Optide-Hunter

Mi-Youn Brusniak, Ph.D.

2016, Oct 7

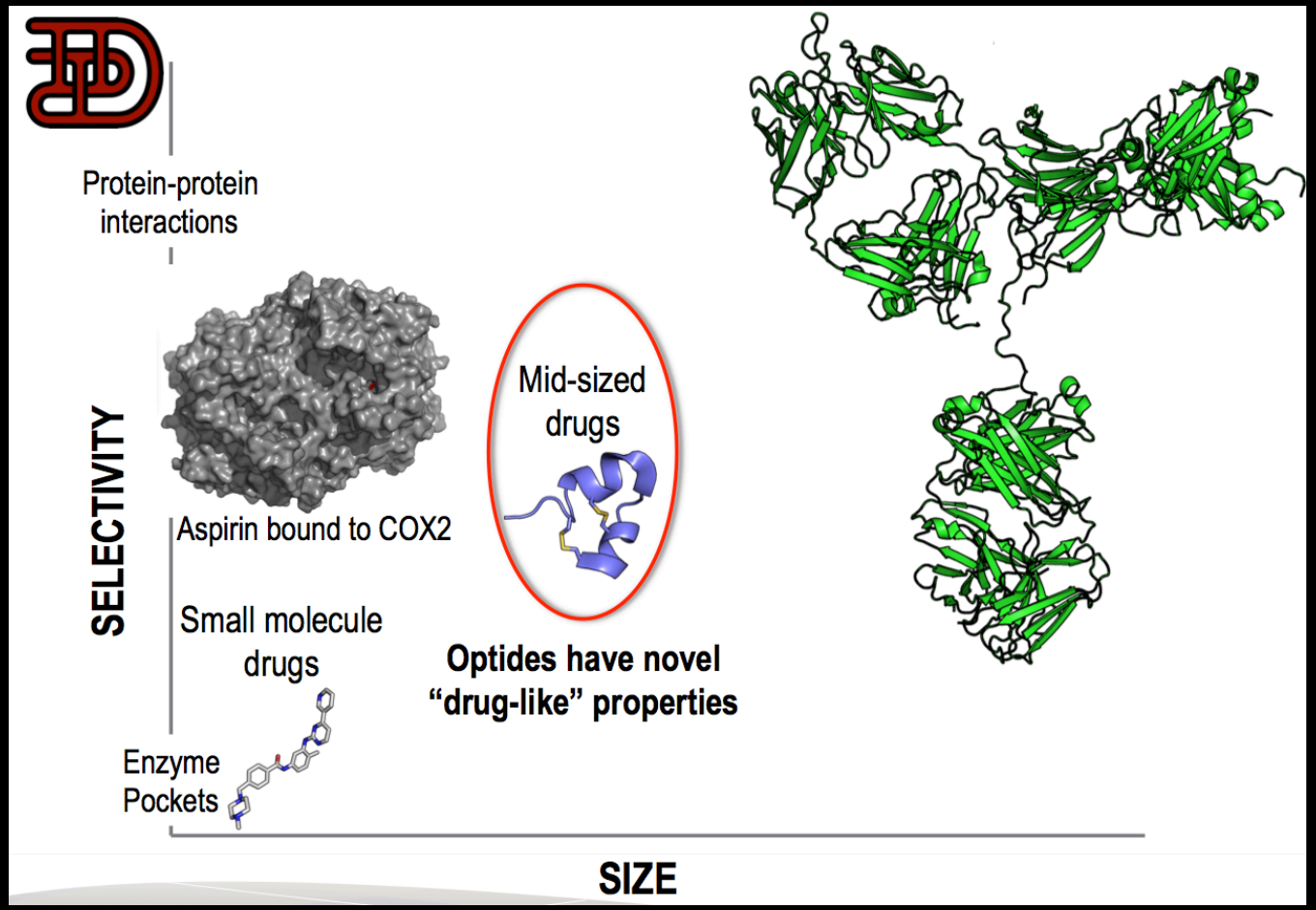
Introduction to Optide Program

Project Violet

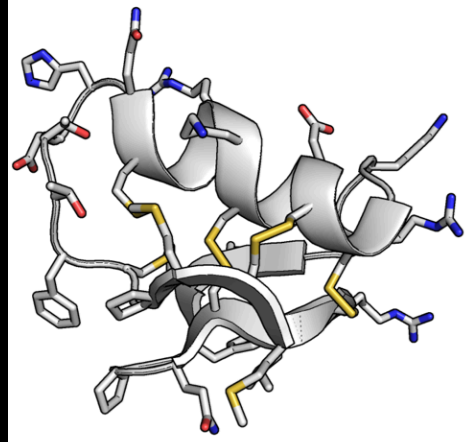
http://www.fredhutch.org/en/labs/clinical/projects/project-violet.html

To fully realize the potential of the optide platform to maximize the benefit to human health.

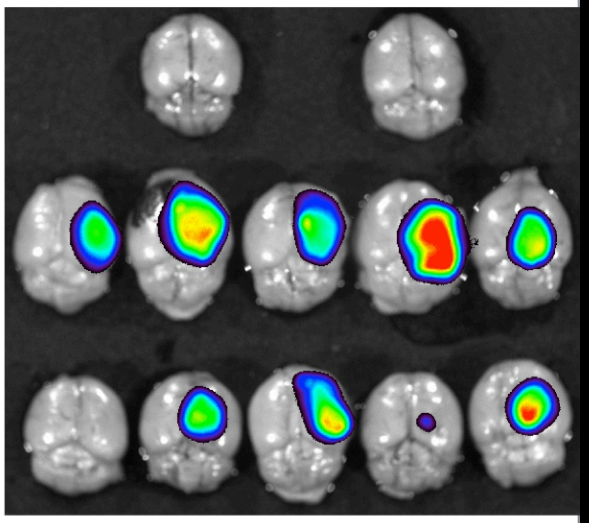
Optimized Peptides (Optides)



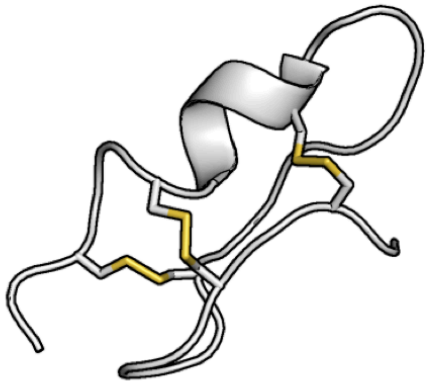
Natural Knottin Biodistribution



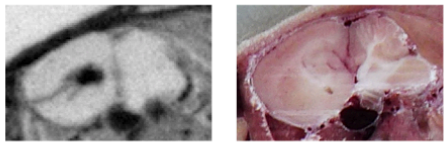
Chlorotoxin



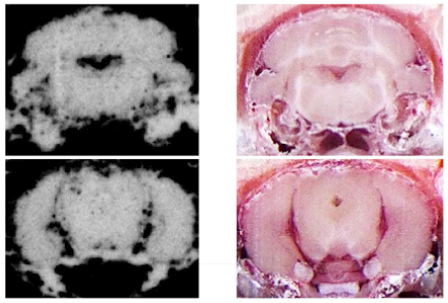
Scorpion Derived Ion Channel Blocker

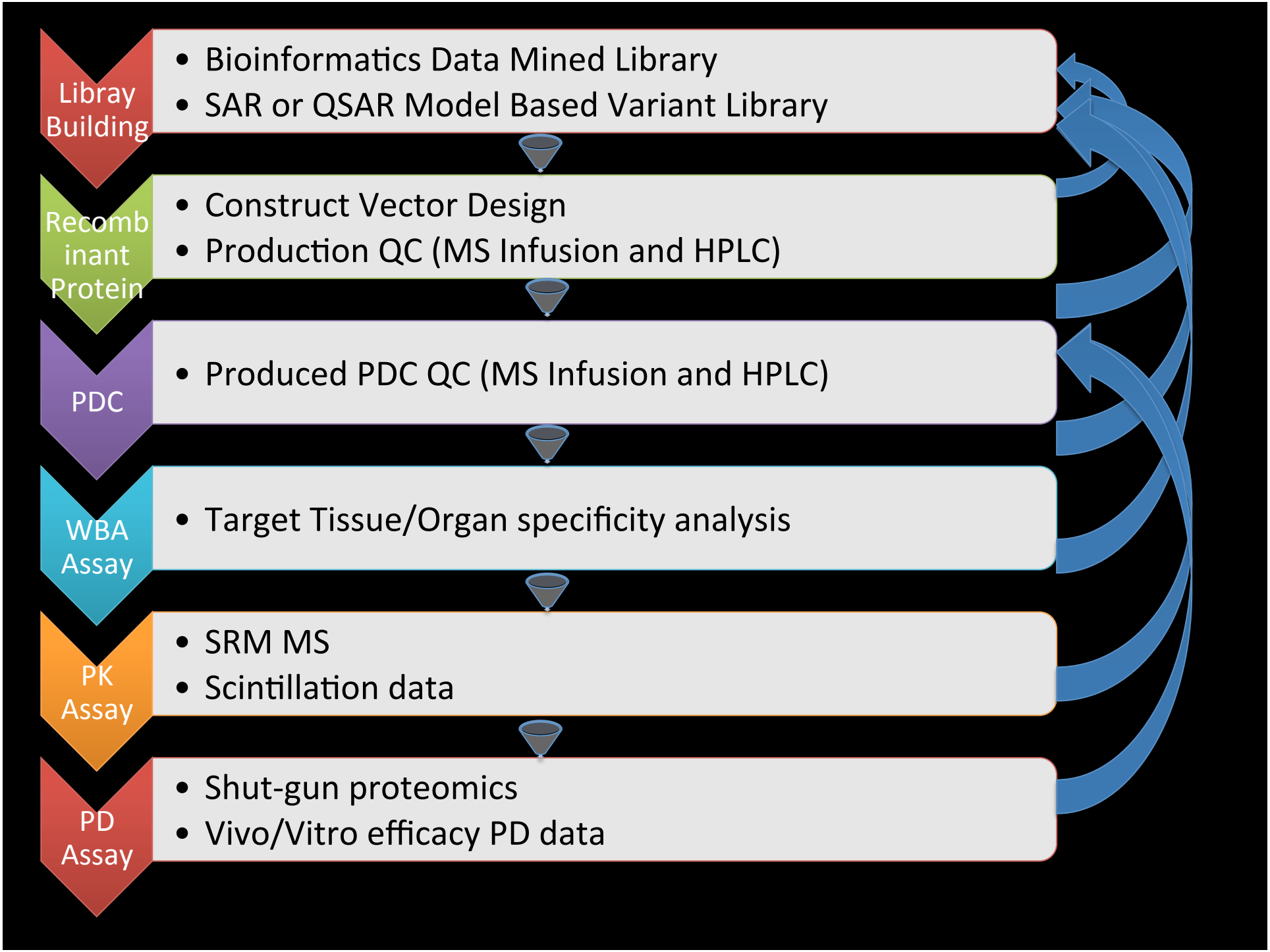


Sagittal Brain Sectioning

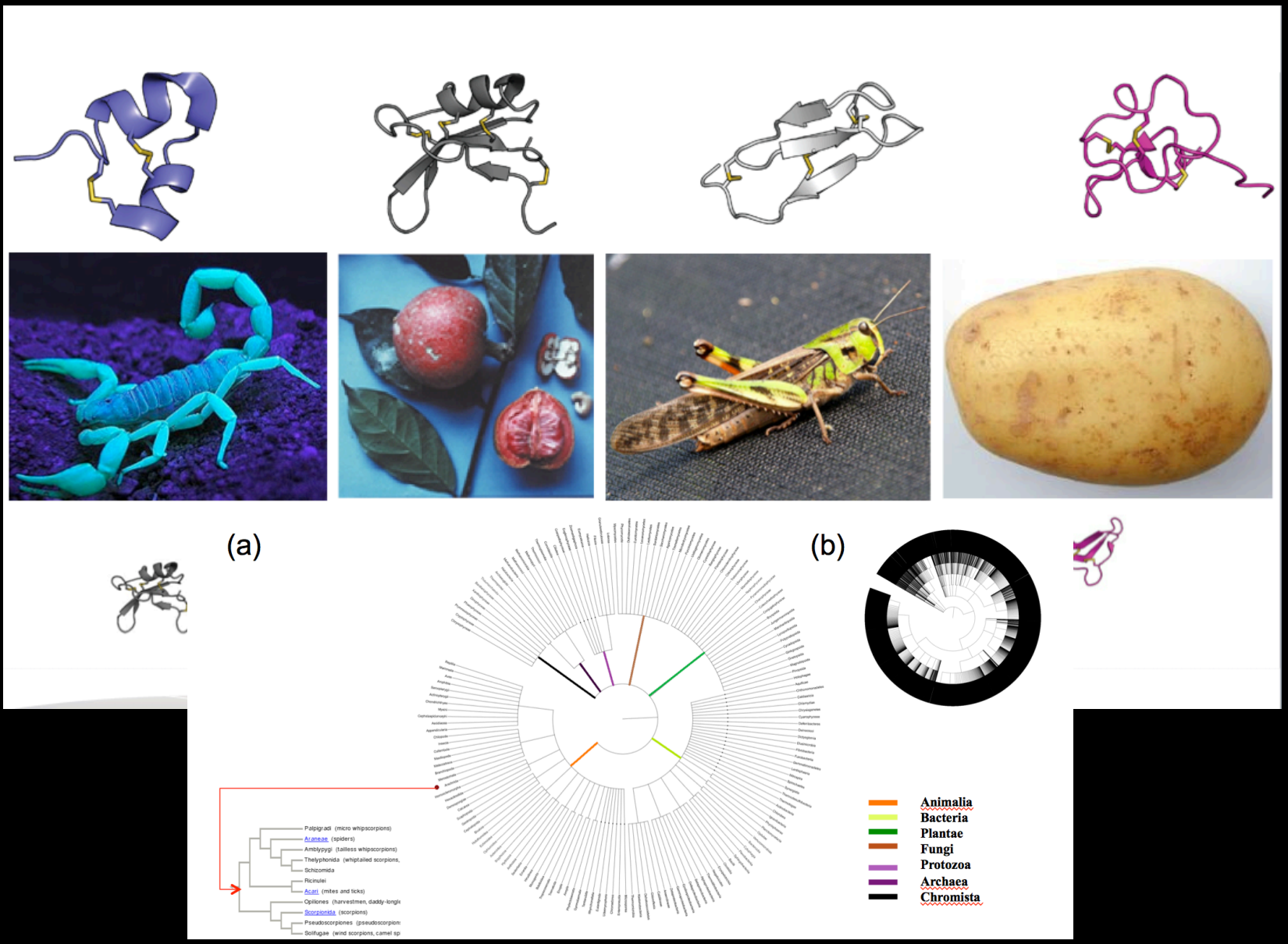


Coronal Brain Sectioning




Libray Building
• Bioinformatics Data Mined Library
• SAR or QSAR Model Based Variant Library
Recomb inant Protein
• Construct Vector Design
• Production QC (MS Infusion and HPLC)
PDC
• Produced PDC QC (MS Infusion and HPLC)
WBA Assay
• Target Tissue/Organ specificity analysis
PK Assay
• SRM MS
• Scintillation data
PD Assay
• Shut-gun proteomics
• Vivo/Vitro efficacy PD data

Design Libraries by Bioinformatics



Optide-Hunter Internal/External Collaboration

Optide-Hunter: Capture Compound Engineering Process

HMG: Homologue (Uniprot/Syn, Species: Entire Protein Sequence)

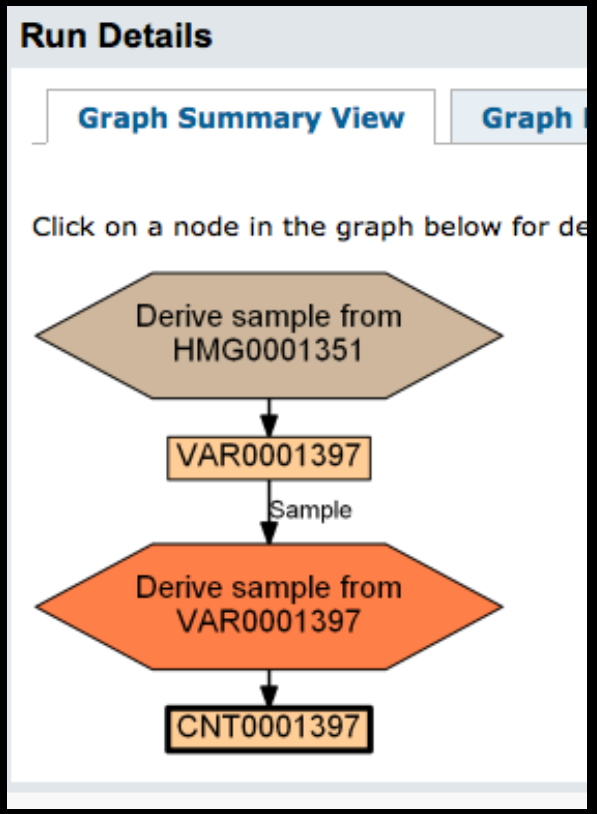
Library Screening

VAR: Variant from Homolgue Sequence

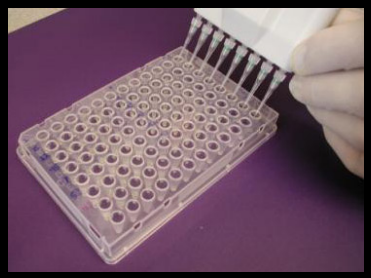
CNT: Construct (prefix, suffix Sequence Modification from Cloning Vector)

OTD and/or HT Production

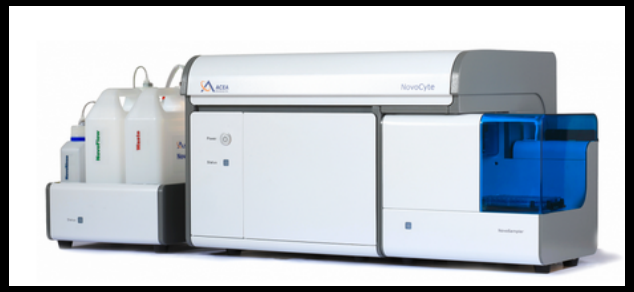
ANA: Analogues of CNT Sequences by Chemical Modification (C14 labeling, Dye or Drug or Linker Conjugates)



Optide-Hunter: Capture Assay Data



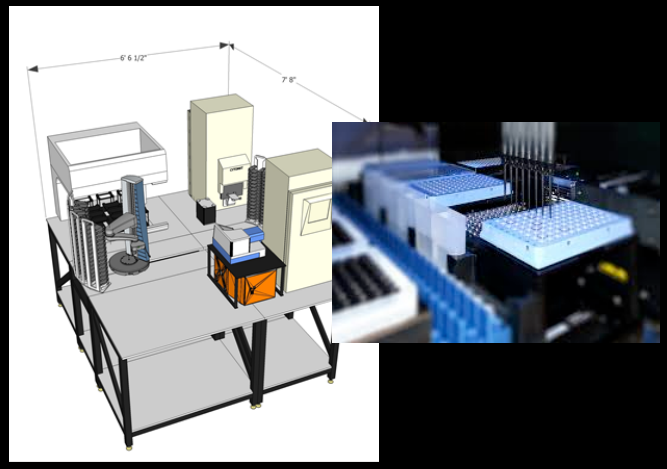
Cloned DNA



Flow Cytometry

HPLC-MS

Large scale protein production



96 well plate based HT production

Whole body autoradiography

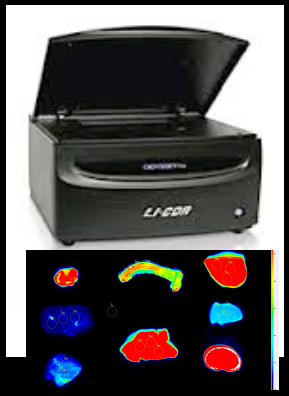
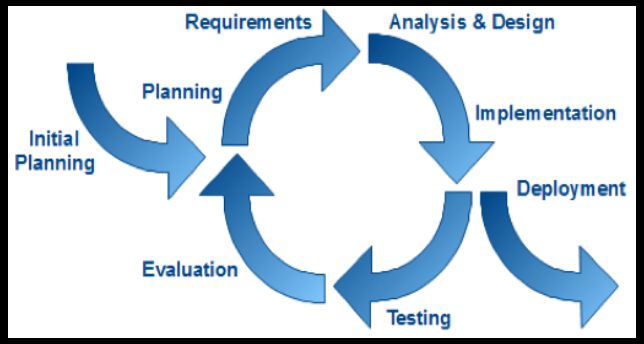


Image scanner

Challenges to Developing Optide-Hunter LIMS

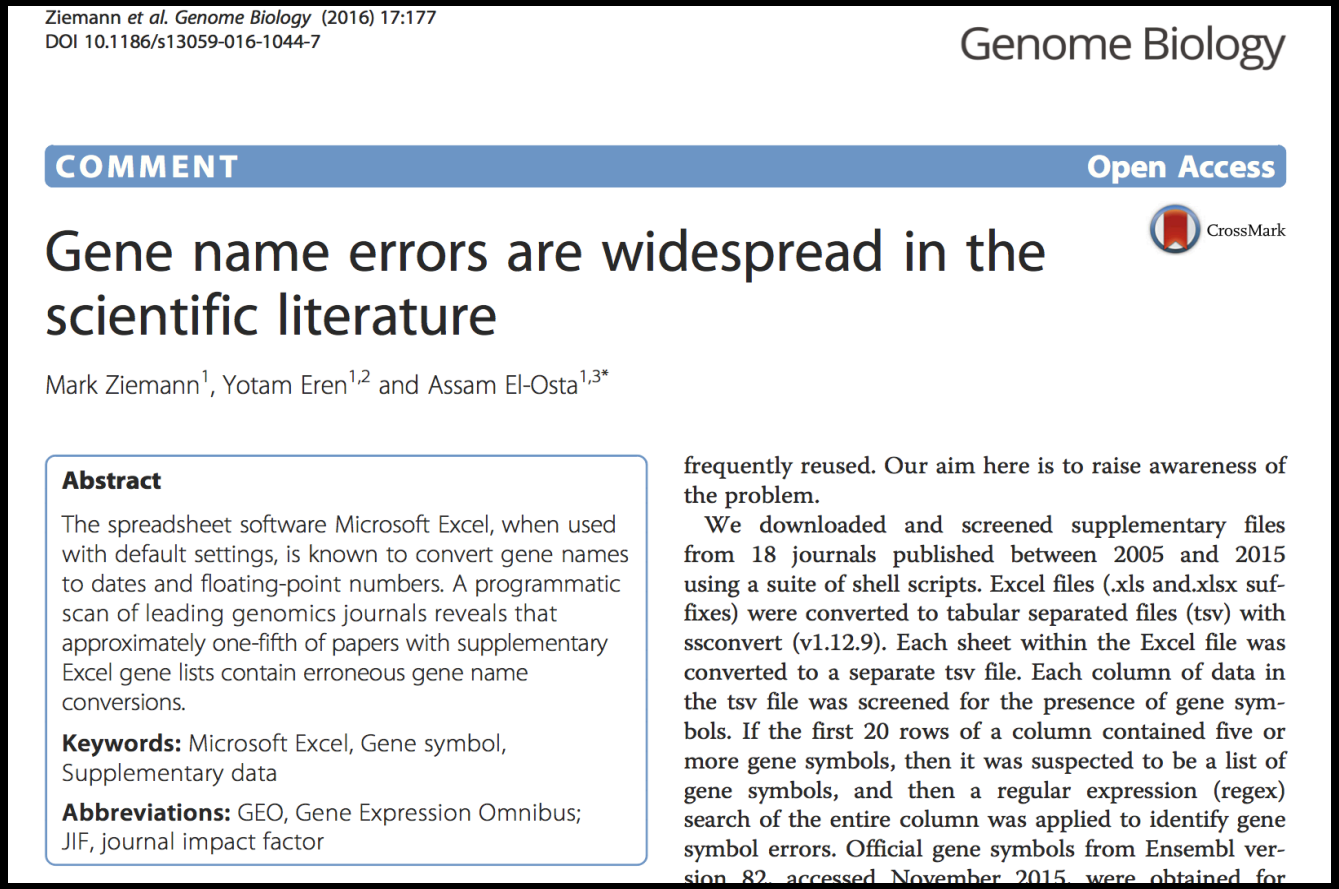
- Heterogeneous data acquisition and vendor software
  - Optimum information capture
  - Optimum usage of vendor software processing
  - Fast adaptation of new vendor data
  - Data governance and data integrity
  - Ongoing training of new and updated features
  - Ongoing automated robust testing update



Development cycle (adapt RAD)

Challenges to Developing Optide-Hunter LIMS

- Retrospect data collection



Ziemann *et al. Genome Biology* (2016) 17:177
DOI 10.1186/s13059-016-1044-7

Genome Biology

COMMENT Open Access

Gene name errors are widespread in the scientific literature

Mark Ziemann[1], Yotam Eren[1,2] and Assam El-Osta[1,3*]

Abstract

The spreadsheet software Microsoft Excel, when used with default settings, is known to convert gene names to dates and floating-point numbers. A programmatic scan of leading genomics journals reveals that approximately one-fifth of papers with supplementary Excel gene lists contain erroneous gene name conversions.

Keywords: Microsoft Excel, Gene symbol, Supplementary data

Abbreviations: GEO, Gene Expression Omnibus; JIF, journal impact factor

frequently reused. Our aim here is to raise awareness of the problem.

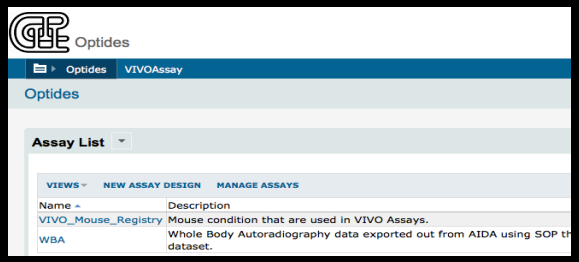
We downloaded and screened supplementary files from 18 journals published between 2005 and 2015 using a suite of shell scripts. Excel files (.xls and.xlsx suffixes) were converted to tabular separated files (tsv) with ssconvert (v1.12.9). Each sheet within the Excel file was converted to a separate tsv file. Each column of data in the tsv file was screened for the presence of gene symbols. If the first 20 rows of a column contained five or more gene symbols, then it was suspected to be a list of gene symbols, and then a regular expression (regex) search of the entire column was applied to identify gene symbol errors. Official gene symbols from Ensembl version 82, accessed November 2015, were obtained for

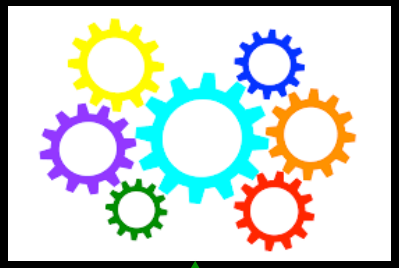
Optide-Hunter LIMS Server

GUI

Processing

Database





SQL Server 2012

Optides-Prod
Optides-Stag

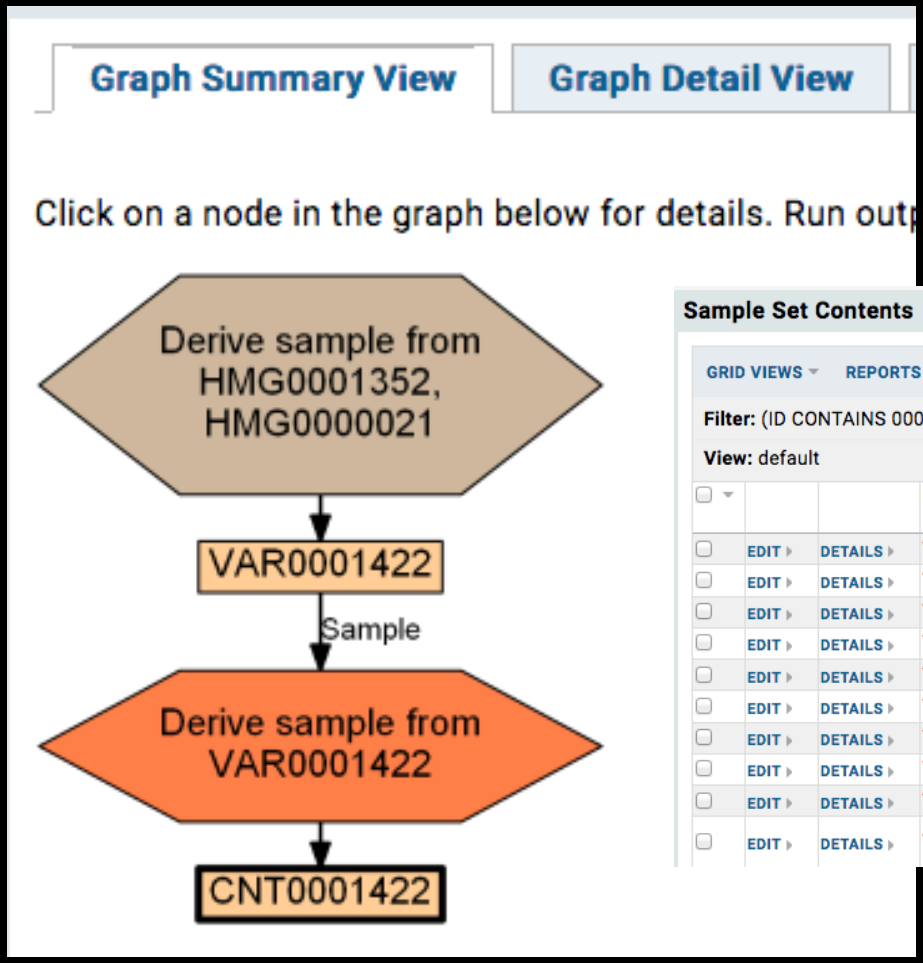
Java version: 1.8
Tomcat: 8.0.33
Labkey: 16.2
Selenium: 2.9.1

SQL Server 2012
11.0.3513.0

Labkey Feature Integration to Optide-Hunter

- Sample set – Compound Lineage Tracking (lookup features)
- FreezerPro integration features
- Schema feature
- Pipeline feature
- Transformation feature
- Wiki for user-guid
- List feature for controlled vocabulary
- User group feature

Optide-Hunter: Sample set – Compound Lineage Tracking (lookup features)



Sample Set Contents

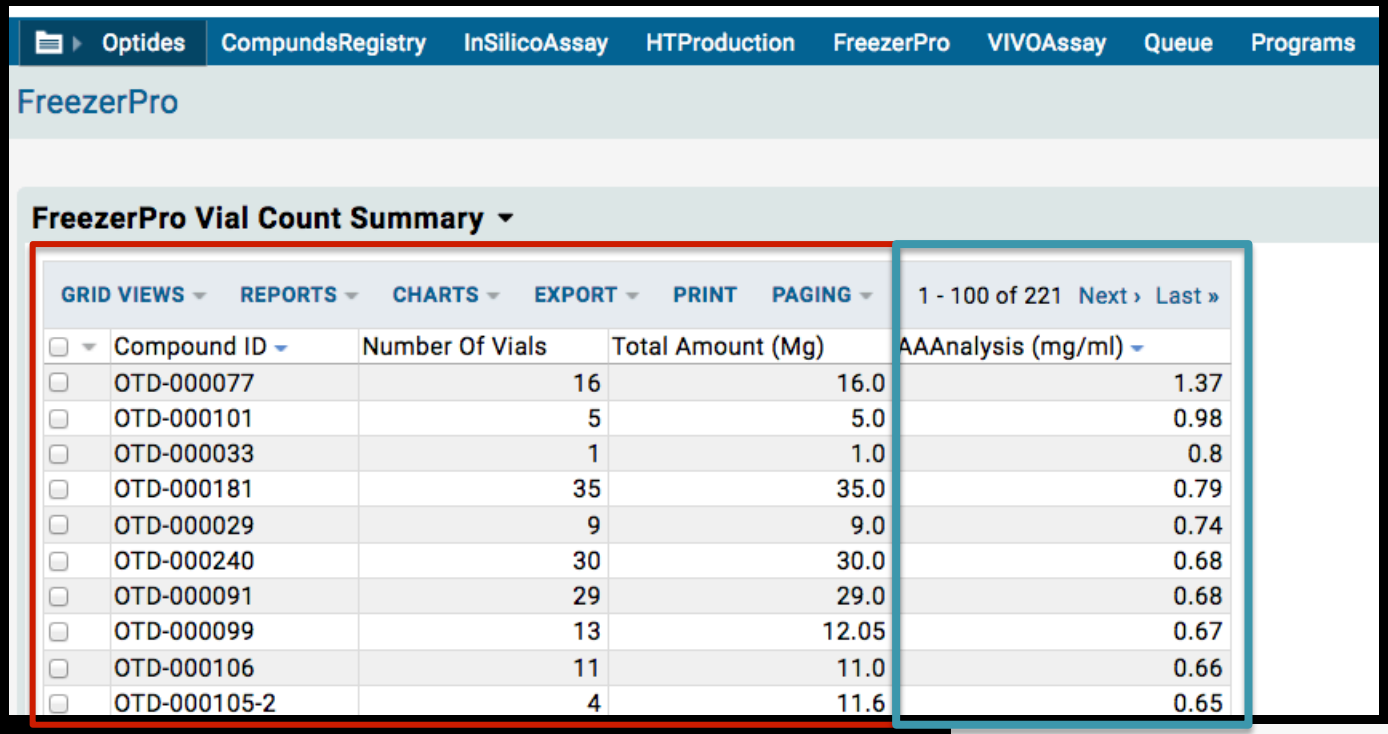
GRID VIEWS | REPORTS | CHARTS | INSERT | EXPORT | PRINT | PAGING | DELETE | DERIVE SAMPLES

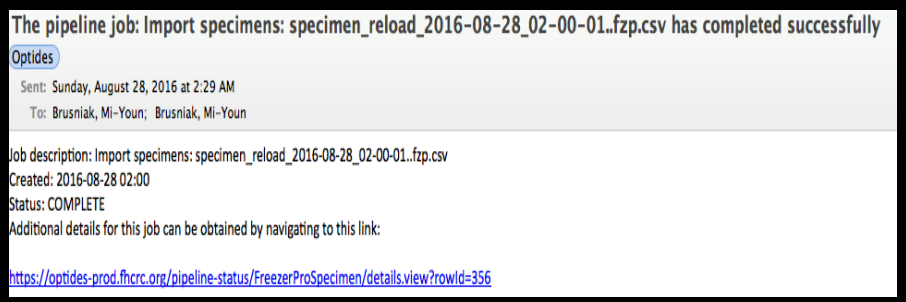
Filter: (ID CONTAINS 000142)

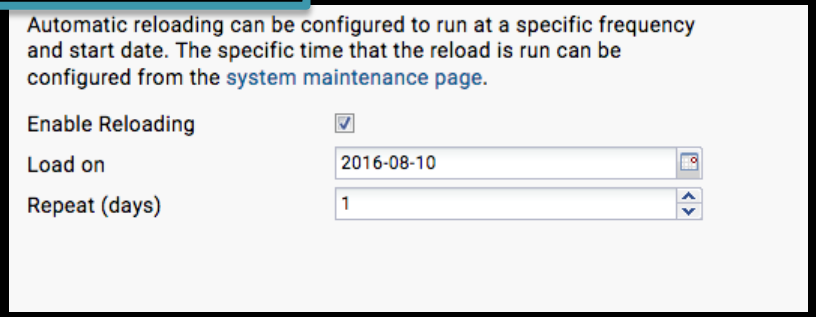
View: default

			Name	Sample Set	Flag	ID	Parent ID	AASeq
	EDIT	DETAILS	VAR0001429	Variant		VAR0001429	HMG0001353	ICIPCFTTDHQIARR
	EDIT	DETAILS	VAR0001428	Variant		VAR0001428	HMG0001353	ICIPCFTTDHQIARR
	EDIT	DETAILS	VAR0001427	Variant		VAR0001427	HMG0001353	MCMPCFTTEQRMA
	EDIT	DETAILS	VAR0001426	Variant		VAR0001426	HMG0001353	MCMPCFTTEQRMA
	EDIT	DETAILS	VAR0001425	Variant		VAR0001425	HMG0001353	MCMPCFTTDTQMQ
	EDIT	DETAILS	VAR0001424	Variant		VAR0001424	HMG0001353	MCMPCFTTDTQMQ
	EDIT	DETAILS	VAR0001423	Variant		VAR0001423	HMG0001353	MCMPCFTTHHRMA
	EDIT	DETAILS	VAR0001422	Variant		VAR0001422	<HMG0000021,HMG0001352>	MKYLLPTAAAGLLL
	EDIT	DETAILS	VAR0001421	Variant		VAR0001421	HMG0001352	MEFGLSWVFLVALF
	EDIT	DETAILS	VAR0001420	Variant		VAR0001420	HMG0001352	MKYLLPTAAAGLLL

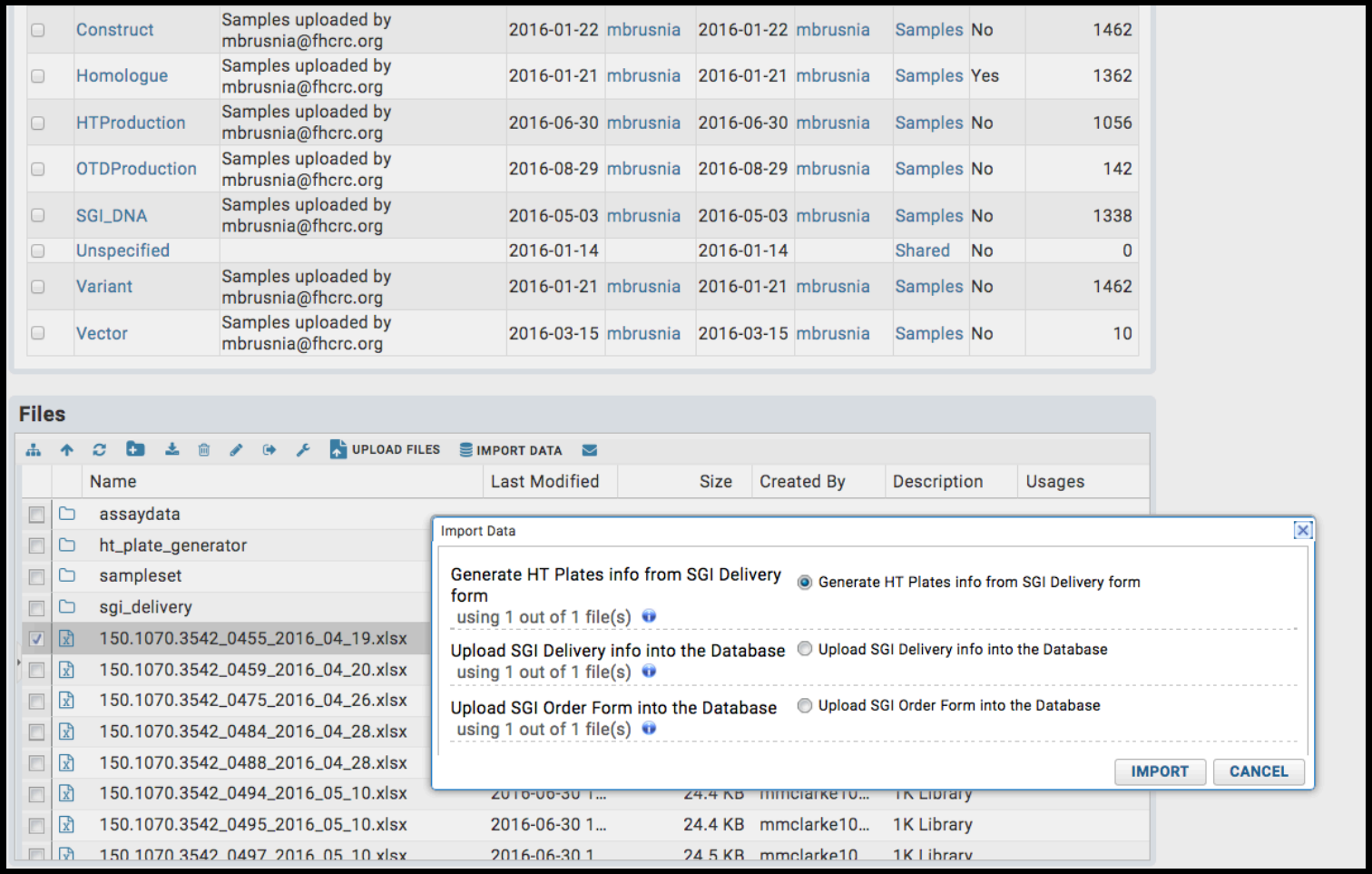
Optide-Hunter: FreezerPro integration and Schema







Optide-Hunter: Pipeline



Construct	Samples uploaded by mbrusnia@fhcrc.org	2016-01-22	mbrusnia	2016-01-22	mbrusnia	Samples	No	1462
Homologue	Samples uploaded by mbrusnia@fhcrc.org	2016-01-21	mbrusnia	2016-01-21	mbrusnia	Samples	Yes	1362
HTProduction	Samples uploaded by mbrusnia@fhcrc.org	2016-06-30	mbrusnia	2016-06-30	mbrusnia	Samples	No	1056
OTDProduction	Samples uploaded by mbrusnia@fhcrc.org	2016-08-29	mbrusnia	2016-08-29	mbrusnia	Samples	No	142
SGI_DNA	Samples uploaded by mbrusnia@fhcrc.org	2016-05-03	mbrusnia	2016-05-03	mbrusnia	Samples	No	1338
Unspecified		2016-01-14		2016-01-14		Shared	No	0
Variant	Samples uploaded by mbrusnia@fhcrc.org	2016-01-21	mbrusnia	2016-01-21	mbrusnia	Samples	No	1462
Vector	Samples uploaded by mbrusnia@fhcrc.org	2016-03-15	mbrusnia	2016-03-15	mbrusnia	Samples	No	10

Optide-Hunter: Transformation and Custom View R Scripts

WBA Results

Whole Body Autoradiography data exported out from AIDA using SOP that associated with the dataset.

MANAGE ASSAY DESIGN | VIEW BATCHES | VIEW RUNS | VIEW RESULTS | VIEW COPY-TO-STUDY HISTORY

GRID VIEWS | REPORTS | CHARTS | DELETE | EXPORT | PRINT | PAGING | COPY TO STUDY | IMPORT DATA | RE-IMPORT RUN | REPLACED FILTER

Filter: (Run = 5901)

View: default

	Grp	GrpName	Tissue	Type	Area [mm2]	Intensity [QL]	Intensity-Bkg [QL]	Intensity/Area [QL/mm2]	Intensity/Area-Bkg [QL/mm2]	Std. Activity [DPM]	Recalc. Activity [DPM]	Fitted [DPM]	Norm [DPM/mm2]
EDIT	0		int cont		23.50266542	1.70696897E8	1.671278554E8	7262873.974	7111017.086		825.1786094	759.504666229816	32.3156821857108
EDIT	0		int cont		23.50266542	1.45139331E8	1.415702894E8	6175441.313	6023584.425		704.4293954	613.520389849127	26.1042898277844

WBA Results

Whole Body Autoradiography data exported out from AIDA using SOP that associated with the dataset.

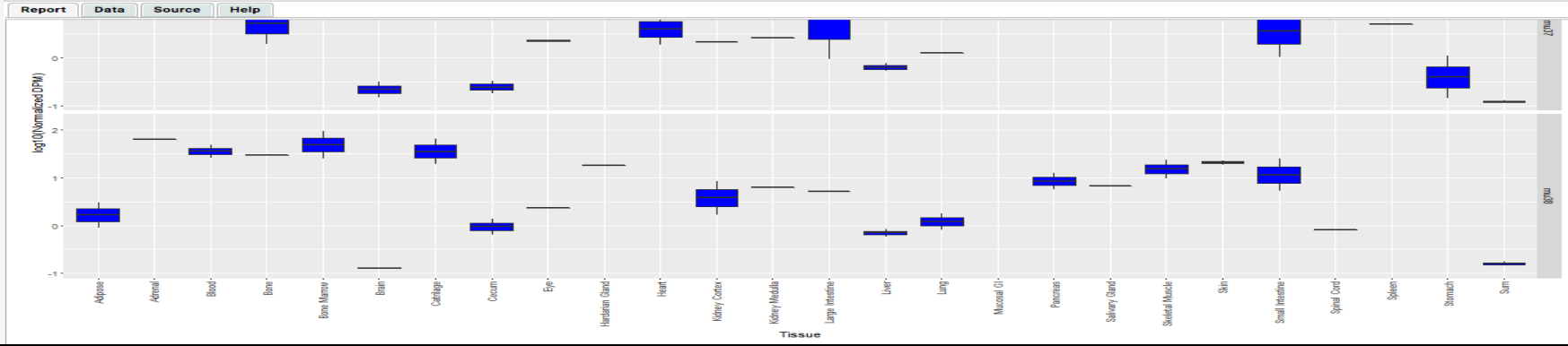
MANAGE ASSAY DESIGN | VIEW BATCHES | VIEW RUNS | VIEW RESULTS | VIEW COPY-TO-STUDY HISTORY

GRID VIEWS | REPORTS | PRINT

Filter: (Run = 5901)

Name: SliceVarianceBoxPlot **Source:** default

Report | Data | Source | Help



Optide-Hunter: file image features

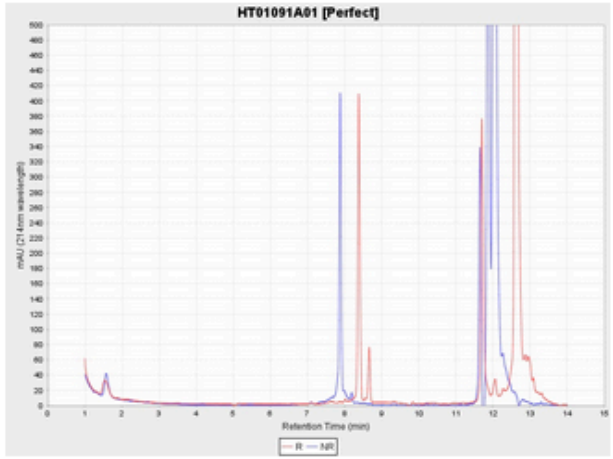
GRID VIEWS ▾ REPORTS ▾ CHARTS ▾ EXPORT ▾ PRINT PAGING ▾ COPY TO STUDY IMPORT DATA RE-IMPORT RUN RE

Filter: (Run = 6249)

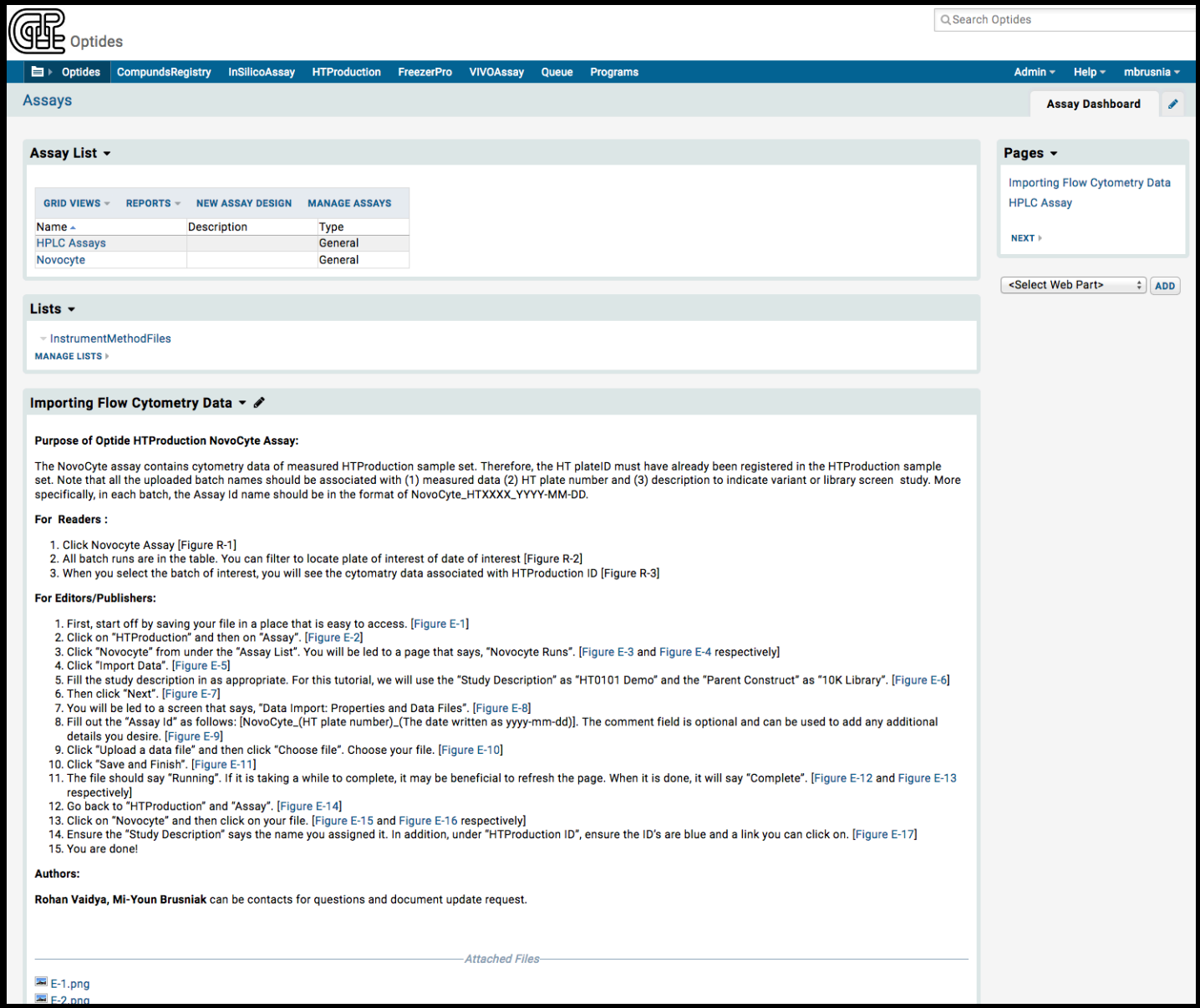
View: default

HTProductionID	Classification	Image	Max Peak NR	Study Description	HPLCProcessing Bat F
HT01091A01	Perfect		410.38	1K Library	assaydata\runGraphingHPLCQ1Q2Q3Q4_
HT01091A02	Complex				HPLCQ1Q2Q3Q4_
HT01091A03	Complex				HPLCQ1Q2Q3Q4_
HT01091A04	Simple				HPLCQ1Q2Q3Q4_
HT01091A05	Complex				HPLCQ1Q2Q3Q4_
HT01091A06	Complex				HPLCQ1Q2Q3Q4_
HT01091B01	Complex				HPLCQ1Q2Q3Q4_
HT01091B02	Perfect				HPLCQ1Q2Q3Q4_
HT01091B03	Perfect				HPLCQ1Q2Q3Q4_
HT01091B04	Perfect				HPLCQ1Q2Q3Q4_
HT01091B05	Complex				HPLCQ1Q2Q3Q4_
HT01091B06	Simple		206.79	1K Library	assaydata\runGraphingHPLCQ1Q2Q3Q4_

HPLC_Assay_Images\HT0109\HT01091A01_Perfect_410.38.jpg ×



Optide-Hunter: Wiki for User Guide



Current Status of Optide-Hunter

- 40+ users
- Role assignment
- Collecting ~5K sequence data
- DNA, Flow Cytometry, HPLC, HP ID
- Wiki
- Several custom schema and external pipeline modules

Future Plan for Optide-Hunter

- Intelligent Report for optimized decision making
- Use HELM notation for sequence description
- FDA 21 CFR Part 11 compliance
- Utilizing Labkey's new features
  - Retrospect usability
  - Testing and roll out resources

THANK YOU

project violet
at FRED HUTCH