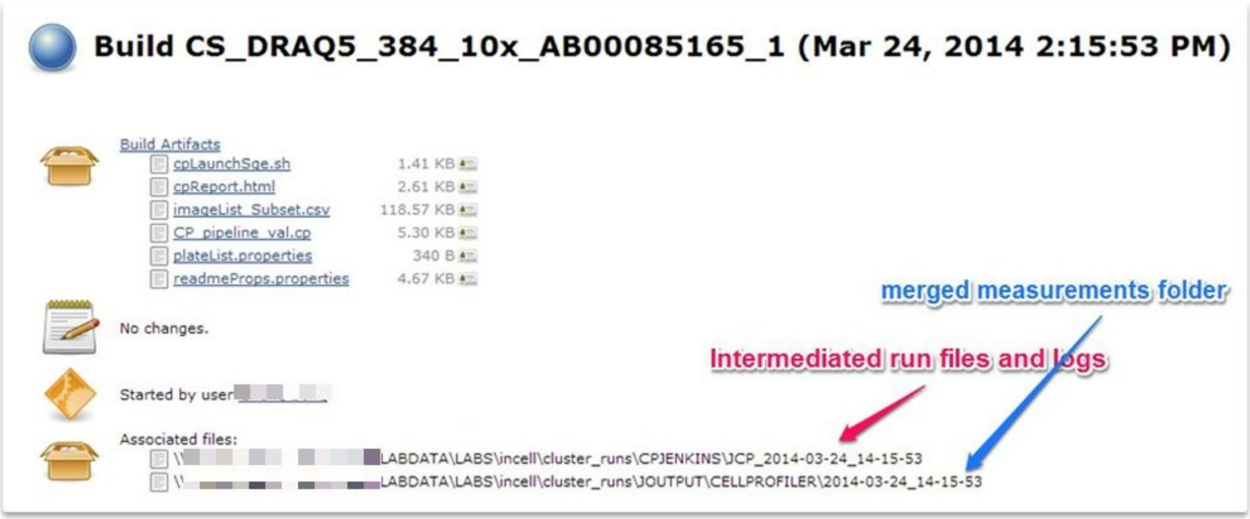
Console Output

```
Started by upstream project "CellProfiler_JClustSelect" build number 126
originally caused by:
 Started by user Josh Ketterman
[EnvInject] - Loading node environment variables.
Building in workspace D:\workspace\Jenkins\Monitor_JCPClusterV2
Run condition [Boolean condition] enabling prebuild for step [Trigger/call builds on other projects]
[Monitor_JCPClusterV2] $ D:\DEVTOOLS\groovy-2.1.1\bin\groovy.bat C:\CWorkspace\Jenkins\IMAGE_DATA\Processing\monitorCsvOutput_build.groovy
Monitoring output for: 384 image sets
	Progress monitoring starting: Thu Mar 27 10:41:25 EDT 2014
	Progress will advance when new output is generated. Please, wait...

[======                                                                                              ] 6%        Thu Mar 27 10:48:43 EDT 2014 (+24)

[===========                                                                                         ] 11%       Thu Mar 27 10:48:56 EDT 2014 (+20)

[===============                                                                                     ] 15%       Thu Mar 27 10:49:07 EDT 2014 (+16)

[============================                                                                        ] 28%       Thu Mar 27 10:49:20 EDT 2014 (+48)

[=================================                                                                   ] 33%       Thu Mar 27 10:49:31 EDT 2014 (+21)

[========================================                                                            ] 40%       Thu Mar 27 10:49:43 EDT 2014 (+25)

[=============================================                                                       ] 45%       Thu Mar 27 10:49:57 EDT 2014 (+20)

[=============================================================                                       ] 61%       Thu Mar 27 10:50:13 EDT 2014 (+61)

[====================================================================                                ] 68%       Thu Mar 27 10:50:22 EDT 2014 (+29)

[=======================================================================                             ] 71%       Thu Mar 27 10:50:32 EDT 2014 (+12)

[===========================================================================                         ] 75%       Thu Mar 27 10:50:41 EDT 2014 (+12)

[==============================================================================                      ] 78%       Thu Mar 27 10:50:51 EDT 2014 (+12)

[=================================================================================                   ] 81%       Thu Mar 27 10:51:02 EDT 2014 (+12)

[=======================================================================================             ] 87%       Thu Mar 27 10:51:13 EDT 2014 (+24)

[==========================================================================================          ] 90%       Thu Mar 27 10:53:56 EDT 2014 (+12)

[=============================================================================================       ] 93%       Thu Mar 27 10:54:08 EDT 2014 (+12)

[================================================================================================    ] 96%       Thu Mar 27 10:54:31 EDT 2014 (+10)

[================================================================================================    ] 96%       Thu Mar 27 10:54:42 EDT 2014 (+2)

[====================================================================================================] 100%      Thu Mar 27 11:11:42 EDT 2014 (+12)

[Monitor_JCPClusterV2] $ D:\DEVTOOLS\groovy-2.1.1\bin\groovy.bat C:\CWorkspace\Jenkins\IMAGE_DATA\Processing\updateHtmlReport.groovy
```

Console output for Monitor_JCPClusterV2 #55

Run Report & Measurement Retrieval

Uses: Associated Files and HTML Publisher plugins



Build CS_DRAQ5_384_10x_AB00085165_1 (Mar 24, 2014 2:15:53 PM)

Build Artifacts

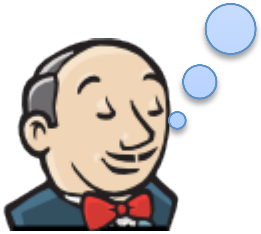
cpLaunchSge.sh	1.41 KB
cpReport.html	2.61 KB
imageList_Subset.csv	118.57 KB
CP_pipeline_val.cp	5.30 KB
plateList.properties	340 B
readmeProps.properties	4.67 KB

No changes.

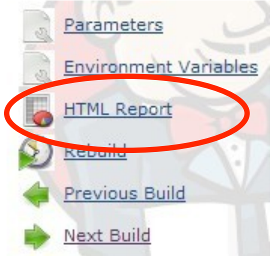
Started by user [redacted]

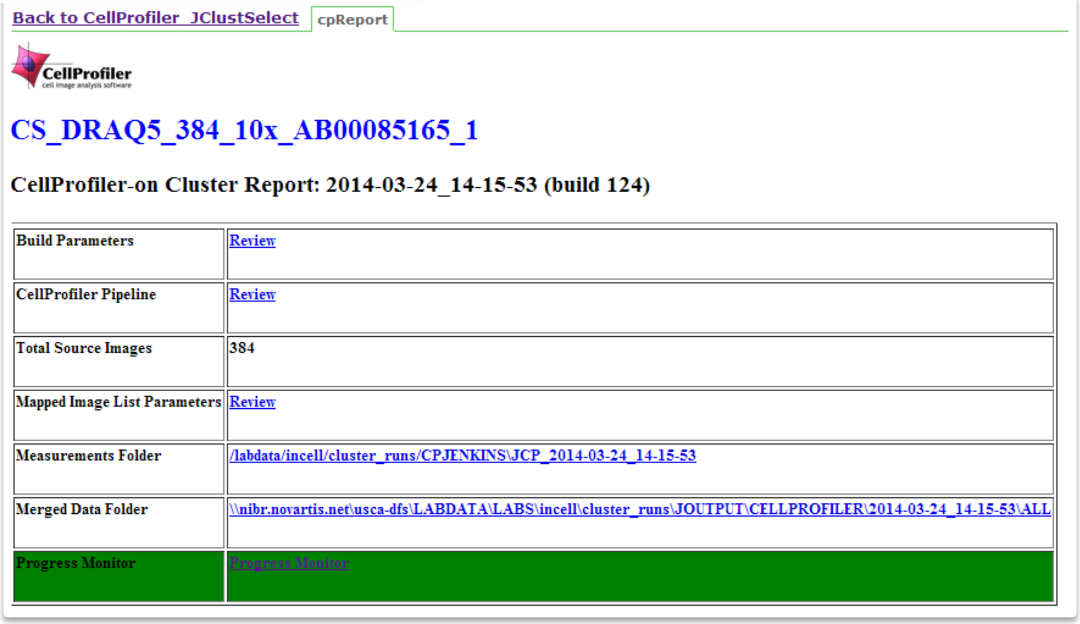
Associated files:

- \\[redacted]LABDATA\LABS\incell\cluster_runs\CPJENKINS\JCP_2014-03-24_14-15-53
- \\[redacted]LABDATA\LABS\incell\cluster_runs\JOUTPUT\CELLPROFILER\2014-03-24_14-15-53



If all goes well final results are found in the **merged measurements** folder





Back to CellProfiler_JClustSelect
cpReport
CellProfiler
cell image analysis software

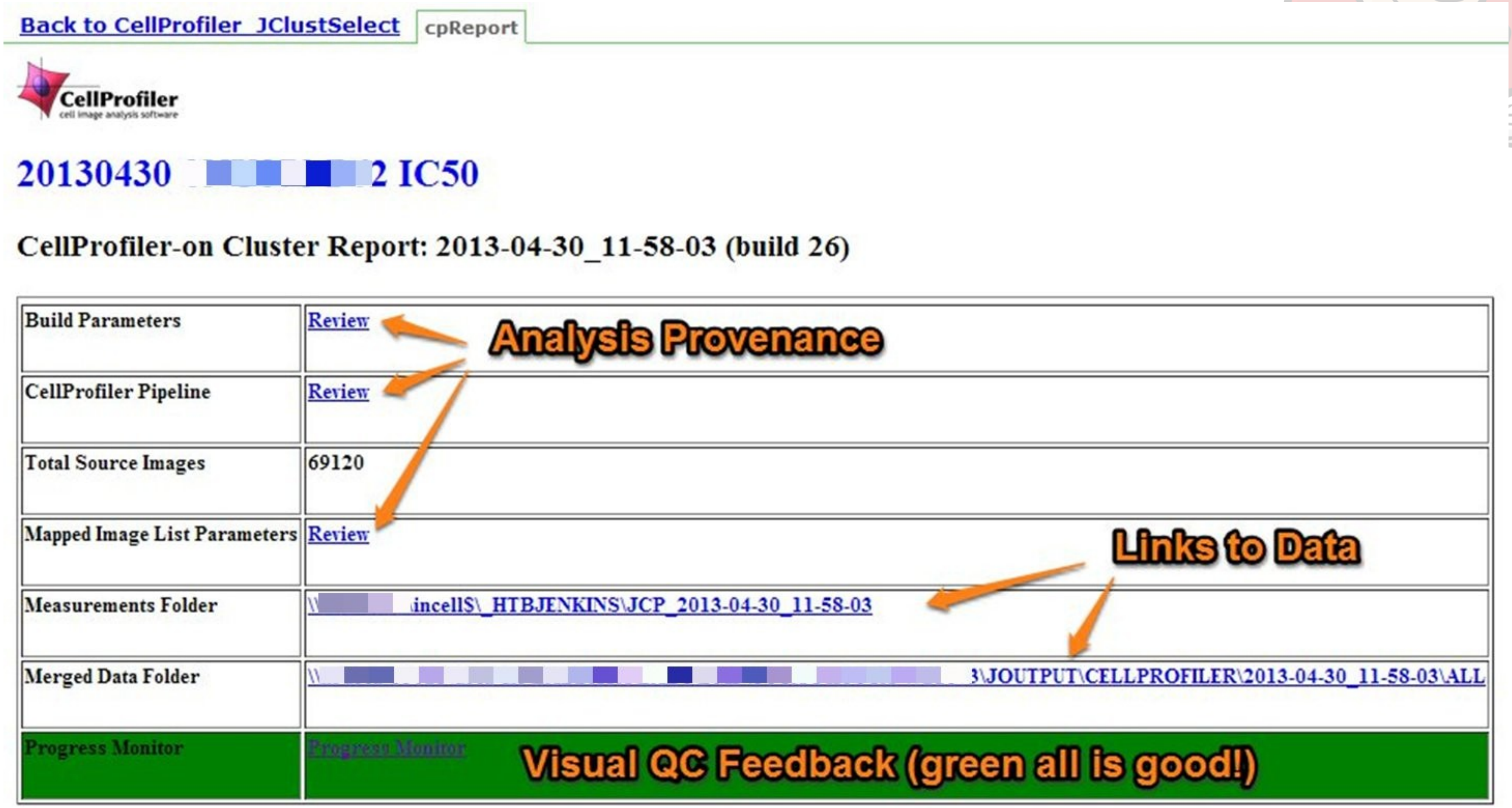
CS_DRAQ5_384_10x_AB00085165_1

CellProfiler-on Cluster Report: 2014-03-24_14-15-53 (build 124)

Build Parameters	Review
CellProfiler Pipeline	Review
Total Source Images	384
Mapped Image List Parameters	Review
Measurements Folder	/labdata/incell/cluster_runs/CPJENKINS\JCP_2014-03-24_14-15-53
Merged Data Folder	\\nibr.novartis.net\usca-dfs\LABDATA\LABS\incell\cluster_runs\JOUTPUT\CELLPROFILER\2014-03-24_14-15-53\ALL
Progress Monitor	Progress Monitor

Jenkins –CI: CellProfiler Image Processing

Uses: HTML Publisher plugin



Back to CellProfiler_JClustSelect | cpReport

CellProfiler
cell image analysis software

20130430 [redacted]2 IC50

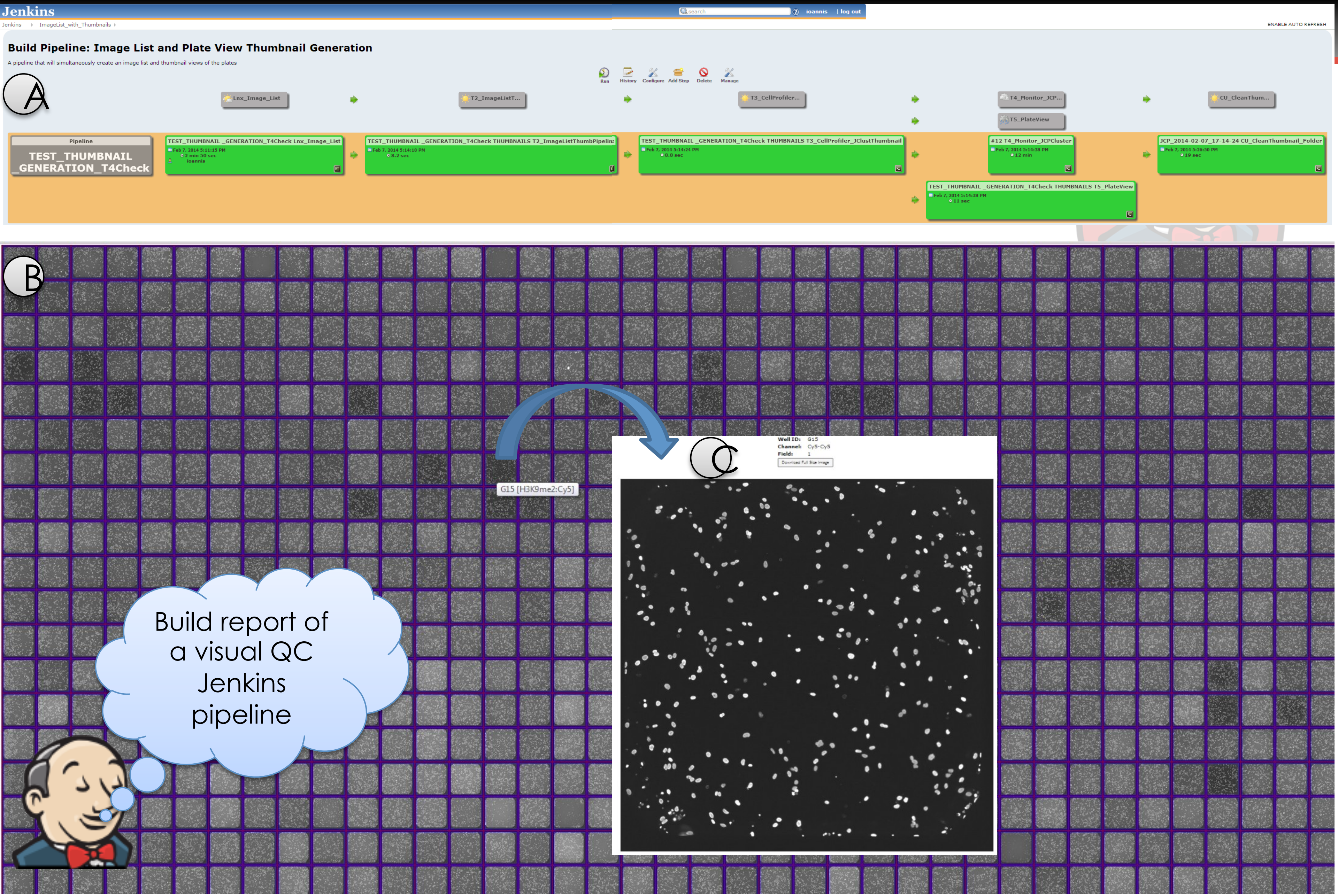
CellProfiler-on Cluster Report: 2013-04-30_11-58-03 (build 26)

Build Parameters	Review
CellProfiler Pipeline	Review
Total Source Images	69120
Mapped Image List Parameters	Review
Measurements Folder	\\[redacted]incell$_HTBJENKINS\JCP_2013-04-30_11-58-03
Merged Data Folder	\\[redacted]3\JOUTPUT\CELLPROFILER\2013-04-30_11-58-03\ALL
Progress Monitor	Progress Monitor

Analysis Provenance

Links to Data

Visual QC Feedback (green all is good!)



Jenkins

Jenkins › ImageList_with_Thumbnails ›

Build Pipeline: Image List and Plate View Thumbnail Generation

A pipeline that will simultaneously create an image list and thumbnail views of the plates

Pipeline

TEST_THUMBNAIL _GENERATION_T4Check

TEST_THUMBNAIL _GENERATION_T4Check Lnx_Image_List
Feb 7, 2014 5:11:15 PM
2 min 50 sec
ioannis

TEST_THUMBNAIL _GENERATION_T4Check THUMBNAILS T2_ImageListThumbPipeline
Feb 7, 2014 5:14:10 PM
8.2 sec

TEST_THUMBNAIL _GENERATION_T4Check THUMBNAILS T3_CellProfiler_JClustThumbnail
Feb 7, 2014 5:14:24 PM
8.8 sec

#12 T4_Monitor_JCPCluster
Feb 7, 2014 5:14:38 PM
12 min

JCP_2014-02-07_17-14-24 CU_CleanThumbnail_Folder
Feb 7, 2014 5:26:50 PM
19 sec

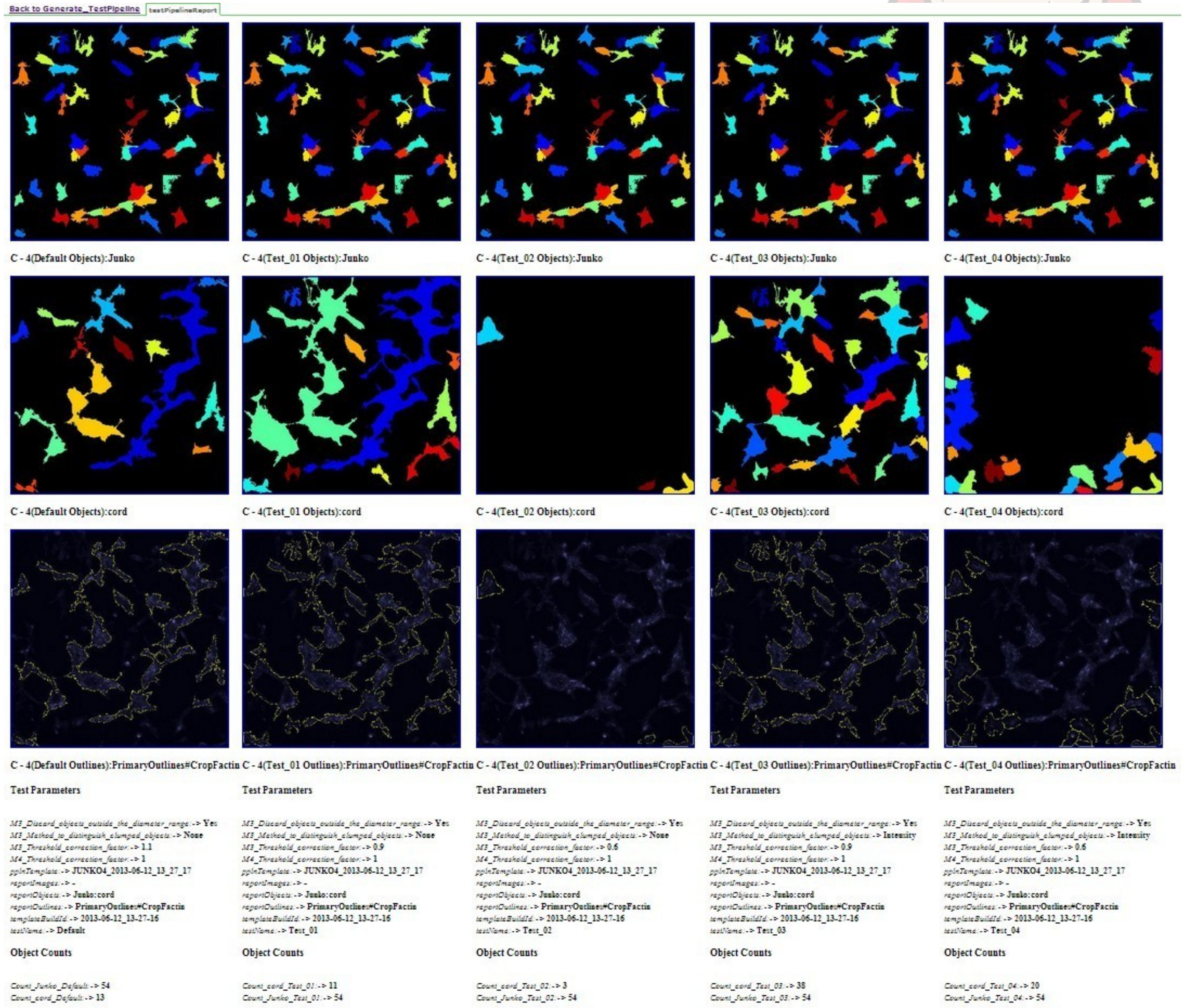
TEST_THUMBNAIL _GENERATION_T4Check THUMBNAILS T5_PlateView
Feb 7, 2014 5:14:38 PM
11 sec

Well ID: G15
Channel: Cy5-Cy5
Field: 1

Advanced/Experimental Functionality

Exploring the parameter space (a.k.a. Test Mosaic)

- Optimization of imaging module parameters
  - A typical pipeline development requirement
- Test Mosaic
  - Allows systematic and documented exploration of the parameter space
  - Evaluation is based on visual and quantitative interpretation of the results



C - 4(Default Objects):Junko

C - 4(Test_01 Objects):Junko

C - 4(Test_02 Objects):Junko

C - 4(Test_03 Objects):Junko

C - 4(Test_04 Objects):Junko

C - 4(Default Objects):cord

C - 4(Test_01 Objects):cord

C - 4(Test_02 Objects):cord

C - 4(Test_03 Objects):cord

C - 4(Test_04 Objects):cord

C - 4(Default Outlines):PrimaryOutlines#CropFactin C - 4(Test_01 Outlines):PrimaryOutlines#CropFactin C - 4(Test_02 Outlines):PrimaryOutlines#CropFactin C - 4(Test_03 Outlines):PrimaryOutlines#CropFactin C - 4(Test_04 Outlines):PrimaryOutlines#CropFactin

Test Parameters

M3_Discard_objects_outside_the_diameter_range: -> Yes
M3_Method_to_distinguish_clumped_objects: -> None
M3_Threshold_correction_factor: -> 1.1
M4_Threshold_correction_factor: -> 1
pplnTemplate: -> JUNKO4_2013-06-12_13_27_17
reportImages: -> -
reportObjects: -> Junko:cord
reportOutlines: -> PrimaryOutlines#CropFactin
templateBuildId: -> 2013-06-12_13-27-16
testName: -> Default

Object Counts

Count_Junko_Default: -> 54
Count_cord_Default: -> 13

Test Parameters

M3_Discard_objects_outside_the_diameter_range: -> Yes
M3_Method_to_distinguish_clumped_objects: -> None
M3_Threshold_correction_factor: -> 0.9
M4_Threshold_correction_factor: -> 1
pplnTemplate: -> JUNKO4_2013-06-12_13_27_17
reportImages: -> -
reportObjects: -> Junko:cord
reportOutlines: -> PrimaryOutlines#CropFactin
templateBuildId: -> 2013-06-12_13-27-16
testName: -> Test_01

Object Counts

Count_cord_Test_01: -> 11
Count_Junko_Test_01: -> 54

Test Parameters

M3_Discard_objects_outside_the_diameter_range: -> Yes
M3_Method_to_distinguish_clumped_objects: -> None
M3_Threshold_correction_factor: -> 0.6
M4_Threshold_correction_factor: -> 1
pplnTemplate: -> JUNKO4_2013-06-12_13_27_17
reportImages: -> -
reportObjects: -> Junko:cord
reportOutlines: -> PrimaryOutlines#CropFactin
templateBuildId: -> 2013-06-12_13-27-16
testName: -> Test_02

Object Counts

Count_cord_Test_02: -> 3
Count_Junko_Test_02: -> 54

Test Parameters

M3_Discard_objects_outside_the_diameter_range: -> Yes
M3_Method_to_distinguish_clumped_objects: -> Intensity
M3_Threshold_correction_factor: -> 0.9
M4_Threshold_correction_factor: -> 1
pplnTemplate: -> JUNKO4_2013-06-12_13_27_17
reportImages: -> -
reportObjects: -> Junko:cord
reportOutlines: -> PrimaryOutlines#CropFactin
templateBuildId: -> 2013-06-12_13-27-16
testName: -> Test_03

Object Counts

Count_cord_Test_03: -> 38
Count_Junko_Test_03: -> 54

Test Parameters

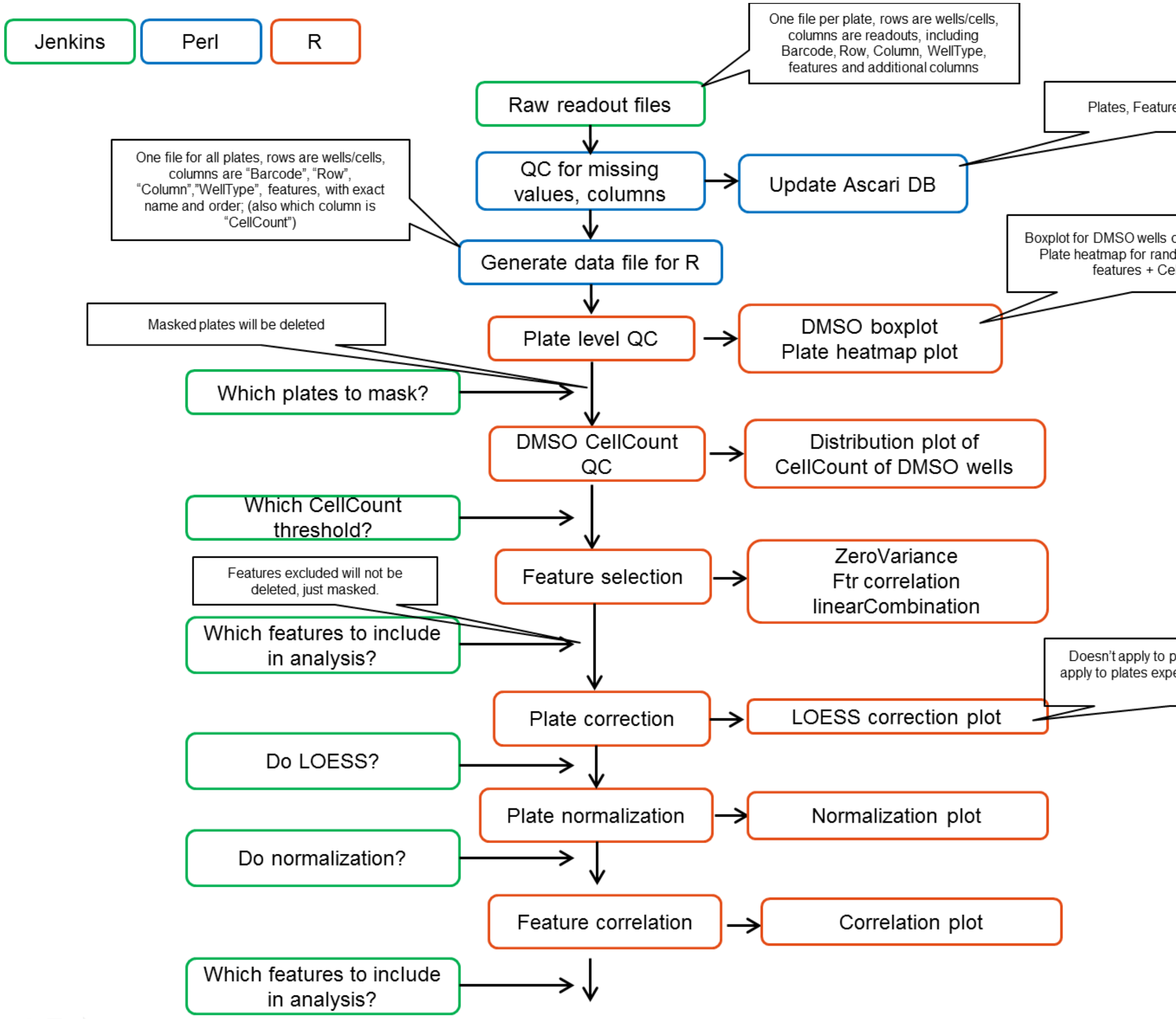
M3_Discard_objects_outside_the_diameter_range: -> Yes
M3_Method_to_distinguish_clumped_objects: -> Intensity
M3_Threshold_correction_factor: -> 0.6
M4_Threshold_correction_factor: -> 1
pplnTemplate: -> JUNKO4_2013-06-12_13_27_17
reportImages: -> -
reportObjects: -> Junko:cord
reportOutlines: -> PrimaryOutlines#CropFactin
templateBuildId: -> 2013-06-12_13-27-16
testName: -> Test_04

Object Counts

Count_cord_Test_04: -> 20
Count_Junko_Test_04: -> 54

HCS-Multi-Parametric Data Analysis

Current Focus: Prototype powerful and easy to use analytics



Jenkins

Perl

R

Raw readout files

One file per plate, rows are wells/cells, columns are readouts, including Barcode, Row, Column, WellType, features and additional columns

QC for missing values, columns

Update Ascari DB

Plates, Features, Annotations

One file for all plates, rows are wells/cells, columns are "Barcode", "Row", "Column","WellType", features, with exact name and order; (also which column is "CellCount")

Generate data file for R

Plate level QC

DMSO boxplot
Plate heatmap plot

Boxplot for DMSO wells only for all features. Plate heatmap for randomly selected N features + CellCount.

Masked plates will be deleted

Which plates to mask?

DMSO CellCount QC

Distribution plot of CellCount of DMSO wells

Which CellCount threshold?

Feature selection

ZeroVariance
Ftr correlation
linearCombination

Features excluded will not be deleted, just masked.

Which features to include in analysis?

Plate correction

LOESS correction plot

Doesn't apply to positive controls; doesn't apply to plates expected to have pattern (eg. DRC)

Do LOESS?

Plate normalization

Normalization plot

Do normalization?

Feature correlation

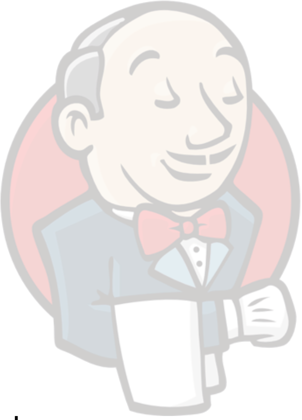
Correlation plot

Which features to include in analysis?

Contributed by Stanley Lazic

Statistics, Visualization, Reporting

My current Jenkins toolkit

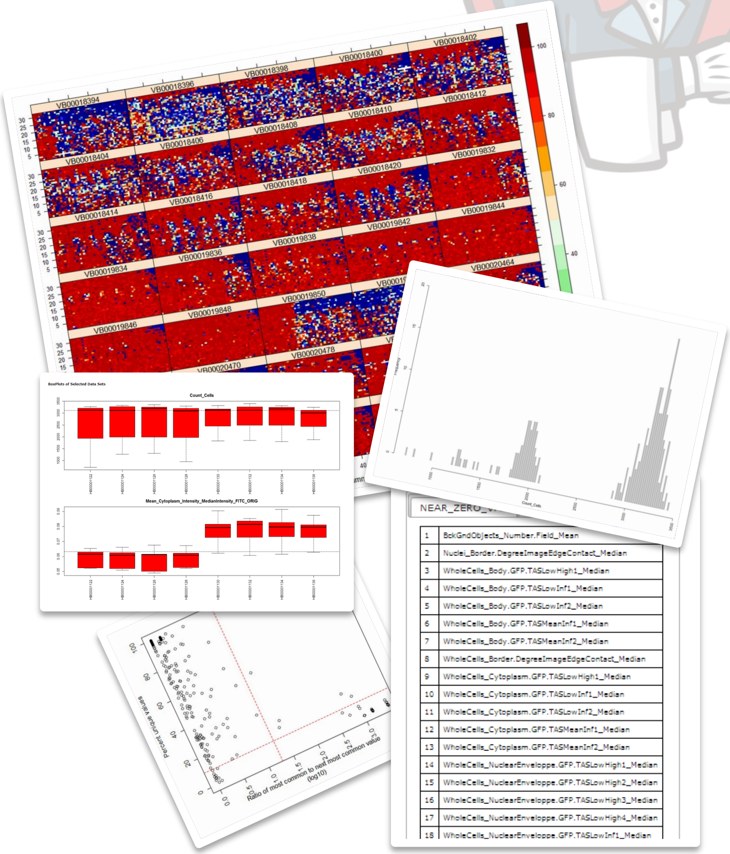


- Jenkins R-Plugin
  - Supplies build step for executing R scripts
    - This plug-in was created by the BioUno project (sponsored by TupiLabs), and released to Jenkins as well.
- Image Gallery Plugin
  - This plug-in reads a job workspace and collects images to produce an image gallery
  - Useful for visualizing various statistical plots and graphs
    - This plug-in was created by the BioUno project (sponsored by TupiLabs), and released to Jenkins as well.
- Reporting Plugins
  - HTML Publisher, Summary Display

Jenkins for Interactive Analytics

Using R in a Jenkins pipeline interactively

- Opportunities
  - Quickly prototype functional analysis for multi-parametric data
    - Improve analysis requirements
    - Experiment with required data management and analysis workflows
  - Provide lab scientists with an easy to use, yet sophisticated, standardized and validated platform for MP data analysis tools



NEAR_ZERO_

1	BckGndObjects_Number.Field_Mean
2	Nuclei_Border.DegreeImageEdgeContact_Median
3	WholeCells_Body.GFP.TASLowHigh1_Median
4	WholeCells_Body.GFP.TASLowInf1_Median
5	WholeCells_Body.GFP.TASLowInf2_Median
6	WholeCells_Body.GFP.TASMeanInf1_Median
7	WholeCells_Body.GFP.TASMeanInf2_Median
8	WholeCells_Border.DegreeImageEdgeContact_Median
9	WholeCells_Cytoplasm.GFP.TASLowHigh1_Median
10	WholeCells_Cytoplasm.GFP.TASLowInf1_Median
11	WholeCells_Cytoplasm.GFP.TASLowInf2_Median
12	WholeCells_Cytoplasm.GFP.TASMeanInf1_Median
13	WholeCells_Cytoplasm.GFP.TASMeanInf2_Median
14	WholeCells_NuclearEnveloppe.GFP.TASLowHigh1_Median
15	WholeCells_NuclearEnveloppe.GFP.TASLowHigh2_Median
16	WholeCells_NuclearEnveloppe.GFP.TASLowHigh3_Median
17	WholeCells_NuclearEnveloppe.GFP.TASLowHigh4_Median
18	WholeCells_NuclearEnveloppe.GFP.TASLowInf1_Median

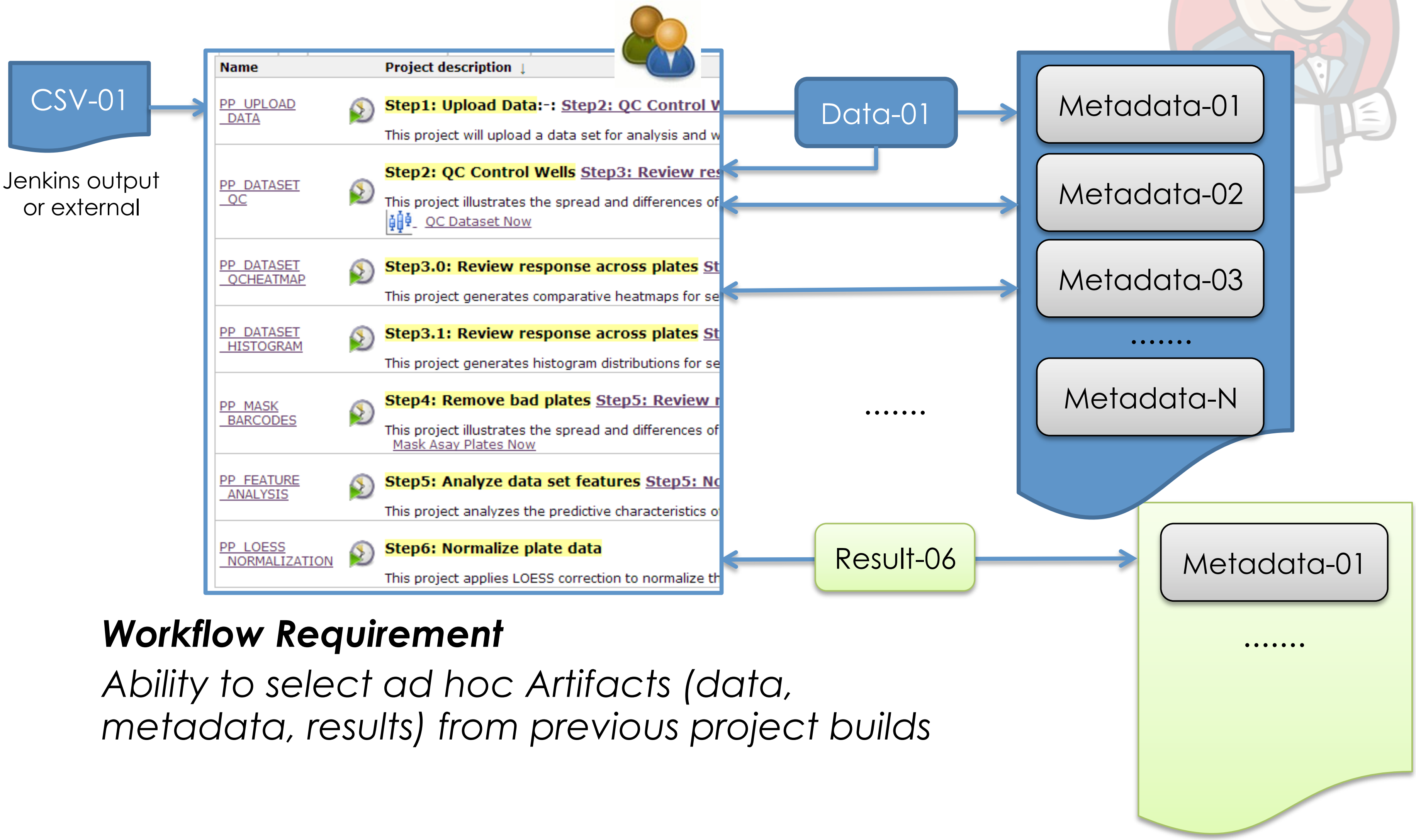
Jenkins for Interactive Analytics

Using R in a Jenkins pipeline interactively

- Challenges
  - Limitations of the Jenkins user interface
    - Limited interaction between UI controls
  - Large and varied HC measurement metadata
    - A challenge for creating HC data schemata as well
- Strategies
    - Open source collaboration with BioUno project
      - Uno-choice UI control greatly facilitates dynamic updating of the UI
    - Initial design supports flexible (but still controlled) data schema
      - Low tech, cumulative, shared key-value Java properties

Analytical Builds

A build may create a new transform of the data or simply add metadata



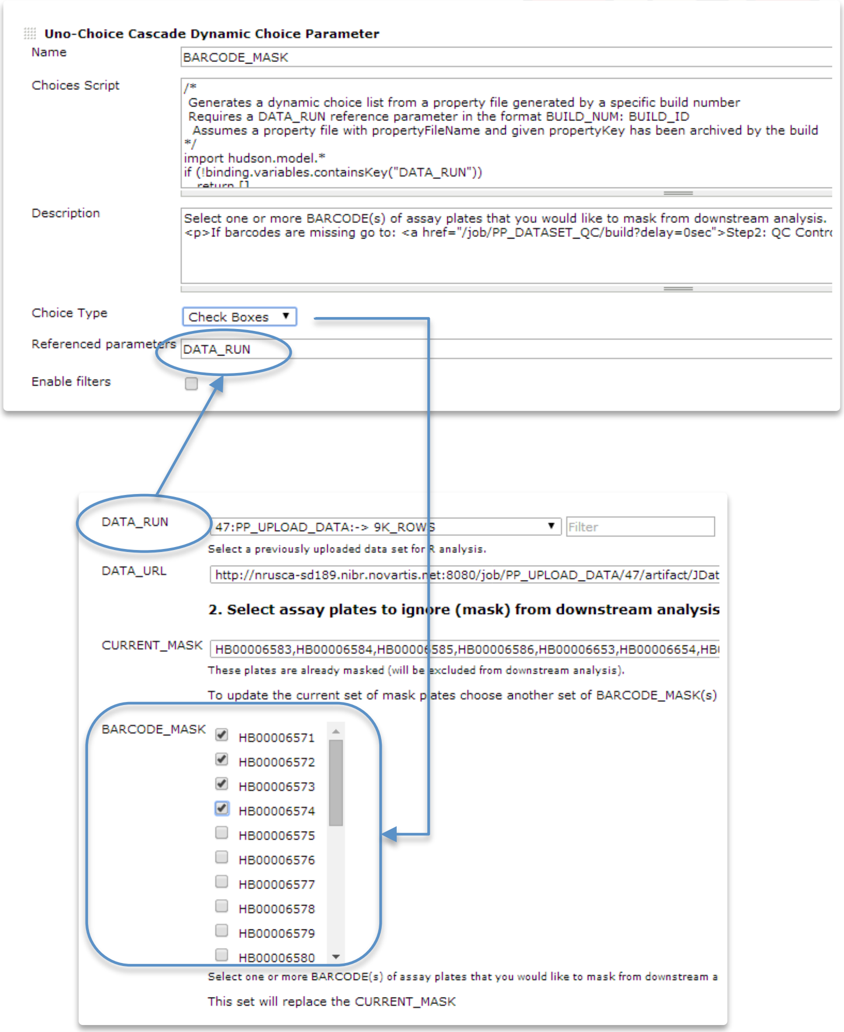
Name	Project description ↓
PP_UPLOAD_DATA	**Step1: Upload Data:-: Step2: QC Control W** This project will upload a data set for analysis and w
PP_DATASET_QC	**Step2: QC Control Wells Step3: Review res** This project illustrates the spread and differences of QC Dataset Now
PP_DATASET_QCHEATMAP	**Step3.0: Review response across plates St** This project generates comparative heatmaps for se
PP_DATASET_HISTOGRAM	**Step3.1: Review response across plates St** This project generates histogram distributions for se
PP_MASK_BARCODES	**Step4: Remove bad plates Step5: Review r** This project illustrates the spread and differences of Mask Asay Plates Now
PP_FEATURE_ANALYSIS	**Step5: Analyze data set features Step5: No** This project analyzes the predictive characteristics o
PP_LOESS_NORMALIZATION	**Step6: Normalize plate data** This project applies LOESS correction to normalize th

Workflow Requirement

Ability to select ad hoc Artifacts (data, metadata, results) from previous project builds

The Uno-Choice plugin

- Provides a list of dynamically generated options
  - Driven by a Groovy script
  - Single/Multi-select (Check Boxes, Radio Buttons)
  - References one or more other UI parameters
  - Dynamically refreshes when referenced UI parameters change



Uno-Choice Cascade Dynamic Choice Parameter

Name: BARCODE_MASK

Choices Script:

```
/*
 Generates a dynamic choice list from a property file generated by a specific build number
 Requires a DATA_RUN reference parameter in the format BUILD_NUM: BUILD_ID
  Assumes a property file with propertyFileName and given propertyKey has been archived by the build
*/
import hudson.model.*
if (!binding.variables.containsKey("DATA_RUN"))
  return []
```

Description:

```
Select one or more BARCODE(s) of assay plates that you would like to mask from downstream analysis.
<p>If barcodes are missing go to: <a href="/job/PP_DATASET_QC/build?delay=0sec">Step2: QC Contr
```

Choice Type: Check Boxes ▼

Referenced parameters: DATA_RUN

Enable filters: □

DATA_RUN: 47:PP_UPLOAD_DATA:-> 9K_ROWS ▼ Filter

Select a previously uploaded data set for R analysis.

DATA_URL: http://nrusca-sd189.nibr.novartis.net:8080/job/PP_UPLOAD_DATA/47/artifact/JDat

2. Select assay plates to ignore (mask) from downstream analysis

CURRENT_MASK: HB00006583,HB00006584,HB00006585,HB00006586,HB00006653,HB00006654,HB

These plates are already masked (will be excluded from downstream analysis).

To update the current set of mask plates choose another set of BARCODE_MASK(s)

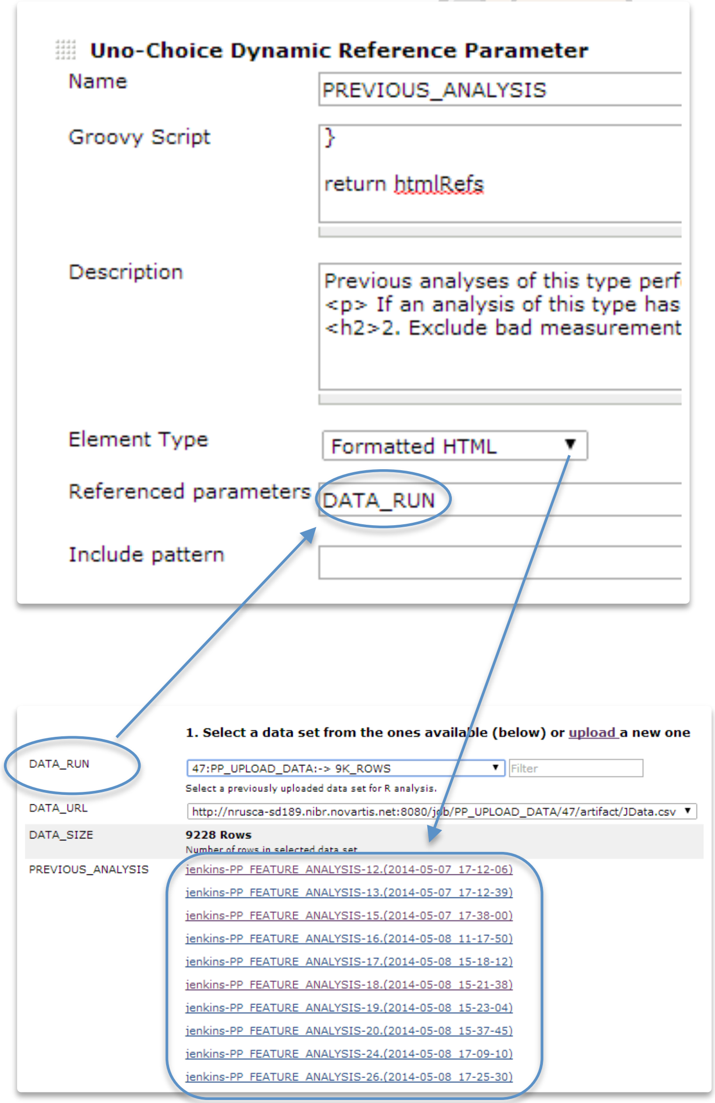
BARCODE_MASK:
☑ HB00006571
☑ HB00006572
☑ HB00006573
☑ HB00006574
□ HB00006575
□ HB00006576
□ HB00006577
□ HB00006578
□ HB00006579
□ HB00006580

Select one or more BARCODE(s) of assay plates that you would like to mask from downstream a

This set will replace the CURRENT_MASK

The Uno-Choice plugin

- Provides reference parameters
  - Dynamically rendered in the UI but not used in the build
  - Rendered as lists, ‘free-form’ HTML, or an image gallery



Uno-Choice Dynamic Reference Parameter

Name	PREVIOUS_ANALYSIS

Groovy Script	}
	return htmlRefs

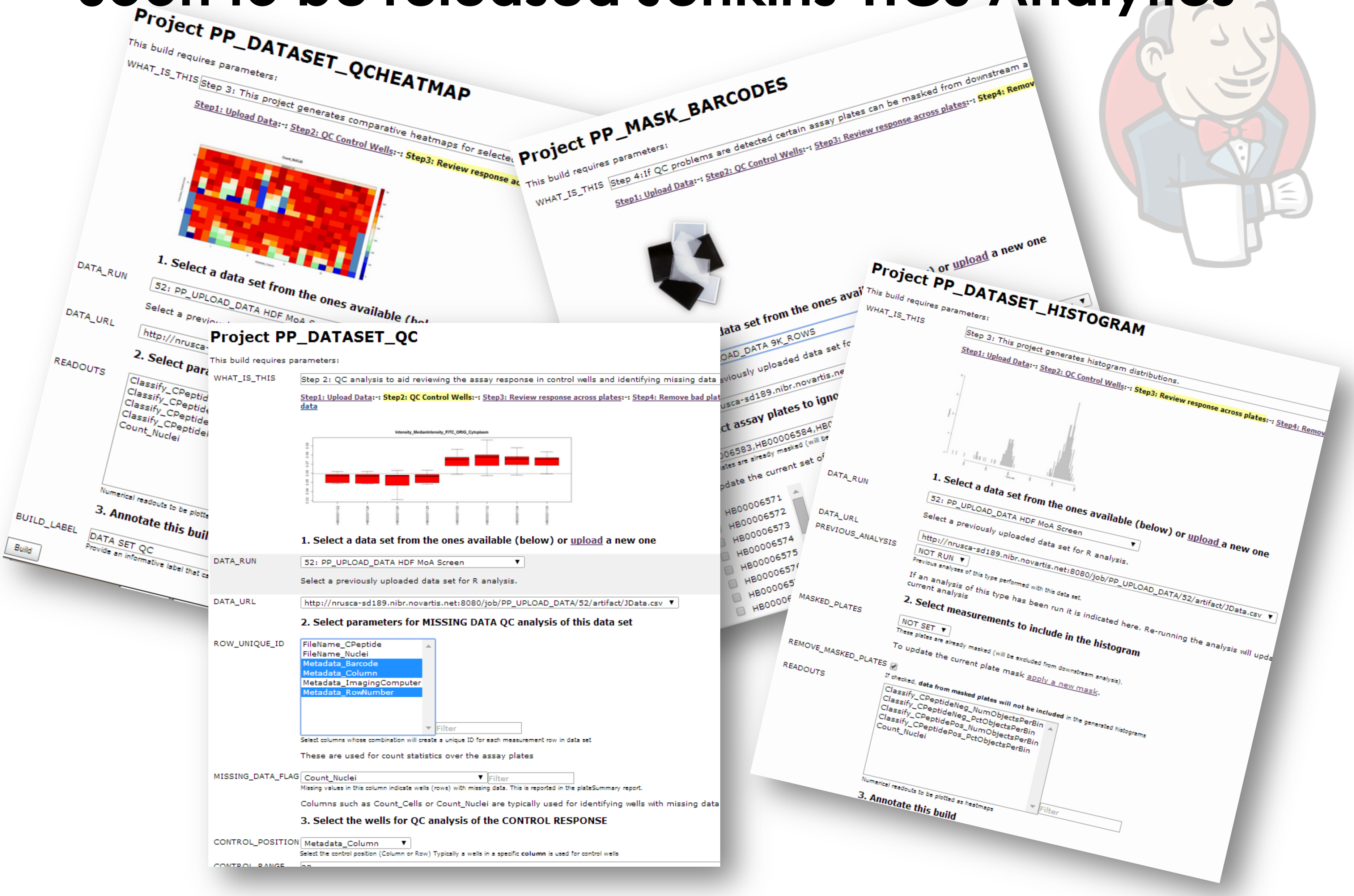
Description	Previous analyses of this type perf <p> If an analysis of this type has <h2>2. Exclude bad measurement
Element Type	Formatted HTML ▼
Referenced parameters	DATA_RUN
Include pattern	

1. Select a data set from the ones available (below) or upload a new one

DATA_RUN	47:PP_UPLOAD_DATA:-> 9K_ROWS ▼ Filter Select a previously uploaded data set for R analysis.
DATA_URL	http://nrusca-sd189.nibr.novartis.net:8080/job/PP_UPLOAD_DATA/47/artifact/JData.csv ▼
DATA_SIZE	**9228 Rows** Number of rows in selected data set

PREVIOUS_ANALYSIS	jenkins-PP_FEATURE_ANALYSIS-12.(2014-05-07_17-12-06)
	jenkins-PP_FEATURE_ANALYSIS-13.(2014-05-07_17-12-39)
	jenkins-PP_FEATURE_ANALYSIS-15.(2014-05-07_17-38-00)
	jenkins-PP_FEATURE_ANALYSIS-16.(2014-05-08_11-17-50)
	jenkins-PP_FEATURE_ANALYSIS-17.(2014-05-08_15-18-12)
	jenkins-PP_FEATURE_ANALYSIS-18.(2014-05-08_15-21-38)
	jenkins-PP_FEATURE_ANALYSIS-19.(2014-05-08_15-23-04)
	jenkins-PP_FEATURE_ANALYSIS-20.(2014-05-08_15-37-45)
	jenkins-PP_FEATURE_ANALYSIS-24.(2014-05-08_17-09-10)
	jenkins-PP_FEATURE_ANALYSIS-26.(2014-05-08_17-25-30)

Soon to be released Jenkins-HCS Analytics



Project PP_DATASET_QCHEATMAP

This build requires parameters:

WHAT_IS_THIS Step 3: This project generates comparative heatmaps for selecte

Step1: Upload Data:-: Step2: QC Control Wells:-: Step3: Review response ac

1. Select a data set from the ones available (bel

DATA_RUN 52: PP_UPLOAD_DATA HDF MoA S

Select a previou

DATA_URL http://nrusca-

2. Select para

READOUTS Classify_CPeptid
Classify_CPeptid
Classify_CPeptide
Classify_CPeptide
Count_Nuclei

Numerical readouts to be plotte

3. Annotate this buil

BUILD_LABEL DATA SET QC

Provide an informative label that ca

Build

Project PP_MASK_BARCODES

This build requires parameters:

WHAT_IS_THIS Step 4:If QC problems are detected certain assay plates can be masked from downstream a

Step1: Upload Data:-: Step2: QC Control Wells:-: Step3: Review response across plates:-: Step4: Remov

) or upload a new one

data set from the ones avai

OAD_DATA 9K_ROWS

eviously uploaded data set fo

usca-sd189.nibr.novartis.ne

ct assay plates to igno

06583,HB00006584,HB0

ates are already masked (will be

date the current set of

HB00006571
HB00006572
HB00006573
HB00006574
HB00006575
HB0000657
HB000065
HB00006

Project PP_DATASET_QC

This build requires parameters:

WHAT_IS_THIS Step 2: QC analysis to aid reviewing the assay response in control wells and identifying missing data

Step1: Upload Data:-: **Step2: QC Control Wells**:-: Step3: Review response across plates:-: Step4: Remove bad plat data

1. Select a data set from the ones available (below) or upload a new one

DATA_RUN 52: PP_UPLOAD_DATA HDF MoA Screen ▼

Select a previously uploaded data set for R analysis.

DATA_URL http://nrusca-sd189.nibr.novartis.net:8080/job/PP_UPLOAD_DATA/52/artifact/JData.csv ▼

2. Select parameters for MISSING DATA QC analysis of this data set

ROW_UNIQUE_ID
FileName_CPeptide
FileName_Nuclei
Metadata_Barcode
Metadata_Column
Metadata_ImagingComputer
Metadata_RowNumber

Filter

Select columns whose combination will create a unique ID for each measurement row in data set

These are used for count statistics over the assay plates

MISSING_DATA_FLAG Count_Nuclei ▼ Filter

Missing values in this column indicate wells (rows) with missing data. This is reported in the plateSummary report.

Columns such as Count_Cells or Count_Nuclei are typically used for identifying wells with missing data

3. Select the wells for QC analysis of the CONTROL RESPONSE

CONTROL_POSITION Metadata_Column ▼

Select the control position (Column or Row) Typically a wells in a specific **column** is used for control wells

CONTROL_RANGE

Project PP_DATASET_HISTOGRAM

This build requires parameters:

WHAT_IS_THIS Step 3: This project generates histogram distributions.

Step1: Upload Data:-: Step2: QC Control Wells:-: **Step3: Review response across plates**:-: Step4: Remov

1. Select a data set from the ones available (below) or upload a new one

DATA_RUN 52: PP_UPLOAD_DATA HDF MoA Screen ▼

Select a previously uploaded data set for R analysis.

DATA_URL http://nrusca-sd189.nibr.novartis.net:8080/job/PP_UPLOAD_DATA/52/artifact/JData.csv ▼

PREVIOUS_ANALYSIS NOT RUN ▼

Previous analyses of this type performed with this data set.

If an analysis of this type has been run it is indicated here. Re-running the analysis will upda current analysis

2. Select measurements to include in the histogram

MASKED_PLATES NOT SET ▼

These plates are already masked (will be excluded from downstream analysis).

To update the current plate mask apply a new mask.

REMOVE_MASKED_PLATES ☑

If checked, **data from masked plates will not be included** in the generated histograms

READOUTS
Classify_CPeptideNeg_NumObjectsPerBin
Classify_CPeptideNeg_PctObjectsPerBin
Classify_CPeptidePos_NumObjectsPerBin
Classify_CPeptidePos_PctObjectsPerBin
Count_Nuclei

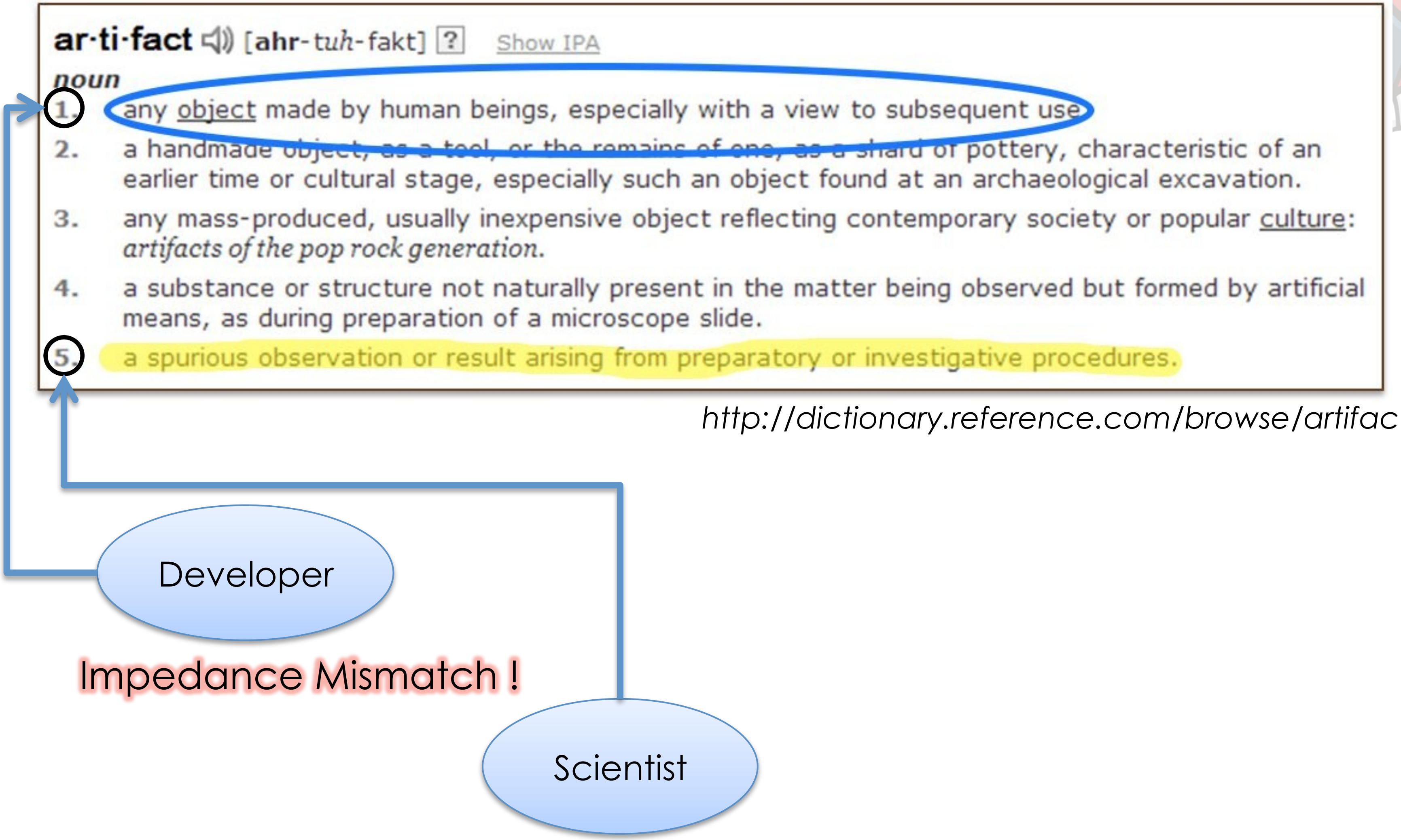
Filter

Numerical readouts to be plotted as heatmaps

3. Annotate this build

Introducing Jenkins to Life Sciences!

Let's start by explaining away 'artifacts'!



ar·ti·fact [ahr-tuh-fakt] ? Show IPA

noun

1. any object made by human beings, especially with a view to subsequent use.
2. a handmade object, as a tool, or the remains of one, as a shard of pottery, characteristic of an earlier time or cultural stage, especially such an object found at an archaeological excavation.
3. any mass-produced, usually inexpensive object reflecting contemporary society or popular culture: *artifacts of the pop rock generation.*
4. a substance or structure not naturally present in the matter being observed but formed by artificial means, as during preparation of a microscope slide.
5. a spurious observation or result arising from preparatory or investigative procedures.

http://dictionary.reference.com/browse/artifact

Developer

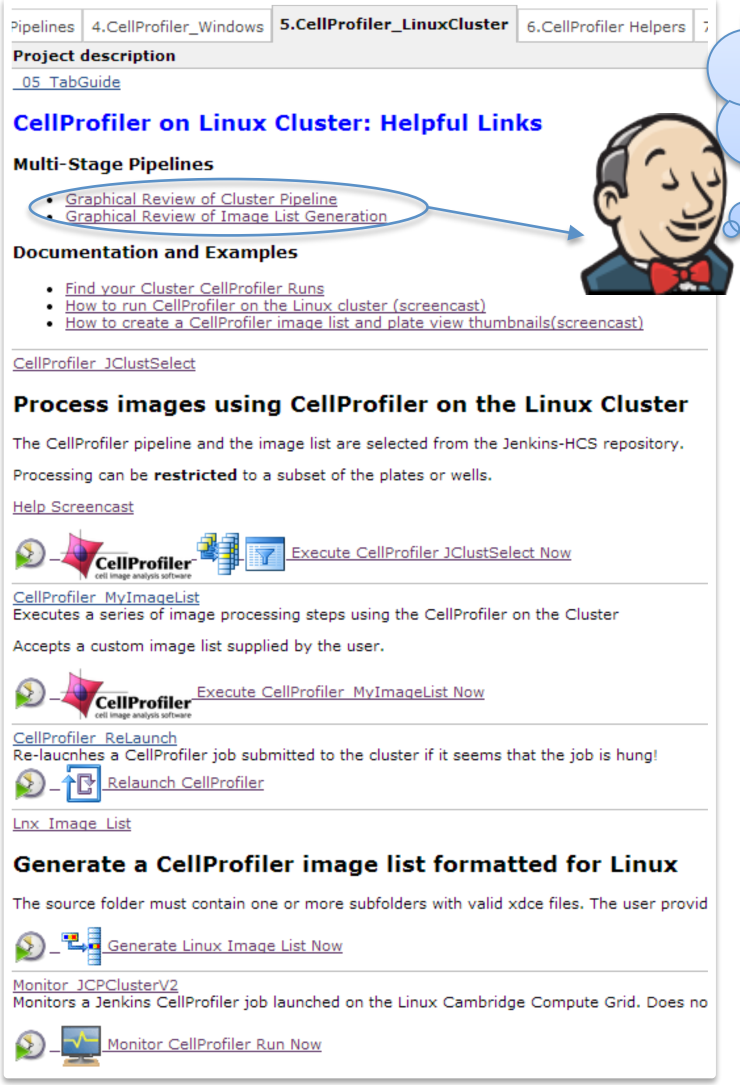
Impedance Mismatch !

Scientist

Introducing Jenkins to Life Sciences

Let's improve the User Interface/Experience

- Let's start by improving the default Jenkins UI
  - Layout
  - Navigation
  - Refreshing
  - Interactivity
- This is an active Jenkins community discussion



Pipelines | 4.CellProfiler_Windows | 5.CellProfiler_LinuxCluster | 6.CellProfiler Helpers | 7

Project description

_05_TabGuide

CellProfiler on Linux Cluster: Helpful Links

Multi-Stage Pipelines

- Graphical Review of Cluster Pipeline
- Graphical Review of Image List Generation

Documentation and Examples

- Find your Cluster CellProfiler Runs
- How to run CellProfiler on the Linux cluster (screencast)
- How to create a CellProfiler image list and plate view thumbnails(screencast)

CellProfiler_JClustSelect

Process images using CellProfiler on the Linux Cluster

The CellProfiler pipeline and the image list are selected from the Jenkins-HCS repository.

Processing can be **restricted** to a subset of the plates or wells.

Help Screencast

Execute CellProfiler JClustSelect Now

CellProfiler_MyImageList

Executes a series of image processing steps using the CellProfiler on the Cluster

Accepts a custom image list supplied by the user.

Execute CellProfiler_MyImageList Now

CellProfiler_ReLaunch

Re-laucnhes a CellProfiler job submitted to the cluster if it seems that the job is hung!

Relaunch CellProfiler

Lnx_Image_List

Generate a CellProfiler image list formatted for Linux

The source folder must contain one or more subfolders with valid xdce files. The user provid

Generate Linux Image List Now

Monitor_JCPClusterV2

Monitors a Jenkins CellProfiler job launched on the Linux Cambridge Compute Grid. Does no

Monitor CellProfiler Run Now

Hyperlinks to Build Pipeline Views of 'Jenkins Helper' user

What We are Missing

Configuration Explorer

- Structured
- Graphical
- Dynamic

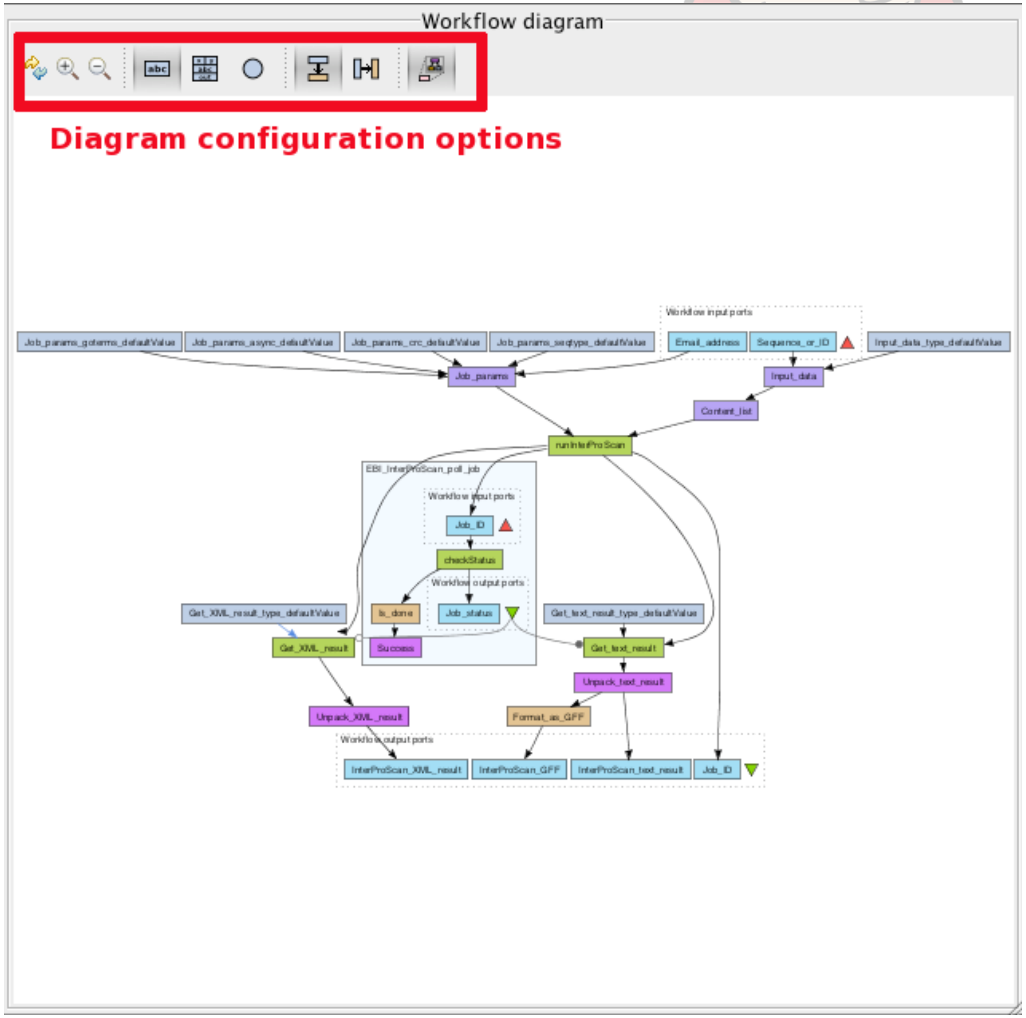
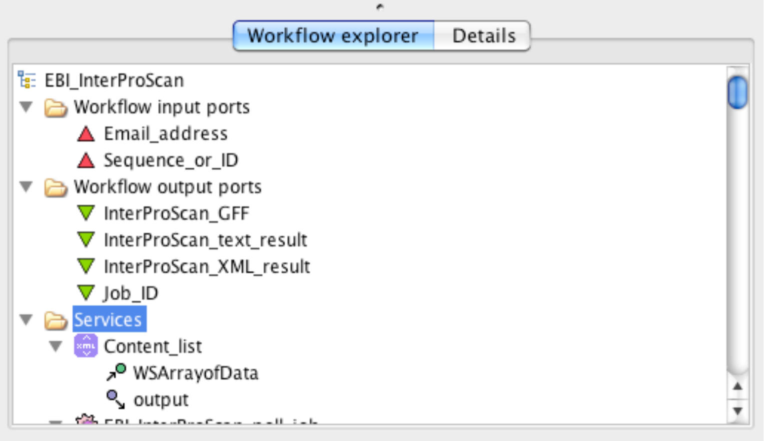


Diagram configuration options



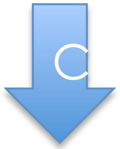
- EBI_InterProScan
  - Workflow input ports
    - Email_address
    - Sequence_or_ID
  - Workflow output ports
    - InterProScan_GFF
    - InterProScan_text_result
    - InterProScan_XML_result
    - Job_ID
  - Services
    - Content_list
      - WSArrayofData
      - output
    - EBI_InterProScan_poll_job

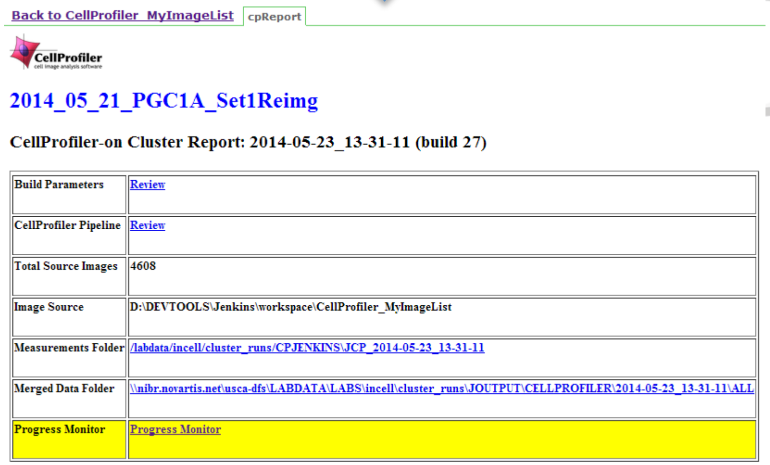
What We are Missing

Bi-Directional Build Interaction

- Build A **uses** artifacts of Build B
  - Limited support by **Run Type** parameter
    - Missing flexible and dynamic filtering
- Build D **modifies** build/ publisher artifacts of Build C
  - Sometimes not do-able
  - Sometimes requires a reload

- Build C produces an intermediate report that will get updated once Build D is finished successfully.





Back to CellProfiler_MyImageList | cpReport

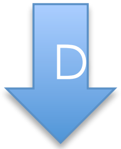
CellProfiler cell image analysis software

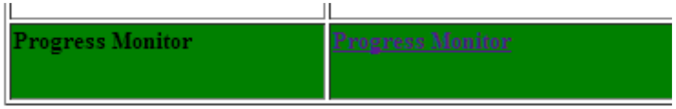
2014_05_21_PGC1A_Set1Reimg

CellProfiler-on Cluster Report: 2014-05-23_13-31-11 (build 27)

Build Parameters	Review
CellProfiler Pipeline	Review
Total Source Images	4608
Image Source	D:\DEVTOOLS\Jenkins\workspace\CellProfiler_MyImageList
Measurements Folder	/labdata/incell/cluster_runs/CPJENKINS\JCP_2014-05-23_13-31-11
Merged Data Folder	\\nibr.novartis.net\usca-dfs\LABDATA\LABS\incell\cluster_runs\JOUTPUT\CELLPROFILER\2014-05-23_13-31-11\ALL
Progress Monitor	Progress Monitor

- Build D monitors output of long running job and updates report of Build C





Progress Monitor	Progress Monitor

‘Progress Monitor’ Link and cell color are updated

What We are Missing

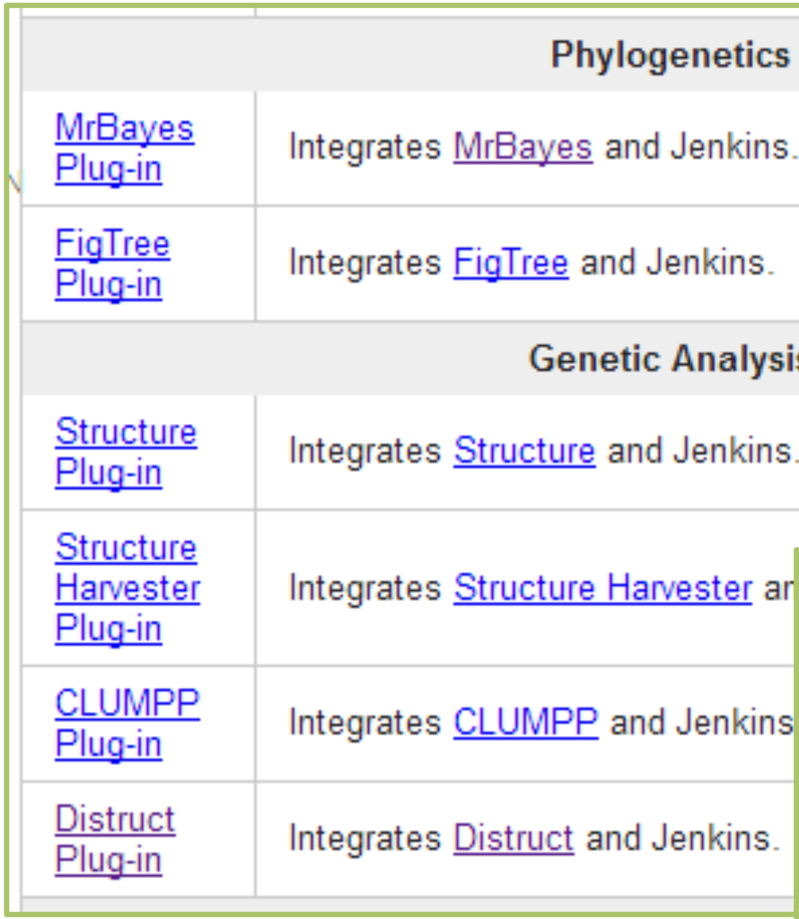
A good, deep search and metadata framework

- Supported
  - View Searches
  - Build Browsing
    - By timeline
    - By view
    - By user

- Missing
  - Build Searching
    - Parameter Search
    - Metadata Search
      - Metadata plugin (currently limited to adding metadata at project level)
    - Artifact Search
  - Tagging
  - Dynamic Metadata

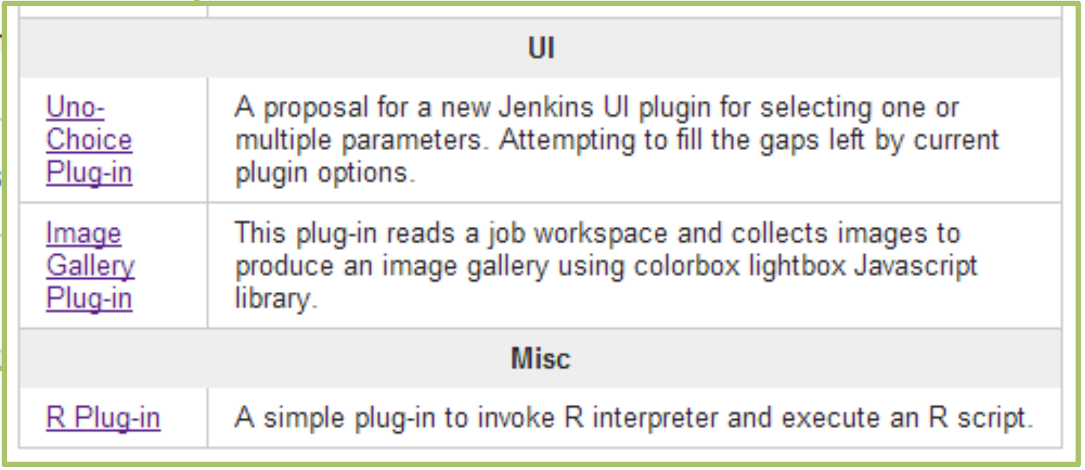
What We are Missing

Life-Sciences Domain Plugins (Bio/Chem Informatic)



	Phylogenetics
MrBayes Plug-in	Integrates MrBayes and Jenkins.
FigTree Plug-in	Integrates FigTree and Jenkins.
	Genetic Analysis
Structure Plug-in	Integrates Structure and Jenkins.
Structure Harvester Plug-in	Integrates Structure Harvester ar
CLUMPP Plug-in	Integrates CLUMPP and Jenkins
Distruct Plug-in	Integrates Distruct and Jenkins.

- The BioUno project is filling the gap
- Interested in plugins that
  - Integrate bio-informatic, statistical and visualization tools
  - Connect to life-science data repositories
  - Generate artifacts and reports in LifeSci formats



	UI
Uno-Choice Plug-in	A proposal for a new Jenkins UI plugin for selecting one or multiple parameters. Attempting to fill the gaps left by current plugin options.
Image Gallery Plug-in	This plug-in reads a job workspace and collects images to produce an image gallery using colorbox lightbox Javascript library.
	Misc
R Plug-in	A simple plug-in to invoke R interpreter and execute an R script.

In Summary

- We have demonstrated that Jenkins-CI can be used for life-science applications
  - Using standard functionality
  - Using domain specific plugins
  - In demanding environments of big data and high performance
- We have observed that scientist are able and willing to use the platform despite it's 'domain impedance mismatch'
- There is some fundamental interest in the larger Jenkins-CI community to expand the boundaries of the framework beyond continuous integration

Where do we want to take Jenkins-CI?

- Discussion
  - No changes?
  - Gradual improvements?
    - User interface
    - API
    - New life-science plugins
  - Fundamental changes?
  - Integration framework for orchestrating more granular pipelines?
    - CellProfiler
    - Galaxy
    - Knime
    - Others?

Acknowledgments

- Novartis
  - Fred Harbinski
  - Christian Parker
  - Stanley Lazic
  - Imtiaz Hossain
  - Josh Snyder
  - Erik Sassaman
- BioUno
  - Bruno Kinoshita
- The Jenkins Community

Thank You To Our Sponsors

Platinum

CloudBees

Gold

black diamond SOFTWARE

MV MidVision™ Release the innovation

choose digital

JFrog

XebiaLabs the deployment automation company

Silver

SOASTA Test Faster. Release Sooner.

ELIASSEN GROUP Agile Consulting Services

KENSHO

SERENA™

assembla