Optide-Hunter

Mi-Youn Brusniak, Ph.D.
2016, Oct 7
Introduction to Optide Program
Project Violet

To fully realize the potential of the optide platform to maximize the benefit to human health.
Optimized Peptides (Optides)

- Protein-protein interactions
- Aspirin bound to COX2
- Small molecule drugs
- Enzyme Pockets

Mid-sized drugs
Optides have novel “drug-like” properties
Natural Knottin Biodistribution
Optide-Hunter Internal/External Collaboration

Optide-Hunter

Labkey

Commercial Collaboration

Academic Collaboration

FHCRC IT/HDC

FHCRC Proteomics Core

Optide Team
Optide-Hunter: Capture Compound Engineering Process

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

VAR: Variant from Homologue 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)

Library Screening
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

Gene name errors are widespread in the scientific literature

Mark Ziemann¹, Yotam Eren¹,² and Assam El-Osta¹,³*

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 sconvert (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

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)
Optide-Hunter: FreezerPro integration and Schema

FreezerPro Vial Count Summary

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

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

Job description: Import specimens: specimen_reload_2016-08-28_02-00-01.csv
Created: 2016-08-28 02:00
Status: COMPLETE

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
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.

Filter: (Run = 5901)
View: default

<table>
<thead>
<tr>
<th>Grp</th>
<th>GrpName</th>
<th>Tissue</th>
<th>Type</th>
<th>Area [mm²]</th>
<th>Intensity [QL]</th>
<th>Intensity-Bkg [QL]</th>
<th>Intensity/Area [QL/mm²]</th>
<th>Intensity/Area-Bkg [QL/mm²]</th>
<th>Std. Activity [DPM]</th>
<th>Recalc. Activity [DPM]</th>
<th>Fitted [DPM]</th>
<th>Norm [DPM/mm²]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>int cont</td>
<td></td>
<td>23.50266542</td>
<td>1.70696897E8</td>
<td>1.671278564E8</td>
<td>7262873.974</td>
<td>7111017.086</td>
<td>825.1786094</td>
<td>759.504666229816</td>
<td>32.3156821857108</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>int cont</td>
<td></td>
<td>23.50266542</td>
<td>1.45139331E8</td>
<td>1.415702894E8</td>
<td>6175441.313</td>
<td>6023844.425</td>
<td>704.4293954</td>
<td>613.520389649127</td>
<td>26.1042898277844</td>
<td></td>
</tr>
</tbody>
</table>

WBA Results

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

Filter: (Run = 5901)
Name: SliceVarianceBoxPlot Source: default
Optide-Hunter: file image features
Optide-Hunter: Wiki for User Guide

Assay List -

GRID VIEWS - REPORTS - NEW ASSAY DESIGN - MANAGE ASSAYS

Name - Description - Type
HPLC Assays - General
Novocye - 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_YYY-MM-DD.

For Readers:
1. Click NovoCyte Assay [Figure E-1]
2. All batch runs are in the table. You can filter to locate plate of interest or date of interest [Figure E-2]
3. When you select the batch of interest, you will see the cytometry data associated with HTProduction ID [Figure E-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 "HT3101 Demo" and the "Parent Construct" as "DK 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 IDs are blue and a link you can click on. [Figure E-17]
15. You are done!

Authors:
Rohan Vadika, Mi-Youn Brusilak 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