

Optide-Hunter

Mi-Youn Brusniak, Ph.D.

2016, Oct 7

Introduction to Optide Program



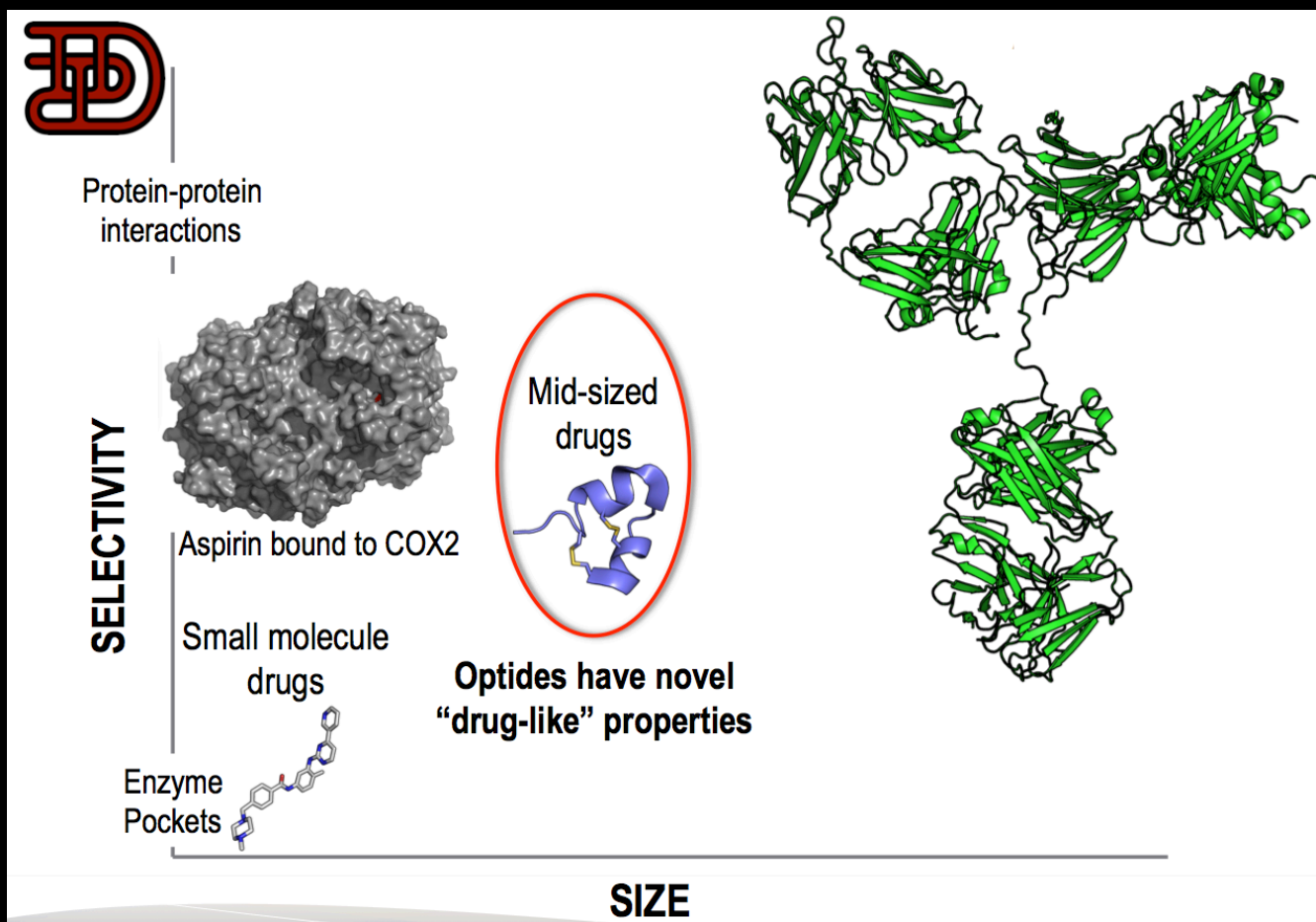
Project Violet

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

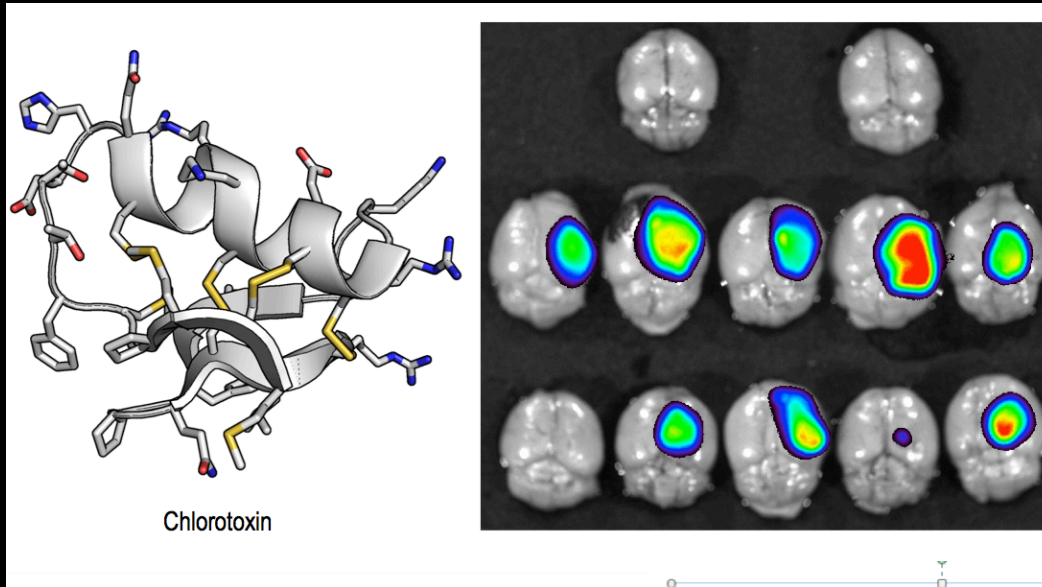


To fully realize the potential of the optide platform to maximize the benefit to human health.

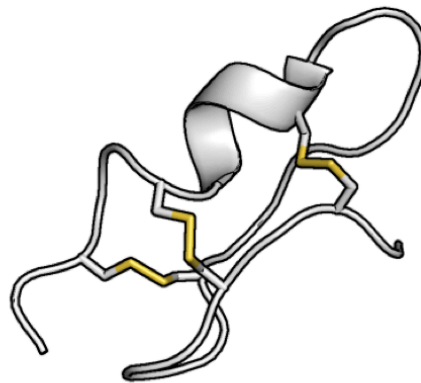
Optimized Peptides (Optides)



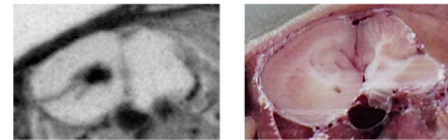
Natural Knottin Biodistribution



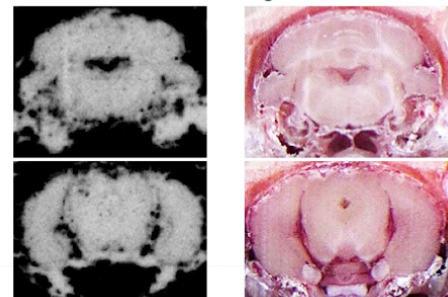
Scorpion Derived Ion Channel Blocker



Sagittal Brain Sectioning



Coronal Brain Sectioning



Library Building

- Bioinformatics Data Mined Library
- SAR or QSAR Model Based Variant Library

Recombinant 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

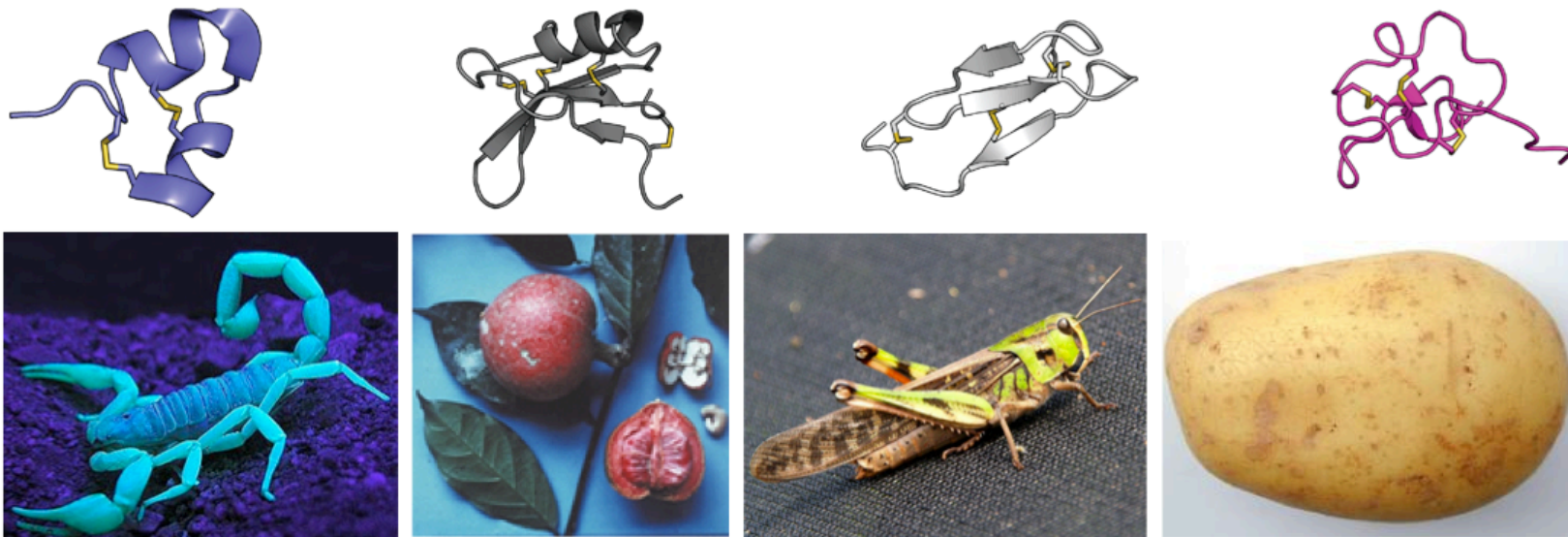
- SRM MS
- Scintillation data

PD Assay

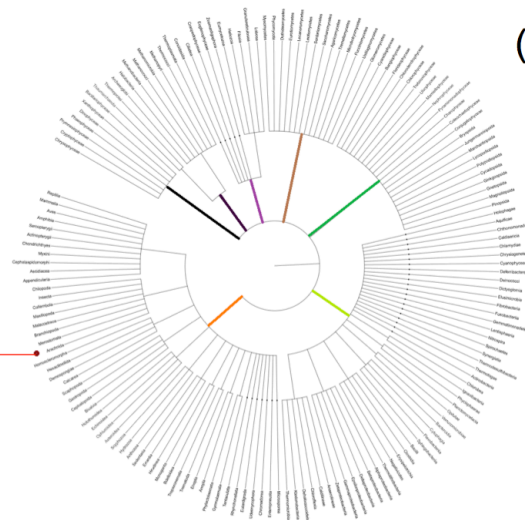
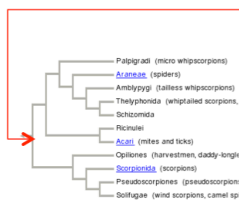
- Shut-gun proteomics
- Vivo/Vitro efficacy PD data



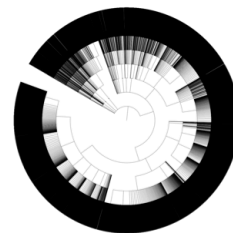
Design Libraries by Bioinformatics



(a)



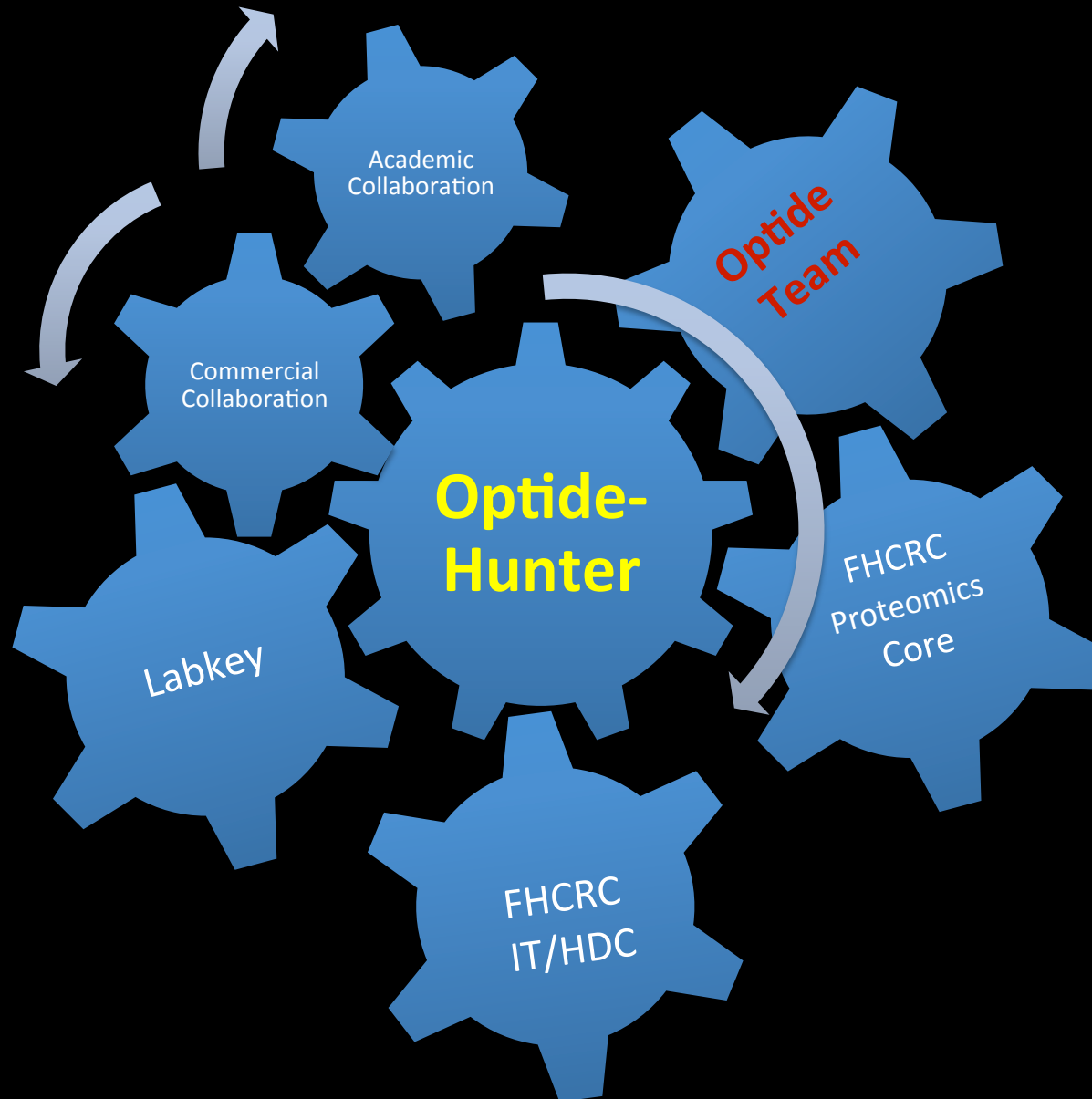
(b)



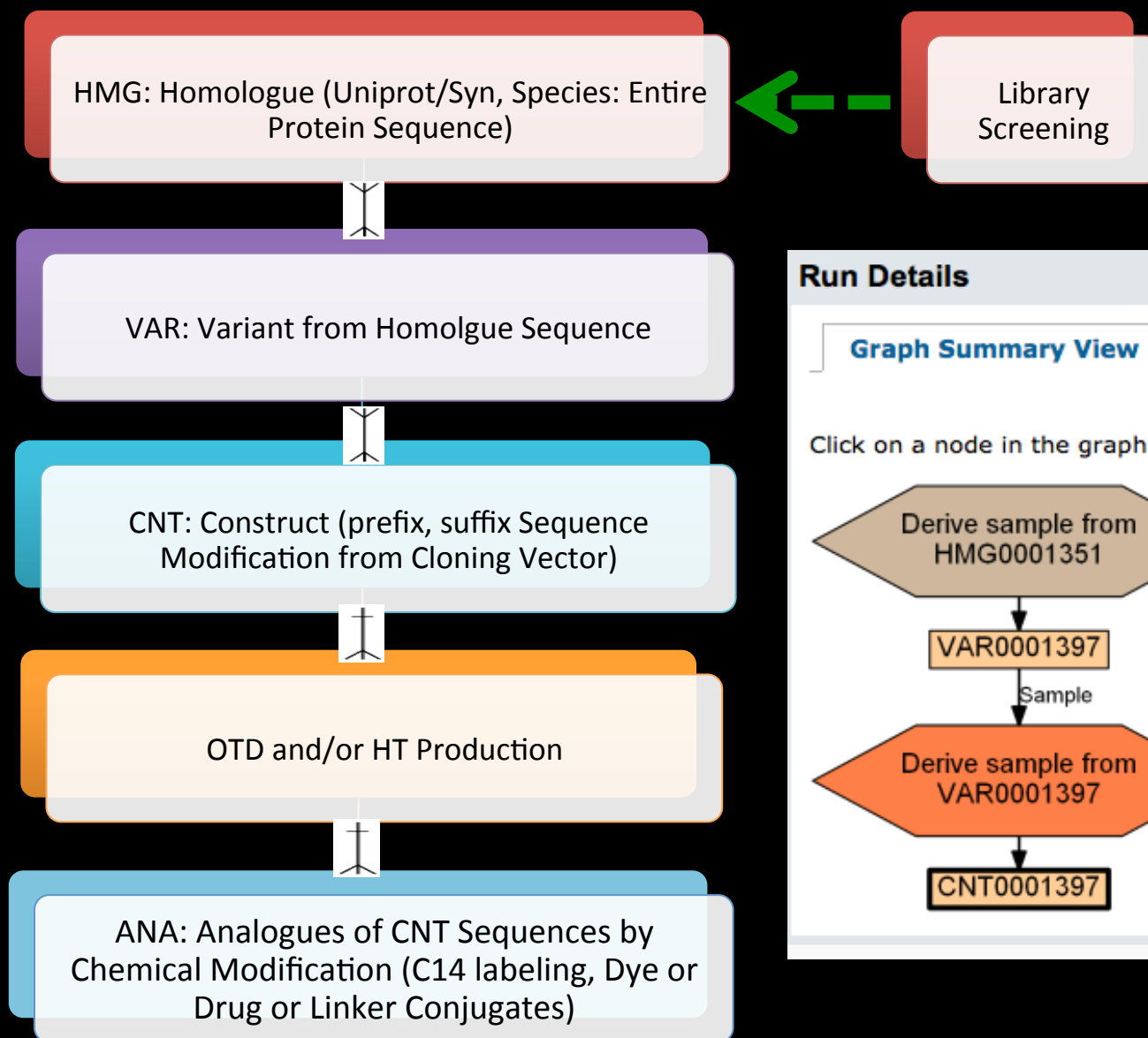
Animalia
Bacteria
Plantae
Fungi
Protozoa
Archaea
Chromista



Optide-Hunter Internal/External Collaboration



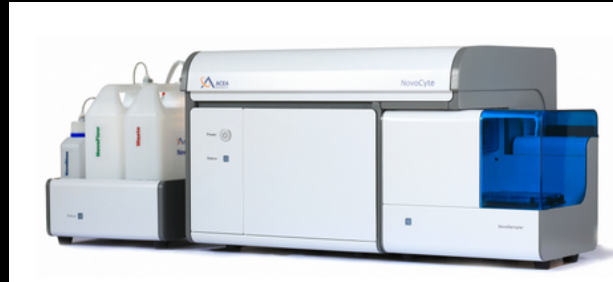
Optide-Hunter: Capture Compound Engineering Process



Optide-Hunter: Capture Assay Data



Cloned DNA



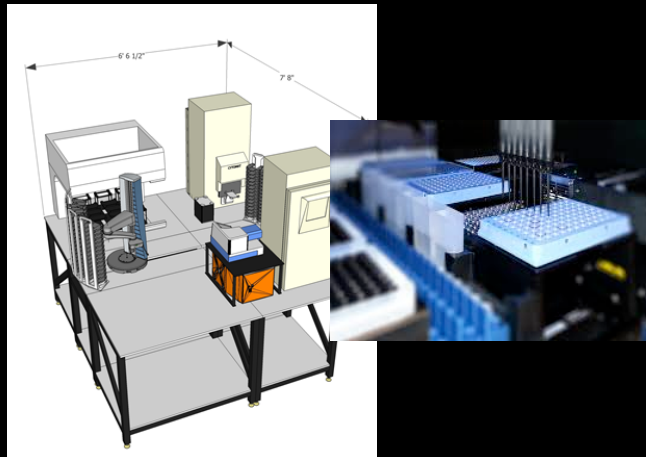
Flow Cytometry



HPLC-MS



Large scale
protein
production



96 well plate based HT
production



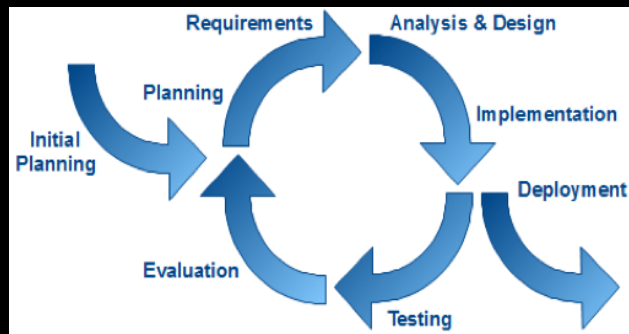
Whole body
autoradiography



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 Ziemann¹, 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

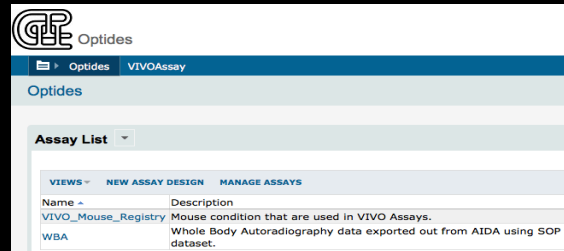
Abbreviations: GEO, Gene Expression Omnibus; JIF, journal impact factor

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

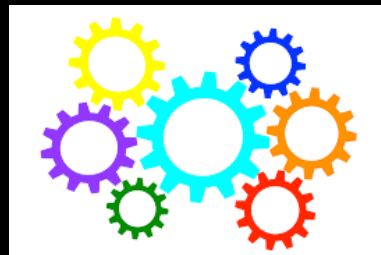
Optide-Hunter LIMS Server

GUI



Optides-Prod
Optides-Stag

Processing



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

Database



SQL Server 2012
11.0.3513.0

Labkey Feature Integration to Optide-Hunter

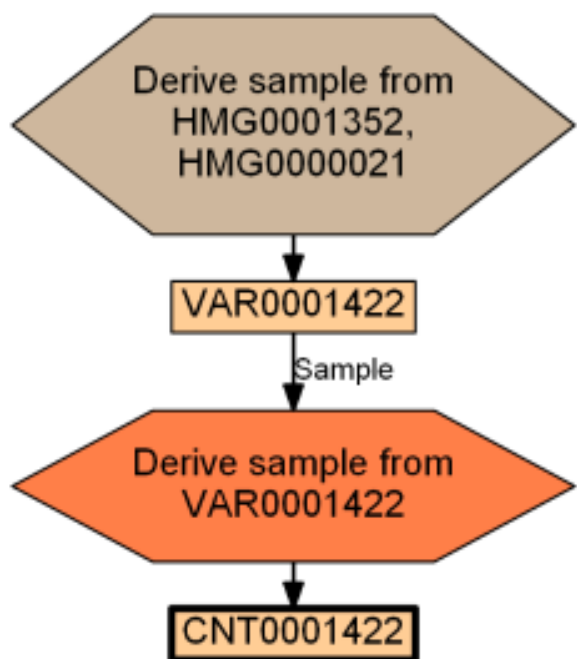
- Sample set – Compound Lineage Tracking (lookup features)
- FreezerPro integration features
- Schema feature
- Pipeline feature
- Transformation feature
- Wiki for user-guid
- List feature for controlled vocabulary
- User group feature

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

Graph Summary View

Graph Detail View

Click on a node in the graph below for details. Run outp



Sample Set Contents

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

Filter: (ID CONTAINS 000142)

View: default

<input type="checkbox"/>			Name	Sample Set	Flag	ID ▾	Parent ID	AASeq
<input type="checkbox"/>	EDIT ▸	DETAILS ▸	VAR0001429	Variant		VAR0001429	HMG0001353	ICIPCFTTDHQIARR0
<input type="checkbox"/>	EDIT ▸	DETAILS ▸	VAR0001428	Variant		VAR0001428	HMG0001353	ICIPCFTTDHQIARR0
<input type="checkbox"/>	EDIT ▸	DETAILS ▸	VAR0001427	Variant		VAR0001427	HMG0001353	MCMPCFTEQRMA
<input type="checkbox"/>	EDIT ▸	DETAILS ▸	VAR0001426	Variant		VAR0001426	HMG0001353	MCMPCFTEQRMA
<input type="checkbox"/>	EDIT ▸	DETAILS ▸	VAR0001425	Variant		VAR0001425	HMG0001353	MCMPCFTTDTQMC
<input type="checkbox"/>	EDIT ▸	DETAILS ▸	VAR0001424	Variant		VAR0001424	HMG0001353	MCMPCFTTDTQMC
<input type="checkbox"/>	EDIT ▸	DETAILS ▸	VAR0001423	Variant		VAR0001423	HMG0001353	MCMPCFTHHRMA
<input type="checkbox"/>	EDIT ▸	DETAILS ▸	VAR0001422	Variant		VAR0001422	<HMG0000021,HMG0001352>	MKYLLPTAAAGLLL
<input type="checkbox"/>	EDIT ▸	DETAILS ▸	VAR0001421	Variant		VAR0001421	HMG0001352	MEFGLSWVFLVALF
<input type="checkbox"/>	EDIT ▸	DETAILS ▸	VAR0001420	Variant		VAR0001420	HMG0001352	MKYLLPTAAAGLLL

Optide-Hunter: FreezerPro integration and Schema

Optides

CompundsRegistry

InSilicoAssay

HTProduction

FreezerPro

VIVOAssay

Queue

Programs

FreezerPro

FreezerPro Vial Count Summary ▾

GRID VIEWS ▾

REPORTS ▾

CHARTS ▾

EXPORT ▾

PRINT

PAGING ▾

1 - 100 of 221

Next ▸

Last »

<input type="checkbox"/> ▾	Compound ID ▾	Number Of Vials	Total Amount (Mg)	AAAnalysis (mg/ml) ▾
<input type="checkbox"/>	OTD-000077	16	16.0	1.37
<input type="checkbox"/>	OTD-000101	5	5.0	0.98
<input type="checkbox"/>	OTD-000033	1	1.0	0.8
<input type="checkbox"/>	OTD-000181	35	35.0	0.79
<input type="checkbox"/>	OTD-000029	9	9.0	0.74
<input type="checkbox"/>	OTD-000240	30	30.0	0.68
<input type="checkbox"/>	OTD-000091	29	29.0	0.68
<input type="checkbox"/>	OTD-000099	13	12.05	0.67
<input type="checkbox"/>	OTD-000106	11	11.0	0.66
<input type="checkbox"/>	OTD-000105-2	4	11.6	0.65

The pipeline job: Import specimens: specimen_reload_2016-08-28_02-00-01.fzp.csv has completed successfully

Optides

Sent: Sunday, August 28, 2016 at 2:29 AM

To: Brusniak, Mi-Youn; Brusniak, Mi-Youn

Job description: Import specimens: specimen_reload_2016-08-28_02-00-01.fzp.csv

Created: 2016-08-28 02:00

Status: COMPLETE

Additional details for this job can be obtained by navigating to this link:

<https://optides-prod.fhcrc.org/pipeline-status/FreezerProSpecimen/details.view?rowId=356>

Automatic reloading can be configured to run at a specific frequency and start date. The specific time that the reload is run can be configured from the [system maintenance page](#).

Enable Reloading

☒

Load on

2016-08-10

Repeat (days)

1

Optide-Hunter: Pipeline

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

Files

UPLOAD FILES

IMPORT DATA

	Name	Last Modified	Size	Created By	Description	Usages
<input type="checkbox"/>	assaydata					
<input type="checkbox"/>	ht_plate_generator					
<input type="checkbox"/>	sampleset					
<input type="checkbox"/>	sgi_delivery					
<input checked="" type="checkbox"/>	150.1070.3542_0455_2016_04_19.xlsx					
<input type="checkbox"/>	150.1070.3542_0459_2016_04_20.xlsx					
<input type="checkbox"/>	150.1070.3542_0475_2016_04_26.xlsx					
<input type="checkbox"/>	150.1070.3542_0484_2016_04_28.xlsx					
<input type="checkbox"/>	150.1070.3542_0488_2016_04_28.xlsx					
<input type="checkbox"/>	150.1070.3542_0494_2016_05_10.xlsx	2016-06-30 1...	24.4 KB	mmclarke10...	1K Library	
<input type="checkbox"/>	150.1070.3542_0495_2016_05_10.xlsx	2016-06-30 1...	24.4 KB	mmclarke10...	1K Library	
<input type="checkbox"/>	150.1070.3542_0497_2016_05_10.xlsx	2016-06-30 1...	24.5 KB	mmclarke10...	1K Library	

Import Data

Generate HT Plates info from SGI Delivery form

☒ Generate HT Plates info from SGI Delivery form

using 1 out of 1 file(s)

Upload SGI Delivery info into the Database

☐ Upload SGI Delivery info into the Database

using 1 out of 1 file(s)

Upload SGI Order Form into the Database

☐ Upload SGI Order Form into the Database

using 1 out of 1 file(s)

IMPORT

CANCEL

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

<input type="checkbox"/>		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]
<input type="checkbox"/>	EDIT	0		int cont		23.50266542	1.70696897E8	1.671278554E8	7262873.974	7111017.086		825.1786094	759.504666229816	32.3156821857108
<input type="checkbox"/>	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.

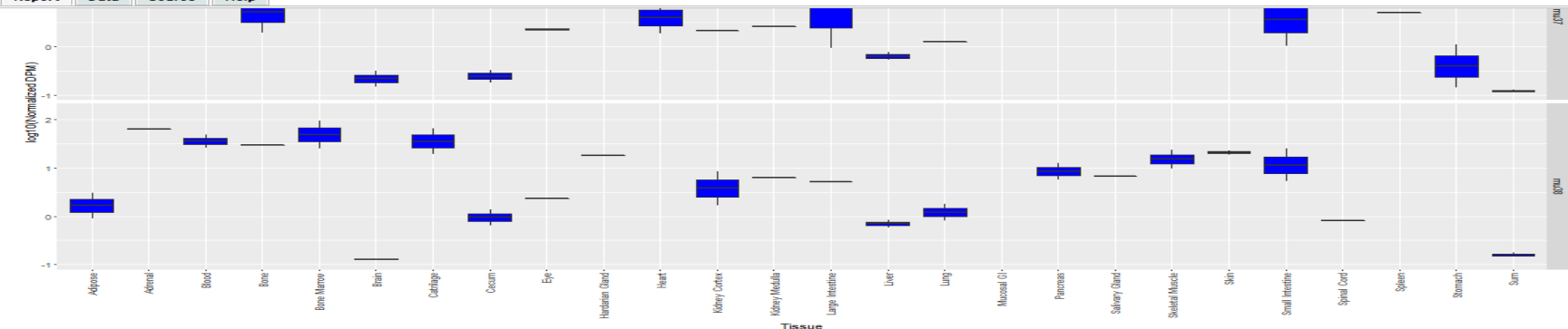
[MANAGE ASSAY DESIGN](#) > [VIEW BATCHES](#) > [VIEW RUNS](#) > [VIEW RESULTS](#) > [VIEW COPY-TO-STUDY HISTORY](#) >

[GRID VIEWS](#) > [REPORTS](#) > [PRINT](#)

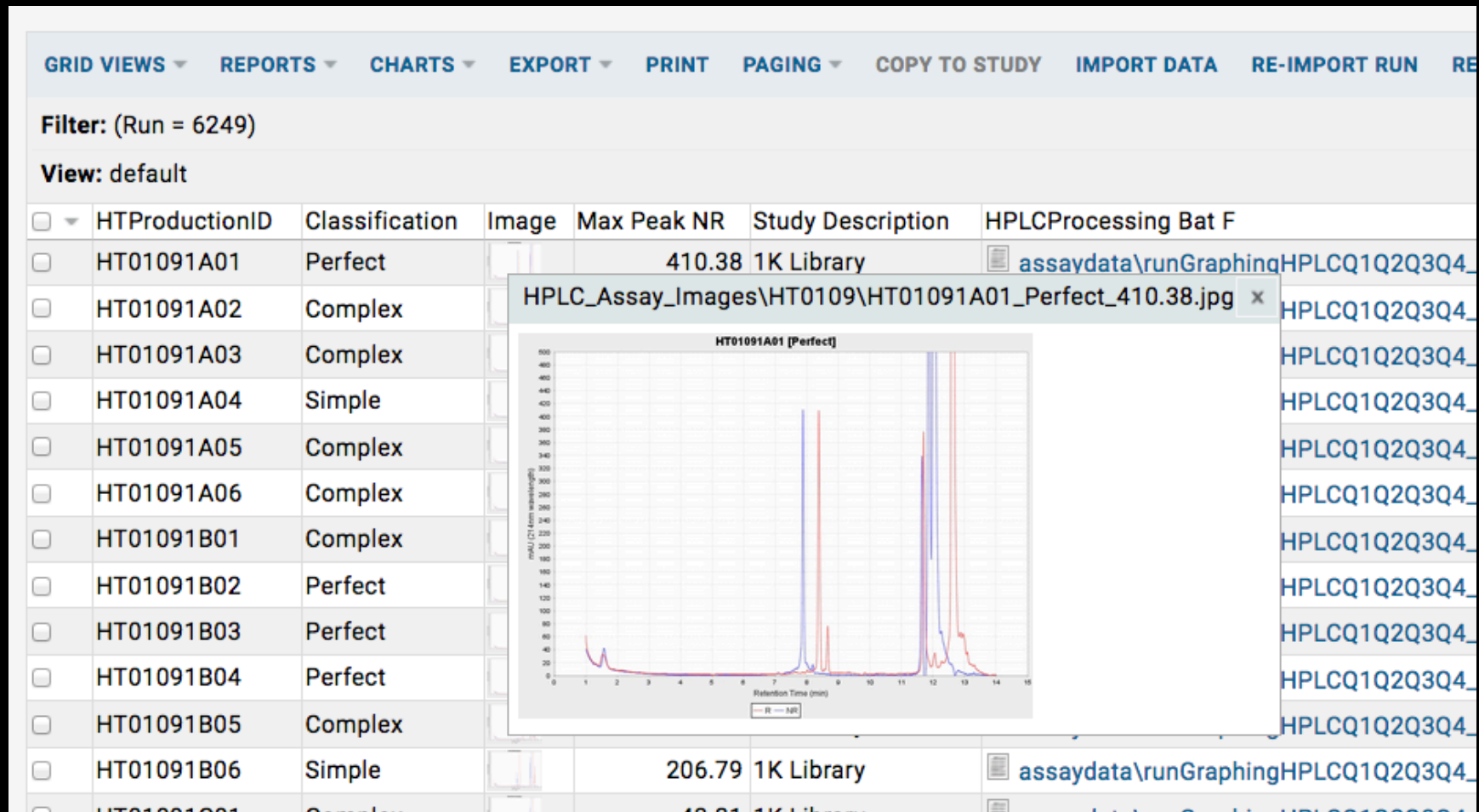
Filter: (Run = 5901)

Name: SliceVarianceBoxPlot Source: default


[Report](#) [Data](#) [Source](#) [Help](#)



Optide-Hunter: file image features



Optide-Hunter: Wiki for User Guide

 Optides

Q Search Optides

OptidesCompundsRegistryInSilicoAssayHTProductionFreezerProVIVOAssayQueueProgramsAdminHelpmbrusnia

AssaysAssay Dashboard

Assay List

GRID VIEWSREPORTSNEW ASSAY DESIGNMANAGE ASSAYS

Name	Description	Type
HPLC Assays		General
Novocyte		General

Lists

InstrumentMethodFiles

MANAGE LISTS

Importing Flow Cytometry Data

Purpose of Optide HTProduction NovoCyte Assay:

The NovoCyte assay contains cytometry data of measured HTProduction sample set. Therefore, the HT plateID must have already been registered in the HTProduction sample set. Note that all the uploaded batch names should be associated with (1) measured data (2) HT plate number and (3) description to indicate variant or library screen study. More specifically, in each batch, the Assay Id name should be in the format of NovoCyte_HTXXXX_YYYY-MM-DD.

For Readers :

1. Click Novocyte Assay [Figure R-1]
2. All batch runs are in the table. You can filter to locate plate of interest of date of interest [Figure R-2]
3. When you select the batch of interest, you will see the cytometry data associated with HTProduction ID [Figure R-3]

For Editors/Publishers:

1. First, start off by saving your file in a place that is easy to access. [Figure E-1]
2. Click on "HTProduction" and then on "Assay". [Figure E-2]
3. Click "Novocyte" from under the "Assay List". You will be led to a page that says, "Novocyte Runs". [Figure E-3 and Figure E-4 respectively]
4. Click "Import Data". [Figure E-5]
5. Fill the study description in as appropriate. For this tutorial, we will use the "Study Description" as "HT0101 Demo" and the "Parent Construct" as "10K Library". [Figure E-6]
6. Then click "Next". [Figure E-7]
7. You will be led to a screen that says, "Data Import: Properties and Data Files". [Figure E-8]
8. Fill out the "Assay Id" as follows: [NovoCyte_(HT plate number)_(The date written as yyyy-mm-dd)]. The comment field is optional and can be used to add any additional details you desire. [Figure E-9]
9. Click "Upload a data file" and then click "Choose file". Choose your file. [Figure E-10]
10. Click "Save and Finish". [Figure E-11]
11. The file should say "Running". If it is taking a while to complete, it may be beneficial to refresh the page. When it is done, it will say "Complete". [Figure E-12 and Figure E-13 respectively]
12. Go back to "HTProduction" and "Assay". [Figure E-14]
13. Click on "Novocyte" and then click on your file. [Figure E-15 and Figure E-16 respectively]
14. Ensure the "Study Description" says the name you assigned it. In addition, under "HTProduction ID", ensure the ID's are blue and a link you can click on. [Figure E-17]
15. You are done!

Authors:

Rohan Vaidya, Mi-Youn Brusniak can be contacts for questions and document update request.

Attached Files

E-1.png

E-2.png

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