Deploying Galaxy for use with High Throughput Screening

Grainne Kerr
Signaling and Functional Genomics
Deploying Galaxy for use with HTS

RNA interference (RNAi)

dsRNA (>27 nt)

Dicer

siRNA (18 – 21 nt)

RISCs

RISC targets
homologous mRNA

Targeted degradation
Target genes with
  - Long double stranded DNA (dsDNA)
  - Small interfering DNA (siDNA)

Success depends on:
  - Specificity (homology to target and non-targets, interferon response, concentration dependant cytotoxic responses)
  - Efficiency (GC content, repeats, target site accessibility etc)

Gene silencing – a reverse engineering approach
Deploying Galaxy for use with HTS

High Throughput Screening

- Systematic silencing of whole genome
  - Every gene is targeted.
  - Phenotypic luminescence readout
Deploying Galaxy for use with HTS

Experimental Steps

RNAi Library Design

Cell based assay format

Large scale experiment

Computational analysis

Library Annotation

Genome Annotation

Screen Description

Plate List

Plate Config

Screen Data Files

Screen Log Files

Analysis Reports
Experimental Steps

- RNAi Library Design
- Cell based assay format
- Large scale experiment
- Computational analysis

- Genome Annotation
- Library Annotation
- Screen Description
- Plate List
- Plate Config
- Screen Data Files
- Screen Log Files
- Analysis Reports
Galaxy’s role in HTS

- Glues different stages together
Computational Analysis

- Design of siRNA or dsRNA reagents
- Normalization of plates
- Determination of significant changes in phenotype
- Grouping and finding patterns in the significant hits
- Retrieval and presentation of results
Deploying Galaxy for use with HTS

Overview

• Design of reagents

nextRNAi + Galaxy

• Analysis of the read-out data

web cellHTS2 + Galaxy

• Selecting and analysis of “hit” genes

Bioconductor cellHTS2 + Galaxy
Automated design and evaluation of RNAi sequences on a genome wide scale

Thomas Horn, Thomas Sandmann and Michael Boutros.
Deploying Galaxy for use with HTS

**NEXT-RNAi Workflow**

1. Target sequence
2. Feature tables
3. Low-complexity filter
4. In silico dicing
5. Prediction of specificity and efficiency
6. Discarding of siRNAs below threshold
7. Design of long dsRNAs primers
8. Ranking of designs
9. Homology, feature and genome mapping
10. Flat files, HTML report, GBrowse
NEXT RNAi Input and dependencies

- Fasta file of target sequences
- Target-group file (e.g. SNP’s, transcript variants)

- Requirements:
  - Bowtie – sequence specificity
  - Primer3 – evaluate the primers
- Optional
  - BLAST
  - BLAT – align longer sequences
  - RNAfold – evaluate the 3d structure
  - Mdust – evaluate primer sequences
Deploying Galaxy for use with HTS

NEXT RNAi in Galaxy

• Why Galaxy?
  • Number of input parameters
  • Number of dependent choices
  • Guidance of choices – conditional inputs in Galaxy
  • GUI - Not command line
  • Incorporation into workflows
Deploying Galaxy for use with HTS

NEXT RNAi in Galaxy

Please select type of run:
- Evaluation of primers for long dsRNAs or siRNAs
- de novo design of RNAi reagents
- Evaluation of primers for long dsRNAs or siRNAs
- Evaluations of long dsRNAs
- Evaluation of long dsRNA and underlying primers

Input Fasta File:
- FASTA file containing target sites for the de novo design of RNAi reagents

Reagent type:
- long dsRNA
- short interfering RNA

TIP: Upload your data using the GET Data link in the tool menu

What it does
NEXT RNAi is a software for the design and evaluation of genome-wide RNAi libraries and performs all steps from the prediction of specific and efficient RNAi target sites to the visualization of designed reagents in their genomic context. The software enables the design and evaluation of siRNAs and long dsRNAs and was implemented in an organism-independent manner allowing designs for all sequenced and annotated genomes. It requires the minimal input of desired target sequences and an off-target database.

NEXT RNAi implements several methods to predict a reagents' quality and offers many special features such as the straight-forward design of independent RNAi reagents. How these quality parameters are assessed and an overview about NEXT RNAi features is available at http://bib10-wiki.dkfz.de/signaling/wiki/display/nextnai/NEXT-RNAi+Features
Deploying Galaxy for use with HTS

NEXT-RNAI Output

NEXT-RNAI results for nextrnaiTest design(s)

Number of queries: 500
Queries covered by design(s): 459 (91.80 %)
Queries not covered by design(s): 41 (8.20 %)

More statistics on designs are here

Links to HTML results
ppn0038397-cr11
ppn0038397-cb6
ppn0034138-cr2
ppn0033388-cr2
ppn00332505-cr2
ppn0005921-cr2

Links to input text files
Database file for off-target evaluation
Database file for mapping of real primers

Links to output report files
Log file
NEXT-RNAI report file
Failed design(s)

Statistics on overall 459 design(s)

Reagent statistics
Length forward primer [nt]: 20.25 +/- 1.3
Length reverse primer [nt]: 20.36 +/- 1.43
GC content forward primer [%]: 49.58 +/- 6.85
GC content reverse primer [%]: 49.18 +/- 7.3
Melting temperature forward primer [°C]: 60.00 +/- 0.8
Melting temperature reverse primer [°C]: 59.97 +/- 0.81
Primer penalty: 2.42 +/- 2.49
Number of efficient siRNAs: 138.05 +/- 53.95

Reagent specificity
23 design(s) with 19 nt off-target effect(s)
0 design(s) have no target at all
1 design(s) with at least one region of low complexity
0 design(s) with at least one 6x CA[ATGC] repeat
436 design(s) with hits to single intended target
12 design(s) with hits to multiple intended targets
9 design(s) with hits to single intended target and other targets
2 design(s) with hits to multiple intended targets and other targets
0 design(s) with no hits to intended target but to other target(s)
0 design(s) with no target at all

Mapping status
457 design(s) located in mapping database

Download complete HTML report as *.tar.gz archive
NEXT RNAi in Galaxy Summary

- Allows for the rapid batch design of RNAi libraries for:
  - Genome wide set
  - Defined batch of genes
- Evaluation of commercial libraries
- Re-evaluation of libraries with release of updated genome annotation
• Systematic analysis of screens
  • Phenotypic results range from single numerical value to multiple dimensional images

• Standardization of experimental information

• Standardization of analysis

Deploying Galaxy for use with HTS

Import raw data files

↓

Per plate quality control

↓

Data normalization

↓

Scoring of phenotypes

↓

Annotation and analysis
Deploying Galaxy for use with HTS

cellHTS in Galaxy

# R code from vignette source 'vignettes/cellHTS2/rsrc/doc/cellhts2Complete.Rnw'

experimentName <- 'KoVLab'
datapath <- system.file(experimentName, package='cellHTS2')
rexdir(datapath)[1:12].
x <- readTablelist("PlateList.txt", name=experimentName, path=datapath).
out <- writeReport(fullfile).
out2 <- writeReport(name=TRUE, outfile=empdir())

x <- configure(x, description="" Description.txt", confFile="Plateconfig.txt", ~
\begin{verbatim}
logFile="ScreenLog.txt" , ~ 
path=dataPath).
xm <- normalizePlates(x,
\end{verbatim}
  scale="multiplicative", ~
log=FALSE, ~
method="median", ~
varianceAdjust=none").
xsc <- scoreReplicates(xm, sign="", method="zscore")
xsc <- summarizeReplicates(xsc, summary="mean")

scores <- data(xsc).
ylim <- quantile(scores, c(0.001, 0.999), na.rm=T)
boxplot(scores ~ cellAnnos$cell, col="lightblue", outline=FALSE, ylim=ylim)
y <- scores2call(xsc, 20.15, lambda=2).
png("cellhts2Complete-callValues.png")
plot(data(xsc), data(y), col="blue", pch=".",

\begin{verbatim}
xlab="scores", ylab="calls", ~ main="expression/(1/2ae\(-\lambda \text{ln}(2) \cdot 2 \cdot \mu.t))\)"\)
\end{verbatim}

devoff()
xsc <- annotate(xsc, geneIDSFile="GeneIDs.dm.1.1.txt", path=dataPath)
setSettings(list(plotList=list(reproducibility=list(include=TRUE, map=TRUE)
intensities=list(include=TRUE, map=TRUE)),
screenSummary=list(scores=list(range=c(-4, 3), ncp=TRUE))).
out <- writeReport(name=x, normalized=xm, score=xsc, force=TRUE).
writeTab(xsc, file="scores.txt")
Deploying Galaxy for use with HTS

cellHTS in Galaxy

Please enter the name of your experiment:

Apply variance Adjustment?:
- None

Scale the data?:
- additive

Is the data log transformed?:
- no

Apply Normalization?:
- median

Which method to score values?:
- zscore

Which method to summarise replicates?:
- min

Please enter the (exact) file name containing the GeneIDs:

TIP: Upload the data files of your experiment using ftp.
Deploying Galaxy for use with HTS

- Web based interactive guide to create input files.
• Annotation Libraries storage
• Standardizing analysis
• Incorporation with other specific screening analysis tools e.g.
  Redundant siRNA activity analysis (RSA)

König et al. A probability based approach for the analysis of large-scale RNAi screens,
Nature Methods, 2007
### HTS Workflows

**Annotation:** Find if human homologs exist for an input of selected KK lines

<table>
<thead>
<tr>
<th>Step</th>
<th>Annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Step 1:</strong> Input dataset</td>
<td>Input your dataset of selected KK lines</td>
</tr>
<tr>
<td>Input Dataset of Selected KK Lines&lt;br&gt;&lt;em&gt;select at runtime&lt;/em&gt;</td>
<td></td>
</tr>
<tr>
<td><strong>Step 2:</strong> Input dataset</td>
<td>Input the following file from the Shared Data Library in Vivo Drosophila Screens&lt;br&gt;&lt;em&gt;KK_TriD_BKN_CG_F8gn_Symbol_Hom_Penninger_EnsembiEnsemblPot.txt&lt;/em&gt;</td>
</tr>
<tr>
<td>Input KK Library Human Homolog Dataset&lt;br&gt;&lt;em&gt;select at runtime&lt;/em&gt;</td>
<td></td>
</tr>
<tr>
<td><strong>Step 3:</strong> Join twoDatasets</td>
<td>Please specify the transformant ID column (e.g., c1) for the joining</td>
</tr>
<tr>
<td>Join&lt;br&gt;Output dataset 'output' from step 1&lt;br&gt;using column with&lt;br&gt;Output dataset 'output' from step 2&lt;br&gt;and column 1 (value not yet validated)&lt;br&gt;Keep lines of first input that do not join with second input&lt;br&gt;Yes&lt;br&gt;Keep lines of first input that are incomplete&lt;br&gt;Yes&lt;br&gt;Fill empty columns&lt;br&gt;Yes&lt;br&gt;Only fill unjoined rows&lt;br&gt;Yes&lt;br&gt;Fill Columns by&lt;br&gt;Single fill value&lt;br&gt;Fill value</td>
<td></td>
</tr>
</tbody>
</table>
• Continue to develop and improve tools
• Expand tools to deal with wider range of screen data e.g. double knockdowns
• Training courses to develop familiarity with other tools Galaxy has to offer
Acknowledgements

- Michael Boutros
- Thomas Horn
- Chen Chen
- Oliver Pelz
- Thomas Sandmann
- The IT team in Signaling and Functional Genomics