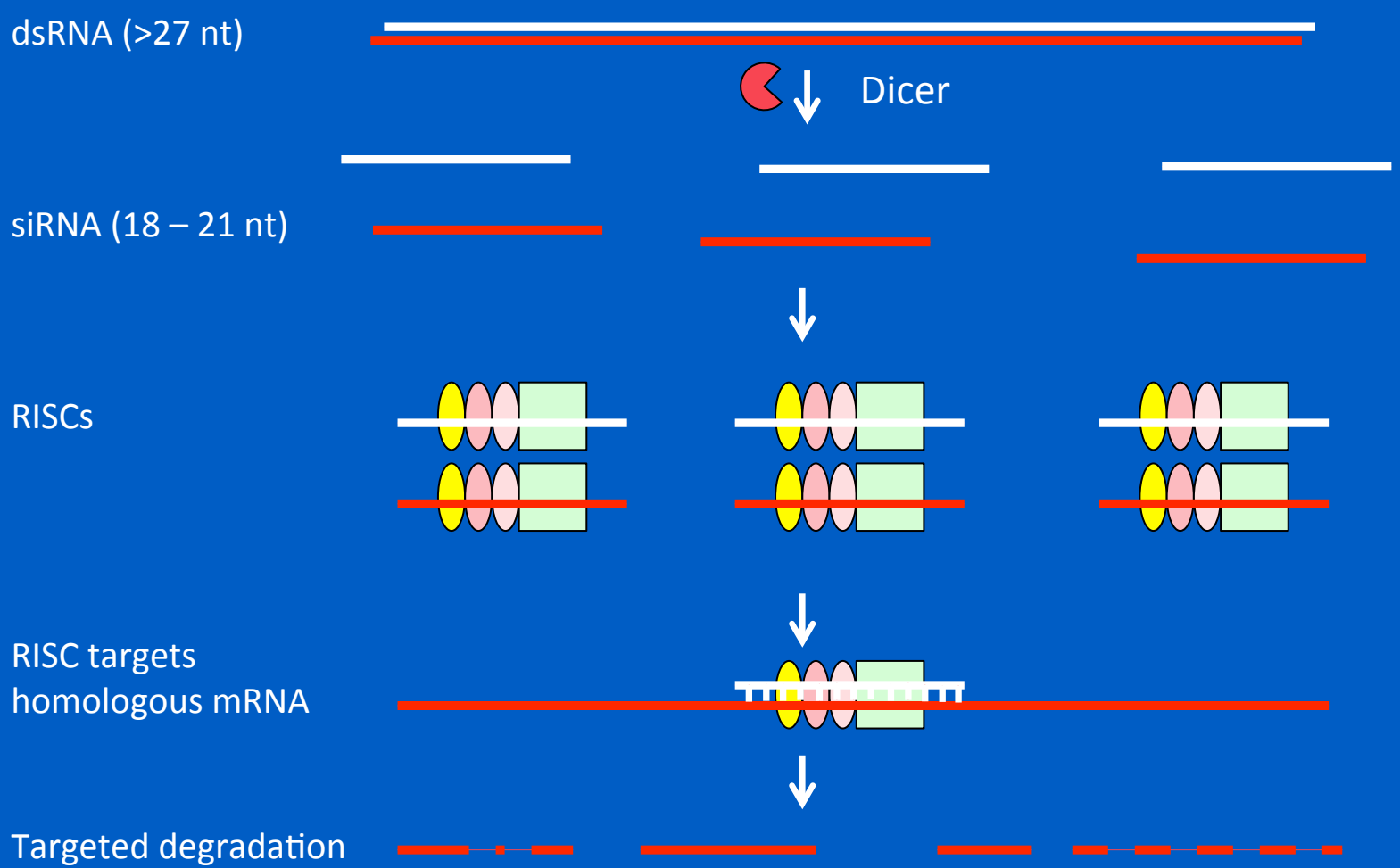
Deploying Galaxy for use with High Throughput Screening

Grainne Kerr
Signaling and Functional Genomics

dkfz.
GERMAN
CANCER RESEARCH CENTER
IN THE HELMHOLTZ ASSOCIATION

RNA interference (RNAi)

dkfz.



RNAi and Screening

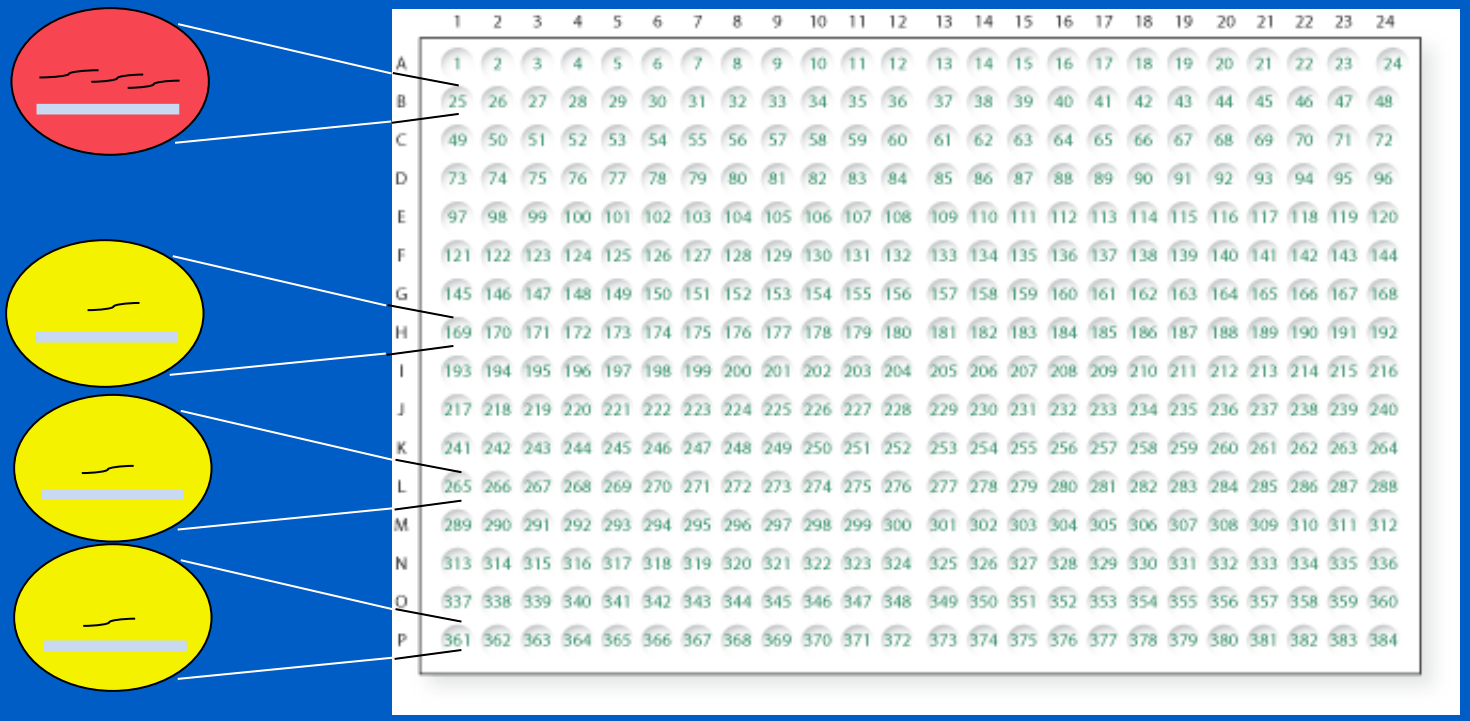
dkfz.

- 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

High Throughput Screening

dkfz.

- Systematic silencing of whole genome
  - Every gene is targeted.
  - Phenotypic luminescence readout



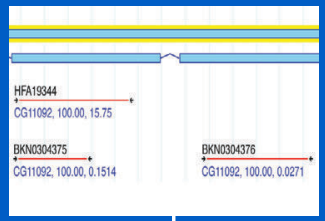
Experimental Steps

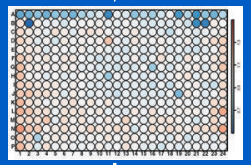
RNAi Library Design

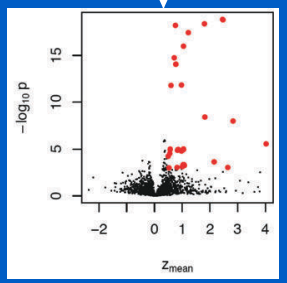
Cell based assay format

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

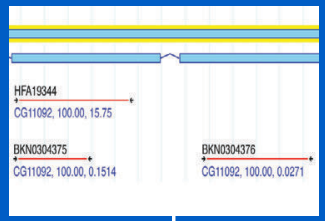
dkfz.

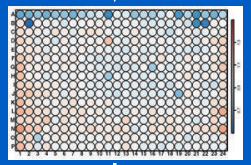
RNAi Library Design

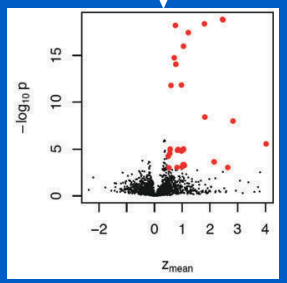
Cell based assay format

Large scale experiment

Computational analysis







Library Annotation

Genome Annotation

Screen Description

Plate List

Plate Config

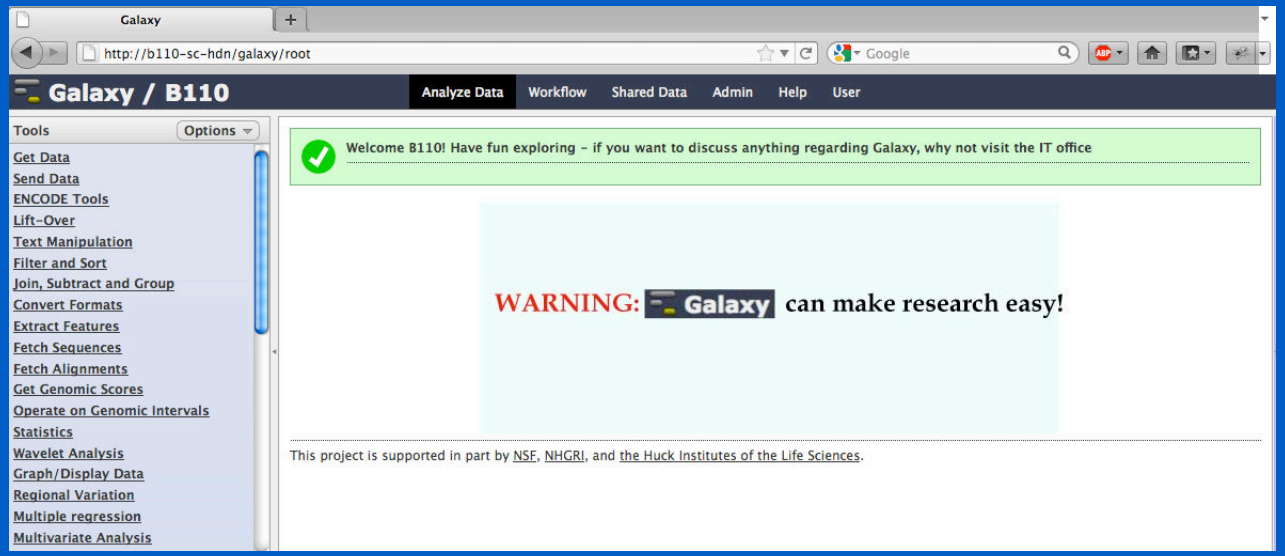
Screen Data Files

Screen Log Files

Analysis Reports

Galaxy's role in HTS

- Glues different stages together



Computational Analysis

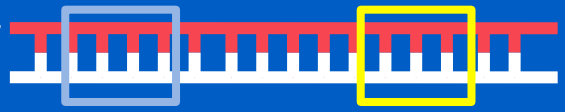
dkfz.

- 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

Overview

dkfz.

- Design of reagents

nextRNAi  + Galaxy

- Analysis of the read-out data

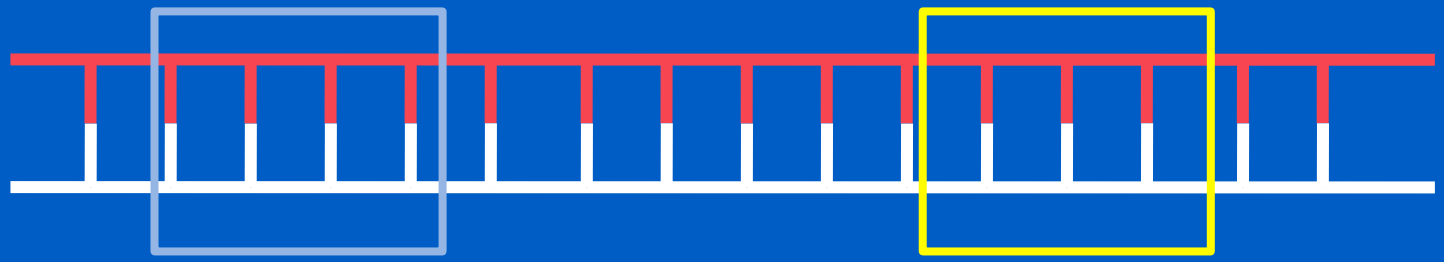
web cellHTS2 + Galaxy

- Selecting and analysis of “hit” genes

Bioconductor OPEN SOURCE SOFTWARE FOR BIOINFORMATICS cellHTS2 + Galaxy

NEXT RNAi

dkfz.

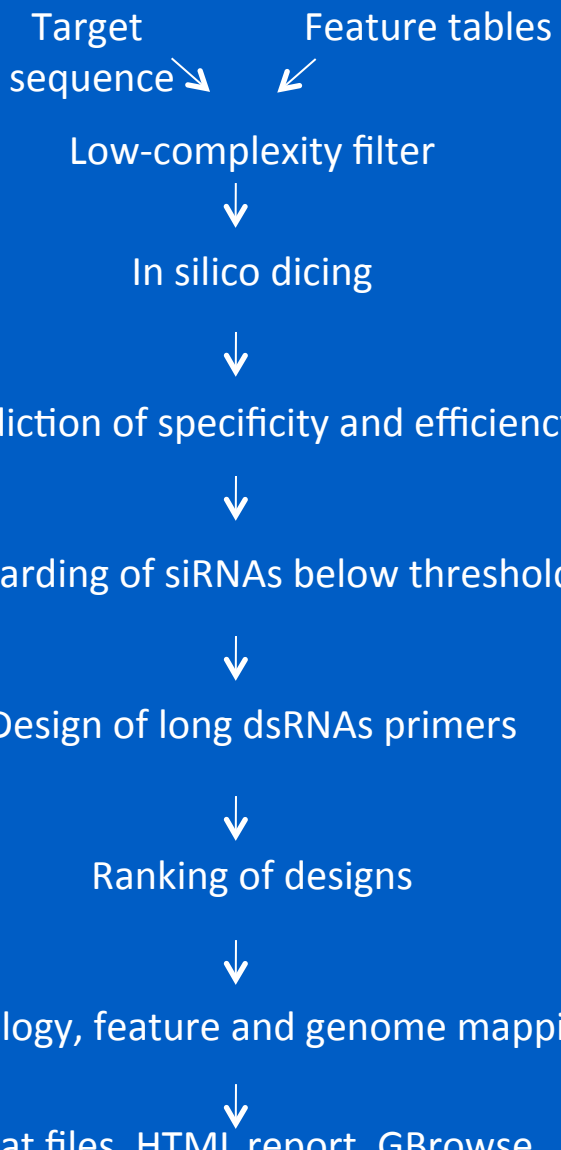


Automated design and evaluation of RNAi sequences on a genome wide scale

Thomas Horn, Thomas Sandmann and Michael Boutros.
Design and evaluation of genome-wide libraries for RNAi screens. Genome Biol. 2010 Jun 15;11(6):R61.

NEXT-RNAi Workflow

dkfz.



NEXT RNAi Input and dependencies

dkfz.

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

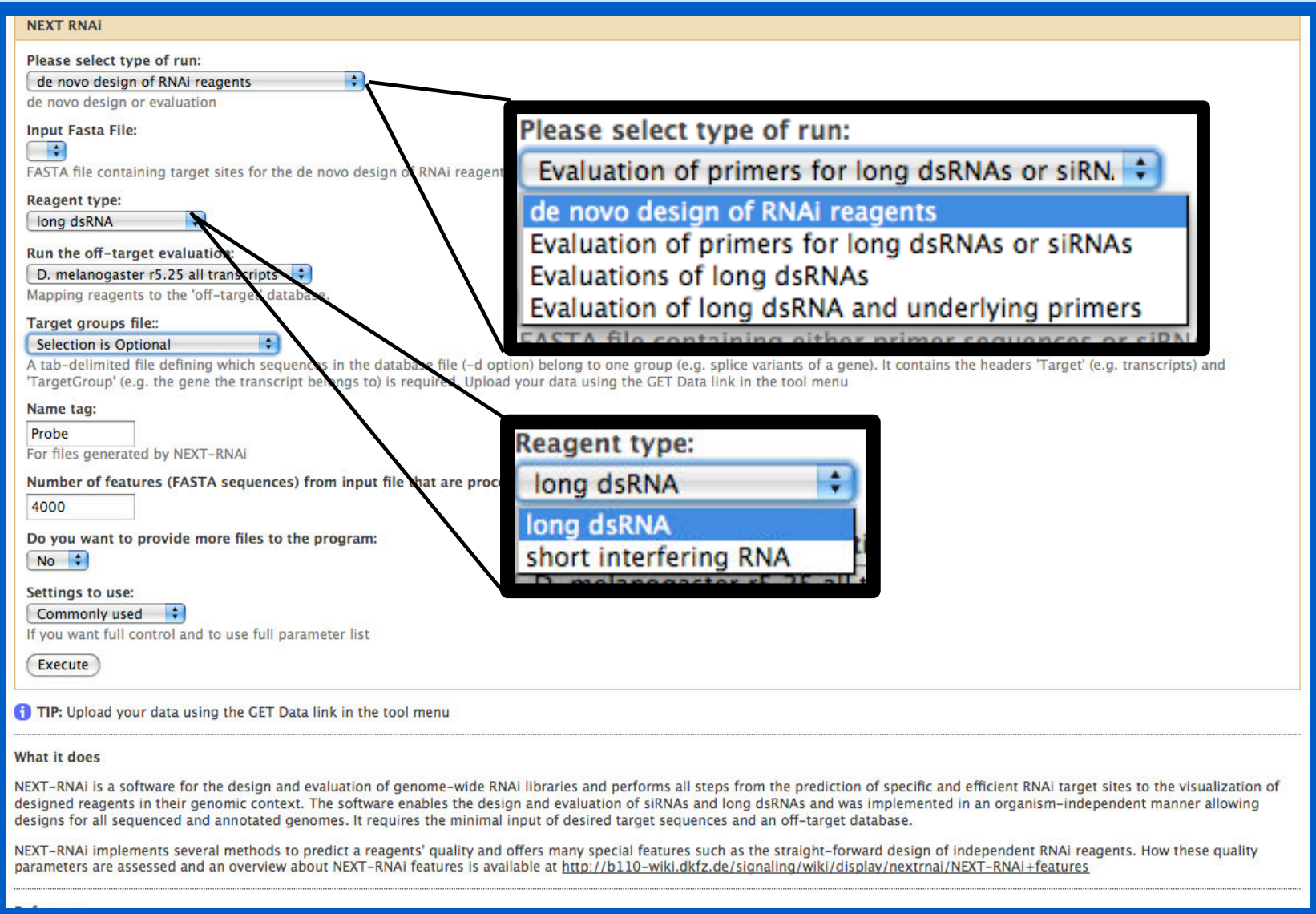
- Requirements:
  - Bowtie – sequence specificity
  - Primer3 – evaluate the primers
- Optional
  - BLAST
  - BLAT – align longer sequences
  - RNAfold – evaluate the 3d structure
  - Mdust – evaluate primer sequences

NEXT RNAi in Galaxy

```
nextrnai_wrapper.pl
        -s xxx
        -r xxx
        -d xxx
        -e xxx
        -n xxx
        --output xxx
        --inputFile xxx
        --targetgroups xxx
        --excluded xxx
        --intended xxx
        --feature xxx
        --seedmatch xxx
        --mirseed xxx
        --pool xxx
        --independent xxx
        --seedmatch xxx
        --sirnalength xxx
        --designwindow xxx
        --designnum xxx
        --outputnum xxx
        --primer3opt xxx
        --primertag xxx
        --efficiency xxx
        --targetseq xxx
        --lowcompeval xxx
        --caneval xxx
        --intron xxx
        --rankd xxx
        --redesign xxx
        --source xxx
        --genomebowtie xxx
        --blatprogram xxx
        --genomefasta xxx
        --blatsplit xxx
        --blathost xxx
        --blatport xxx
        --blatalign xxx
        --txnfasta xxx
        --gff xxx
        --gbrowsetrack xxx
        --aff xxx
        --gbrowsebase xxx
```

- Why Galaxy?
  - Number of input parameters
  - Number of dependent choices
  - Guidance of choices – conditional inputs in Galaxy
  - GUI - Not command line
  - Incorporation into workflows

NEXT RNAi in Galaxy



NEXT-RNAi Output

dkfz.

Links to result files

Tab-delimited result file
Statistics result file
FASTA result file
GFF result file
Annotations result file
Location(s) of mapped reagents
Oligo(s) that could not be mapped
dsRNAs that could not be mapped
Homology of RNAi reagents
miRNA seeds in RNAi reagents

Links to input text files

Database file used for off-target evalu
Database file used for mapping of rea
Reagent sequence input file (FASTA)
Validated reagent sequence input file (FA
Options input file
Targetgroups input file
Database file for homology evaluation

Links to output report files

Error log file
NEXT-RNAi report file
Failed design(s)

NEXT-RNAi results for nextrnaiTest design(s)

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

More statistics on designs are here

Links to HTML results

FBgn0038397_cr11 FBgn0038397_cr8
FBgn0034138_cr2 FBgn0003388_cr6
FBgn0032906_cr4 FBgn0005632_cr5
321_cr2
38397_cr
043_cr1
217_cr5
005_cr3
362_cr1
37895_cr
568_cr3
290_cr2
39207_cr
05632_cr
39207_cr6
674_cr1
860_cr3
233_cr2
233_cr4
825_cr1
187_cr1
109_cr2
1156_cr3

Statistics on overall 459 design(s)

Reagent statistics

Length forward primer [nt]: **20.25 +/- 1.31**
Length reverse primer [nt]: **20.36 +/- 1.43**
GC content forward primer [%]: **49.58 +/- 6.85**
GC content reverse primer [%]: **49.18 +/- 7.30**
Melting temperature forward primer [°C]: **60.00 +/- 0.80**
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

Query ID: FBgn0038397_cr11

Design 1: FBgn0038397_cr11_1

dsRNA information

Primer forward		Amplicon sequence
Sequence	GATGGGACCGAACTTTATCG	GATGGGACCGAACTTTATCGTGTTTCTGGGTCAGAAGTACGGCTATCGGC CCATTCCCACTTACATAGTGTCCTCTGAATTGGCCCTAATCTGCGAAGAA CTCACATCGATGGGAGTGGACCGTGCCATCCTCGATCTGTGGTACAAAAA GGACAGCAACGCAGTGCCACCCATCTCGGTGCTGCAGCCCATCTCATCGA TTTT
Length [nt]	20	
Tm[°C]	59.387	
GC[%]	50.000	
Primer reverse		**Amplicon length [nt]**
		204
Sequence	AAAATCGATGAGATGGGCTG	**Amplicon location**
Length [nt]	20	3R:11892654..11892857(+)
Tm[°C]	60.036	
GC[%]	45.000	

Primer pair penalty 0.6494

Target information

Intended target gene	FBgn0038397
Intended target transcripts (hits)	FBtr0083235 (186)
Other targeted gene(s)	NA
Other targeted transcripts (hits)	NA

Reagent quality

siRNAs [19 nt]	On-target	Off-target	No-target	mirSeed	Efficient siRNAs	Avg efficiency score	LowComplexRegions	CAN
186	186	0	0	0	186	50.77	0	0

Additional quality evaluation

Sequence homology (e-value) FBgn0038397(1e-112)

Genome Browser

NEXT RNAi in Galaxy Summary

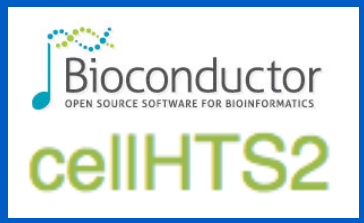
dkfz.

- 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

cellHTS

dkfz.

- Systematic analysis of screens
  - Phenotypic results range from single numerical value to multiple dimensional images
- Standardization of experimental information
- Standardization of analysis



Boutros, M., L. Bras, and W. Huber. (2006). Analysis of cell-based RNAi screens. *Genome Biology* 7:R66.

cellHTS

dkfz.

Import raw data files

↓

Per plate quality control

↓

Data normalization

↓

Scoring of phenotypes

↓

Annotation and analysis

cellHTS in Galaxy

dkfz.

cellHTS2

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:

Execute

TIP: Upload the data files of your experiment using ftp.

Vs.

```
# R code from vignette source 'vignettes/cellHTS2/inst/doc/cellhts2Complete.Rnw'
¬
 experimentName <- "KcViab"¬
 dataPath <- system.file(experimentName, package="cellHTS2") ¬
 rev(dir(dataPath))[1:12]¬
 x <- readPlateList("Platelist.txt", name=experimentName, path=dataPath)¬
 out <- writeReport(raw=x)¬
 out <- writeReport(raw=x, force=TRUE, outdir=tempdir())¬
 browseURL(out)¬
 x <- configure(x, descripFile="Description.txt", confFile="Plateconf.txt", ¬
Δ  Δ  Δ  Δ     logFile="Screenlog.txt", ¬
              path=dataPath)¬
 xn <- normalizePlates(x, ¬
                       scale="multiplicative", ¬
                       log=FALSE, ¬
                       method="median", ¬
                       varianceAdjust="none")¬
 xsc <- scoreReplicates(xn, sign="-", method="zscore") ¬
 xsc <- summarizeReplicates(xsc, summary="mean") ¬
 scores <- Data(xsc)¬
 ylim <- quantile(scores, c(0.001, 0.999), na.rm=TRUE)¬
 boxplot(scores ~ wellAnno(x), col="lightblue", outline=FALSE, ylim=ylim)¬
 y <- scores2calls(xsc, z0=1.5, lambda=2)¬
 png("cellhts2Complete-callvalues.png")¬
 plot(Data(xsc), Data(y), col="blue", pch=".",¬
      xlab="z-scores", ylab="calls", ¬
      main=expression(1/(1+e^{-lambda *(z-z[0])})))¬
 dev.off()¬
 xsc <- annotate(xsc, geneIDFile="GeneIDs_Dm_HFA_1.1.txt", path=dataPath)¬
 setSettings(list(plateList=list(reproducibility=list(include=TRUE, map=TRUE), ¬
                                 intensities=list(include=TRUE, map=TRUE)),¬
                  screenSummary=list(scores=list(range=c(-4, 8), map=TRUE))))¬
 out <- writeReport(raw=x, normalized=xn, scored=xsc, force=TRUE)¬
 writeTab(xsc, file="Scores.txt")¬
```

cellHTS in Galaxy

dkfz.

```
Wells: 384
Plates: 26
Plate   Well    Content
*   *   sample
*   A01 neg
*   A02 neg
*   B01 pos
*   B02 pos
```

```
RA01D2  A01 2390
RA01D1  A01 2300
RA02D1  A01 3040
RA02D2  A01 2430
RA03D1  A01 1020
RA03D1  A02 1030
RA03D1  A03 1450
RA03D1  A04 210
RA03D1  A05 1150
RA03D1  A06 850
RA03D1  A07 360
RA03D1  A08 430
RA03D1  A09 480
RA03D1  A10 350
RA03D1  A11 830
RA03D1  A12 310
RA03D1  A13 410
RA03D1  A14 380
RA03D1  A15 550
```

Filename	Plate	Replicate	Channel
RB01D1.TXT	1	1	2
RA02D1.TXT	2	1	1
RB02D1.TXT	2	1	2
RA03D2.TXT	3	2	1
RB01D2.TXT	1	2	2
RA01D2.TXT	1	2	1
RB03D1.TXT	3	1	2
RA03D1.TXT	3	1	1
RA02D2.TXT	2	2	1
RC03D2.TXT	3	2	2
RA01D1.TXT	1	1	1
RB02D2.TXT	2	2	2

cellHTS2

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

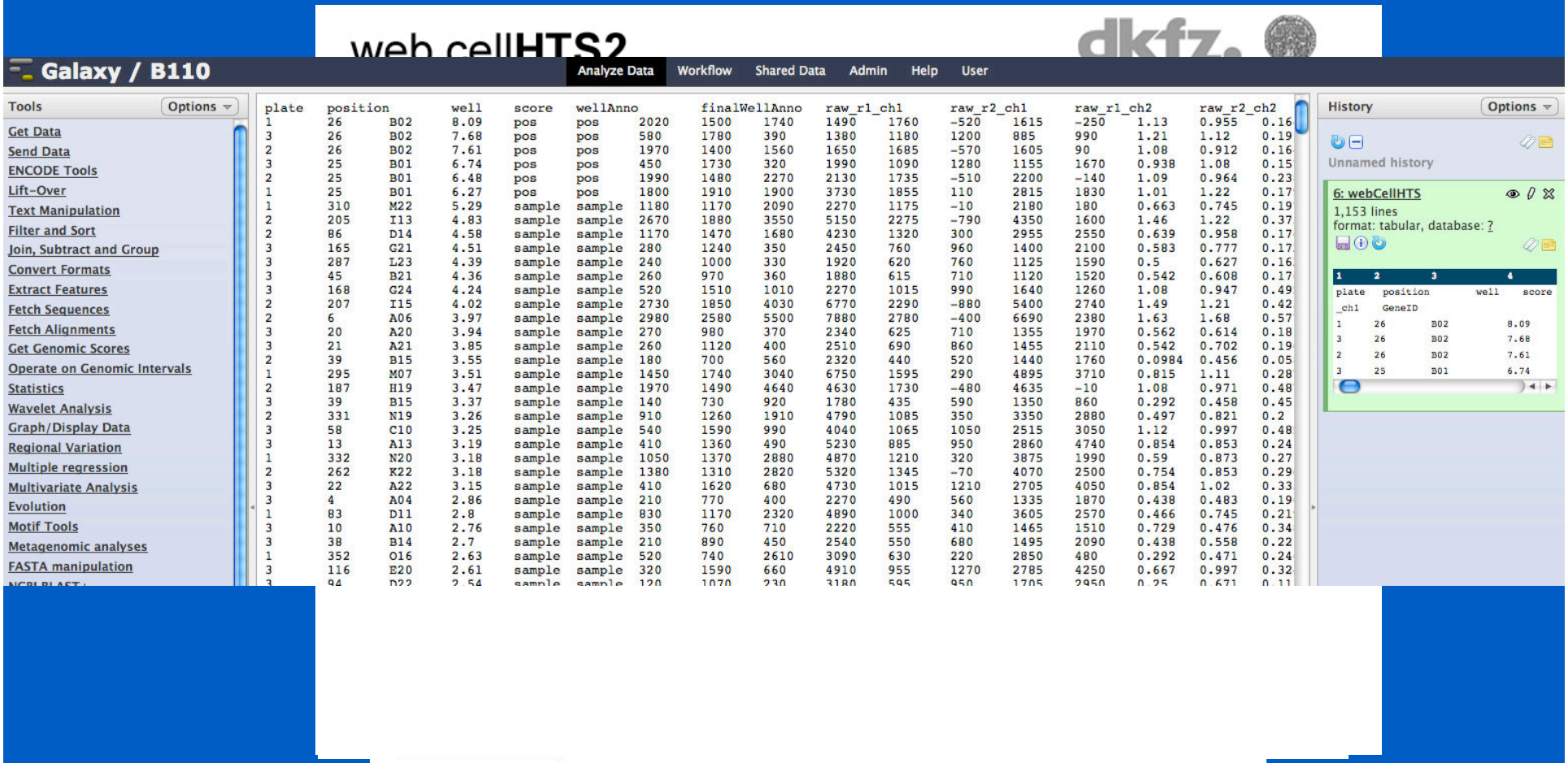
Please enter the (exact) file name containing the GeneIDs:

Execute

TIP: Upload the data files of your experiment using ftp.

webcellHTS

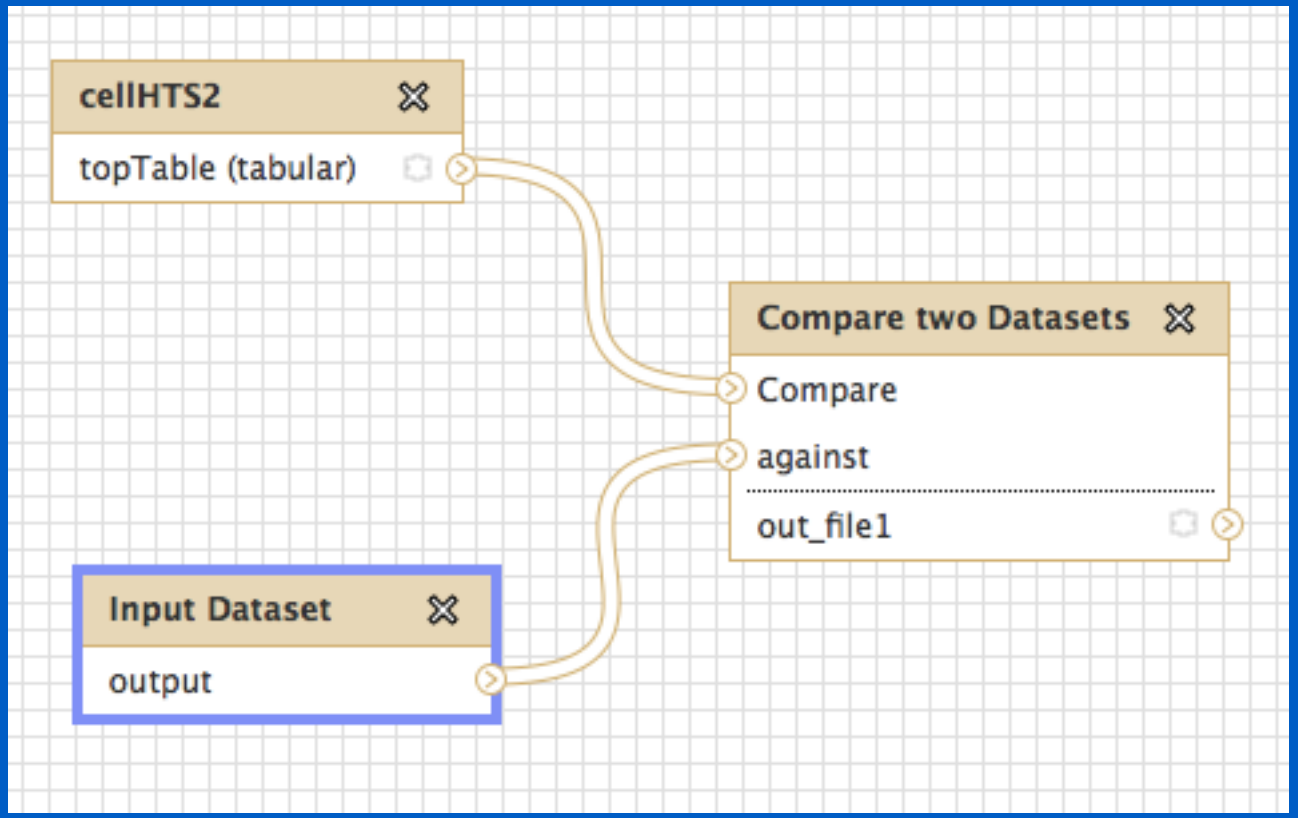
- Web based interactive guide to create input files.



HTS Workflows

dkfz.

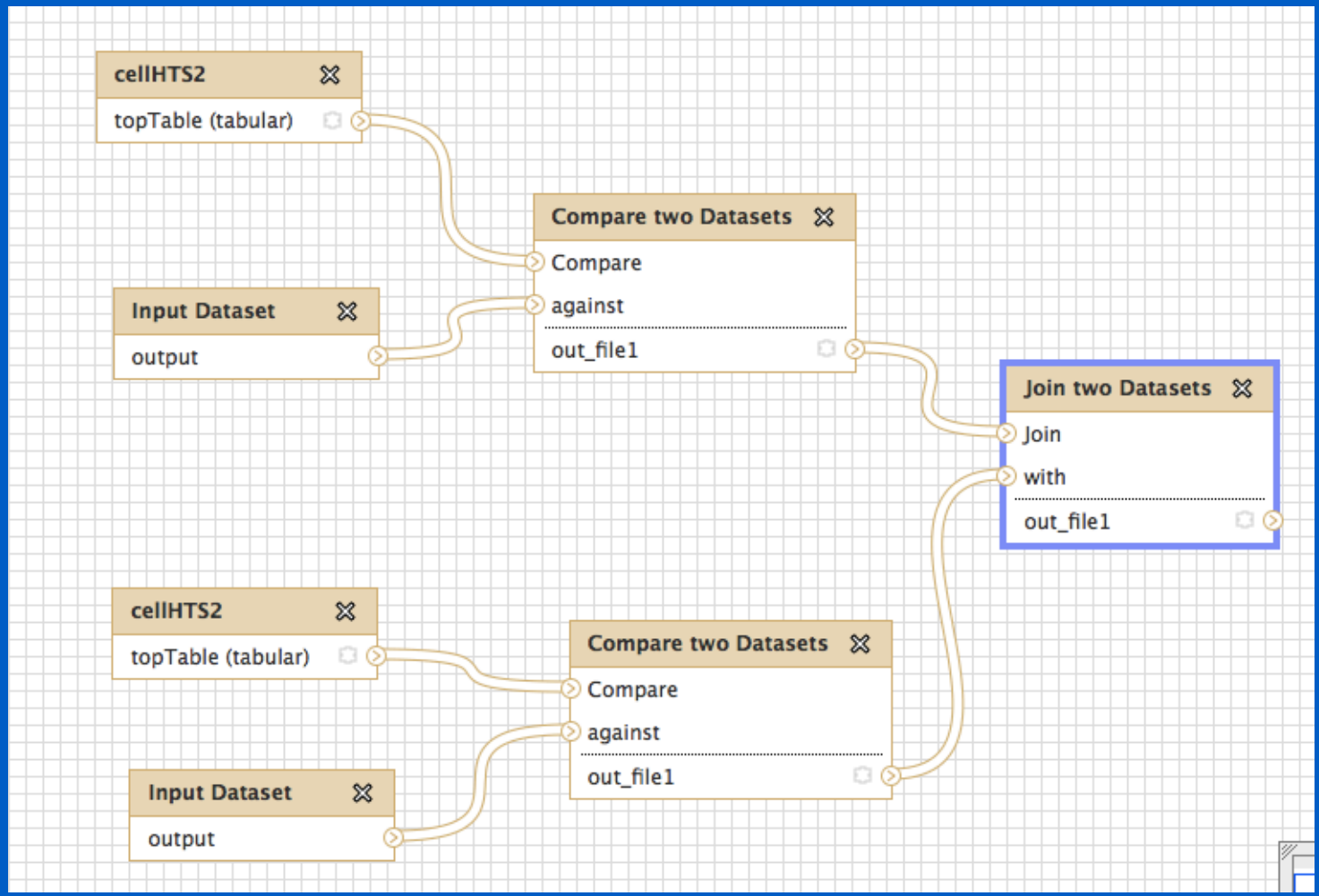
- Annotation Libraries storage



HTS Workflows

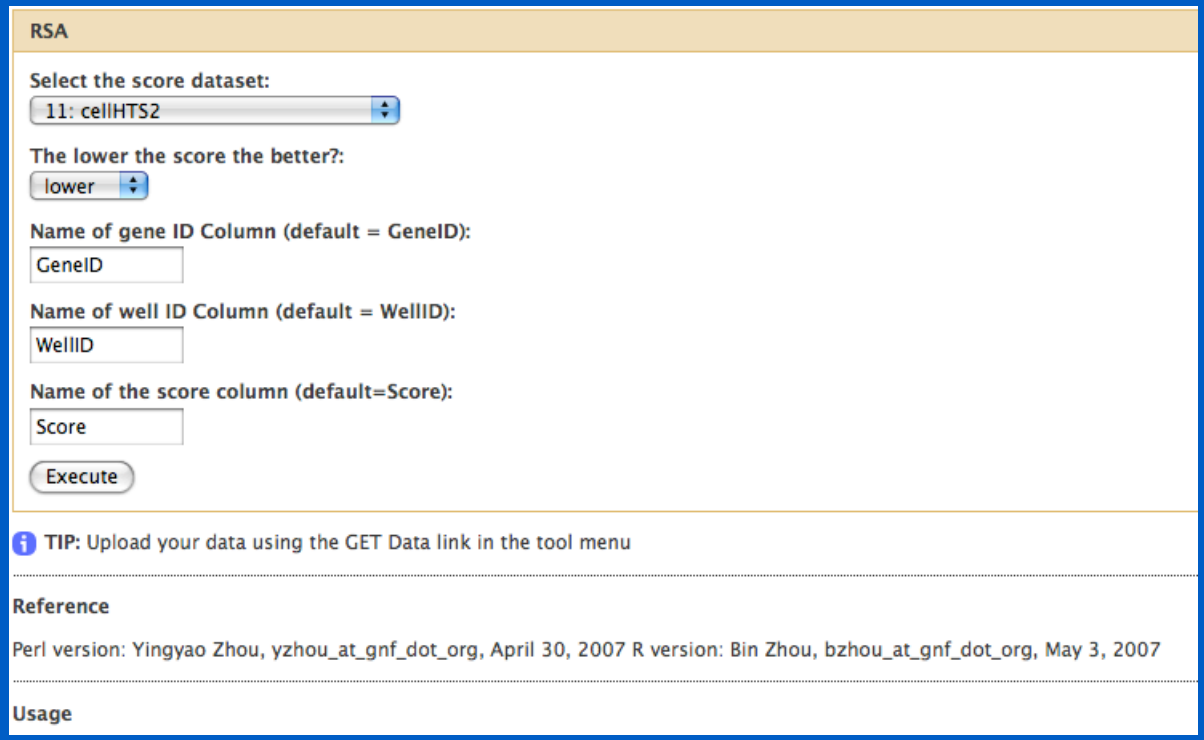
dkfz.

- Standardizing analysis



HTS Workflows

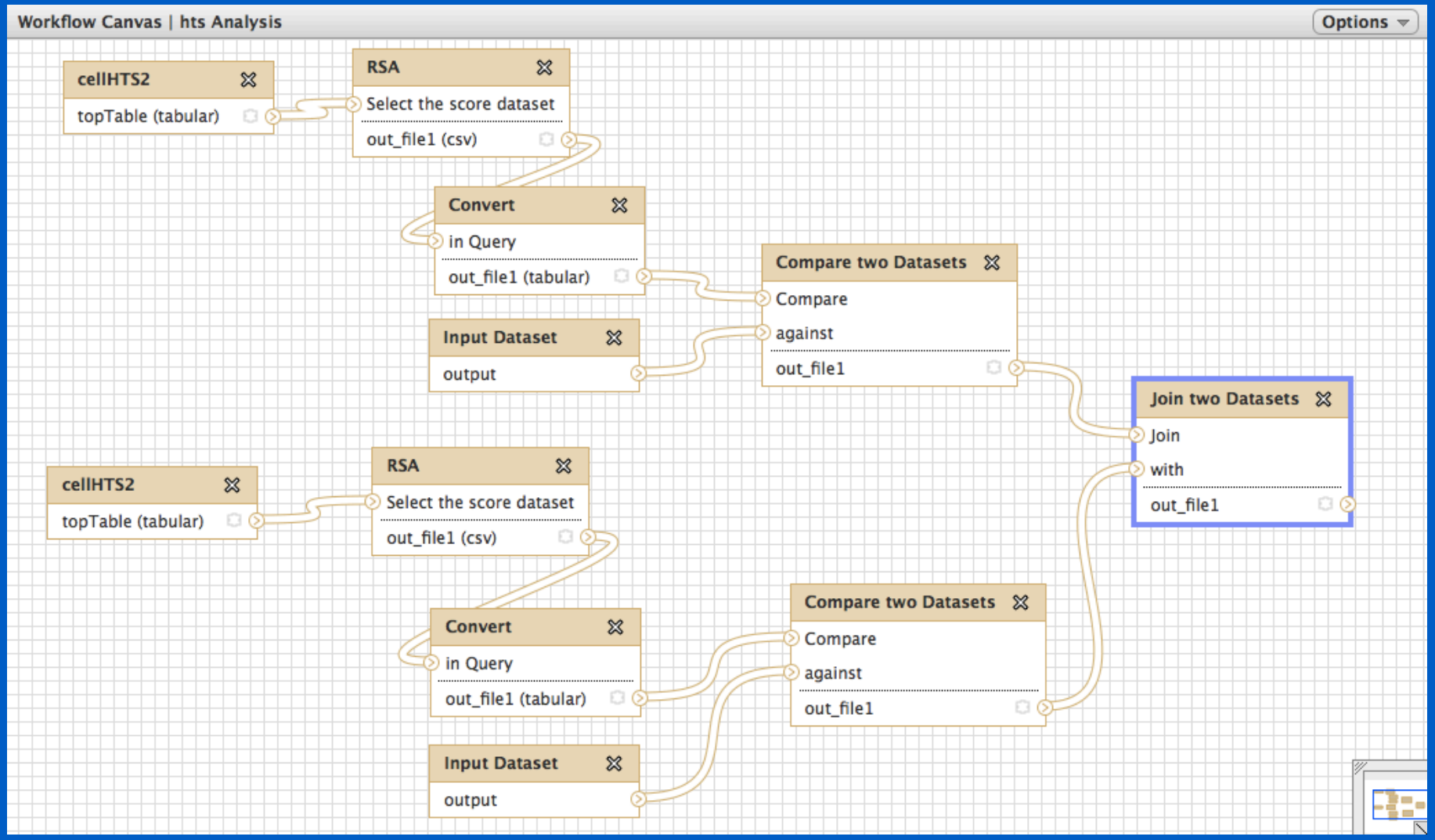
- Incorporation with other specific screening analysis tools e.g. Redundant siRNA activity analysis (RSA)



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

HTS Workflows

dkfz.



HTS Workflows

dkfz.

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

Step	Annotation
Step 1: Input dataset **Input Dataset of Selected KK Lines** *select at runtime*	Input your dataset of selected KK lines
Step 2: Input dataset **Input KK Library Human Homolog Dataset** *select at runtime*	Input the following file from the Shared Data Library In Vivo Drosophila Screens KK_TrID_BKN_CG_FBgn_Symbol_Hom_Penninger_Ensembl_EnsemblPot.txt
Step 3: Join two Datasets **Join** Output dataset 'output' from step 1 **using column** **with** Output dataset 'output' from step 2 **and column** 1 (value not yet validated) **Keep lines of first input that do not join with second input** Yes **Keep lines of first input that are incomplete** Yes **Fill empty columns** Yes **Only fill unjoined rows** Yes **Fill Columns by** Single fill value **Fill value**	Please specify the transformant ID column (e.g. c1) for the joining

Outlook

- 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