

Galaxy and A High Throughput Screening Lab

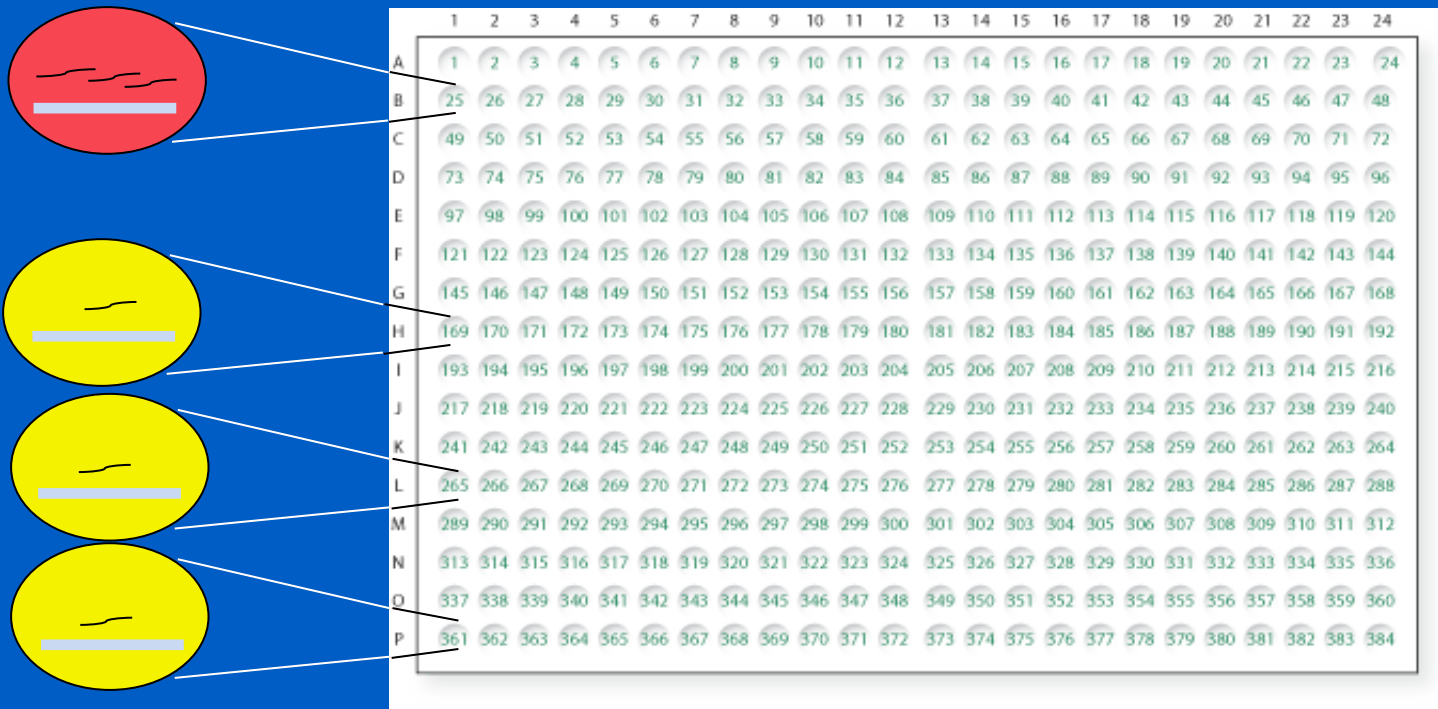


dkfz.

**GERMAN
CANCER RESEARCH CENTER
IN THE HELMHOLTZ ASSOCIATION**

High Throughput Screening Lab?

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



RNA interference (RNAi)

dkfz.

dsRNA (>27 nt)

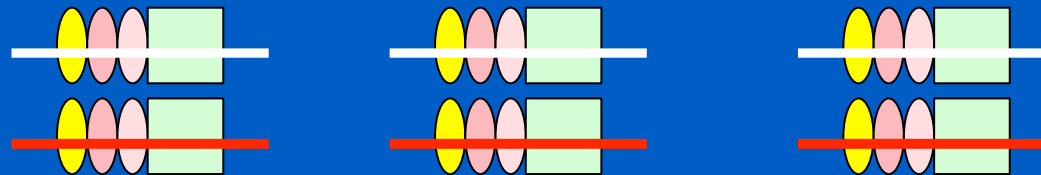


Dicer

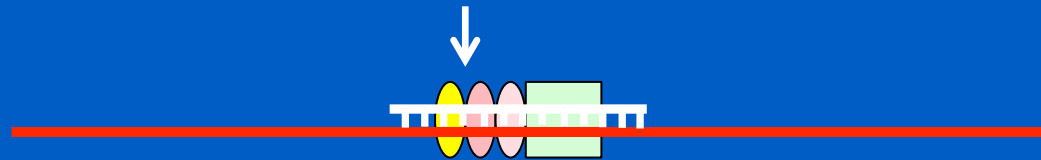
siRNA (18 – 21 nt)



RISCs



RISC targets homologous mRNA



Targeted degradation

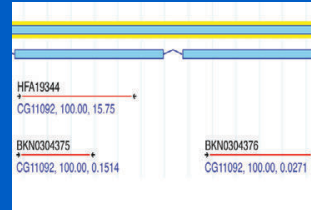


 Galaxy

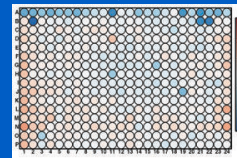


Experimental Steps

RNAi Library Design



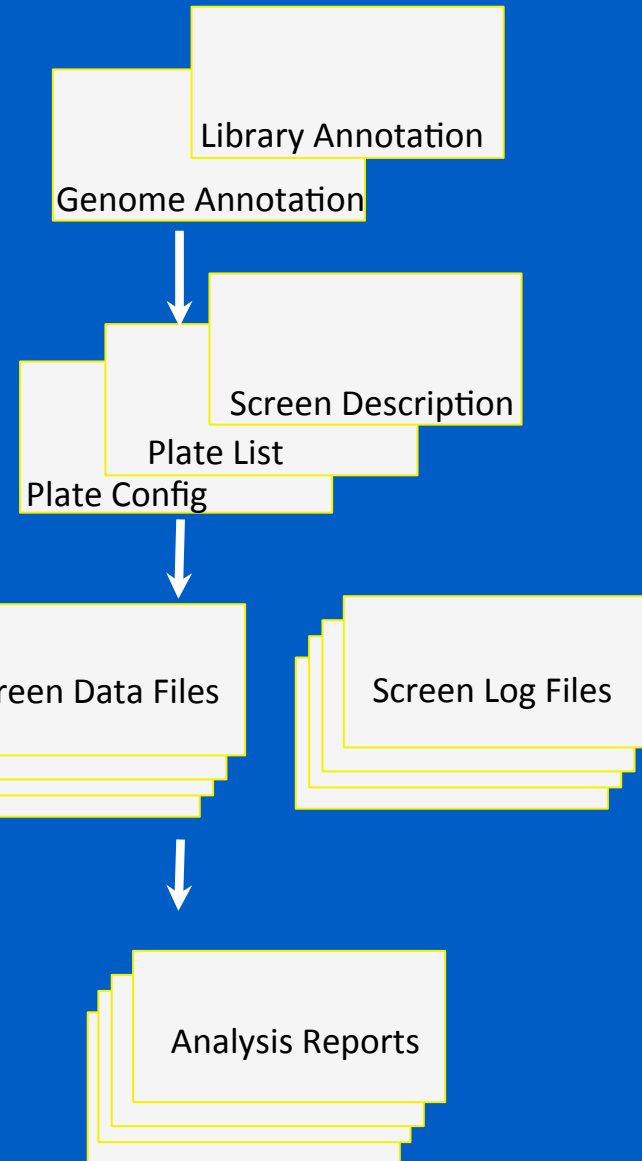
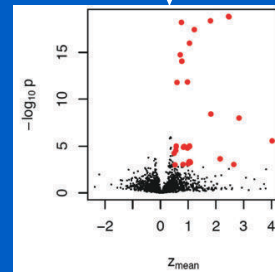
Cell based assay format



Large scale experiment

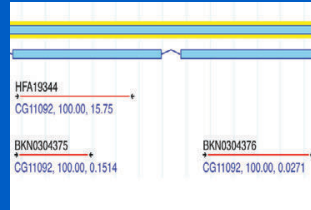


Computational analysis

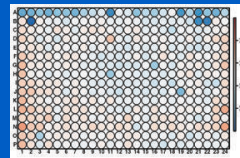


Experimental Steps

RNAi Library Design



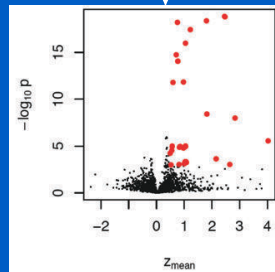
Cell based assay format



Large scale experiment



Computational analysis



Library Annotation

Genome Annotation

Screen Description

Plate List

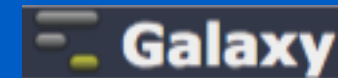
Plate Config

Screen Data Files

Screen Log Files

Analysis Reports

- Design of reagents

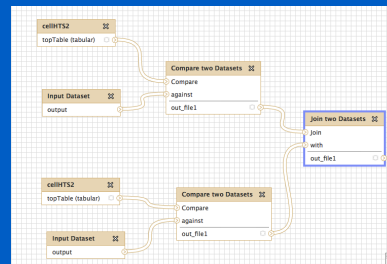


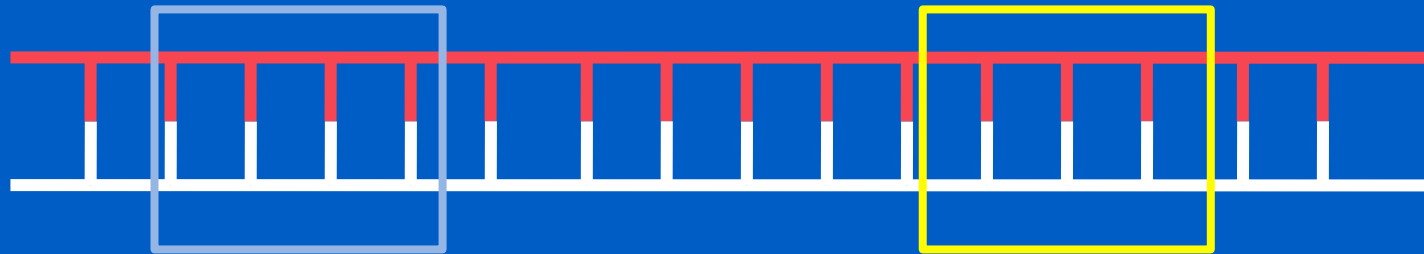
- Analysis of the read-out data

web cellHTS2



- Selecting and analysis of “hit” genes





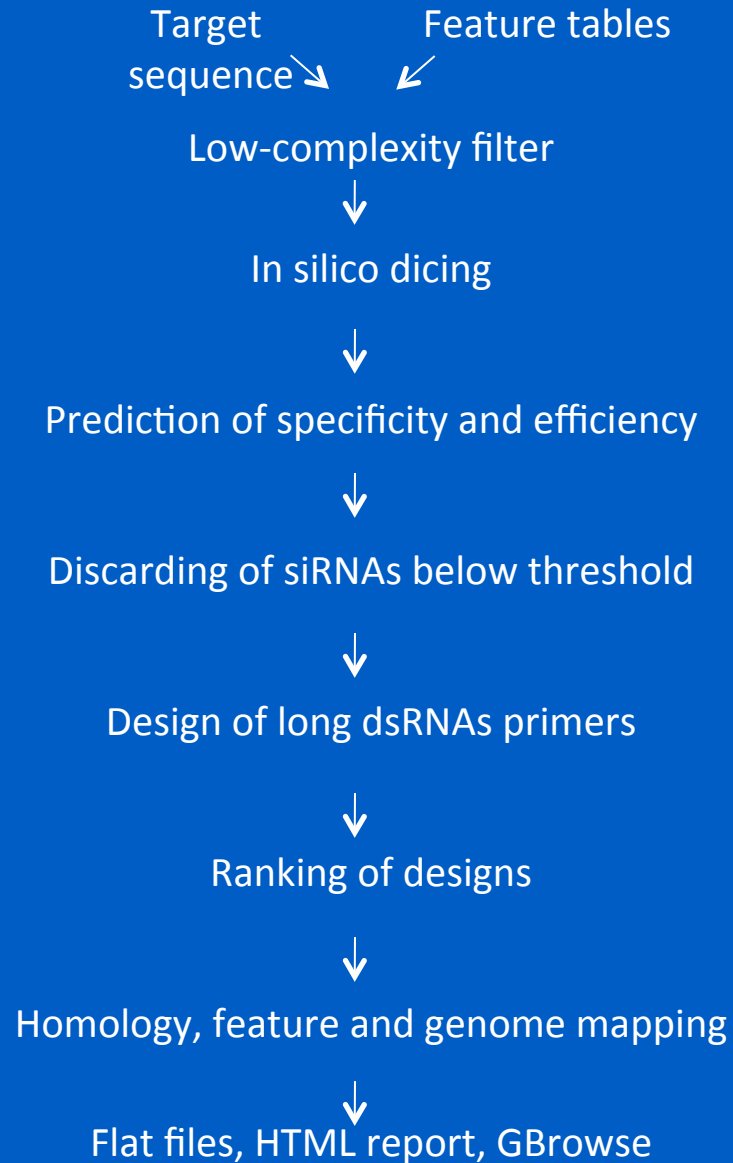
Automated design and evaluation of RNAi sequences
on a genome wide scale

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.

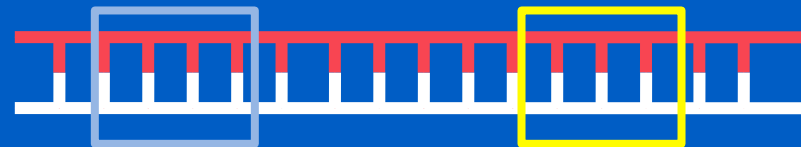


Target Sequences

Target Group Files

Annotation

nextRNAi



Bowtie

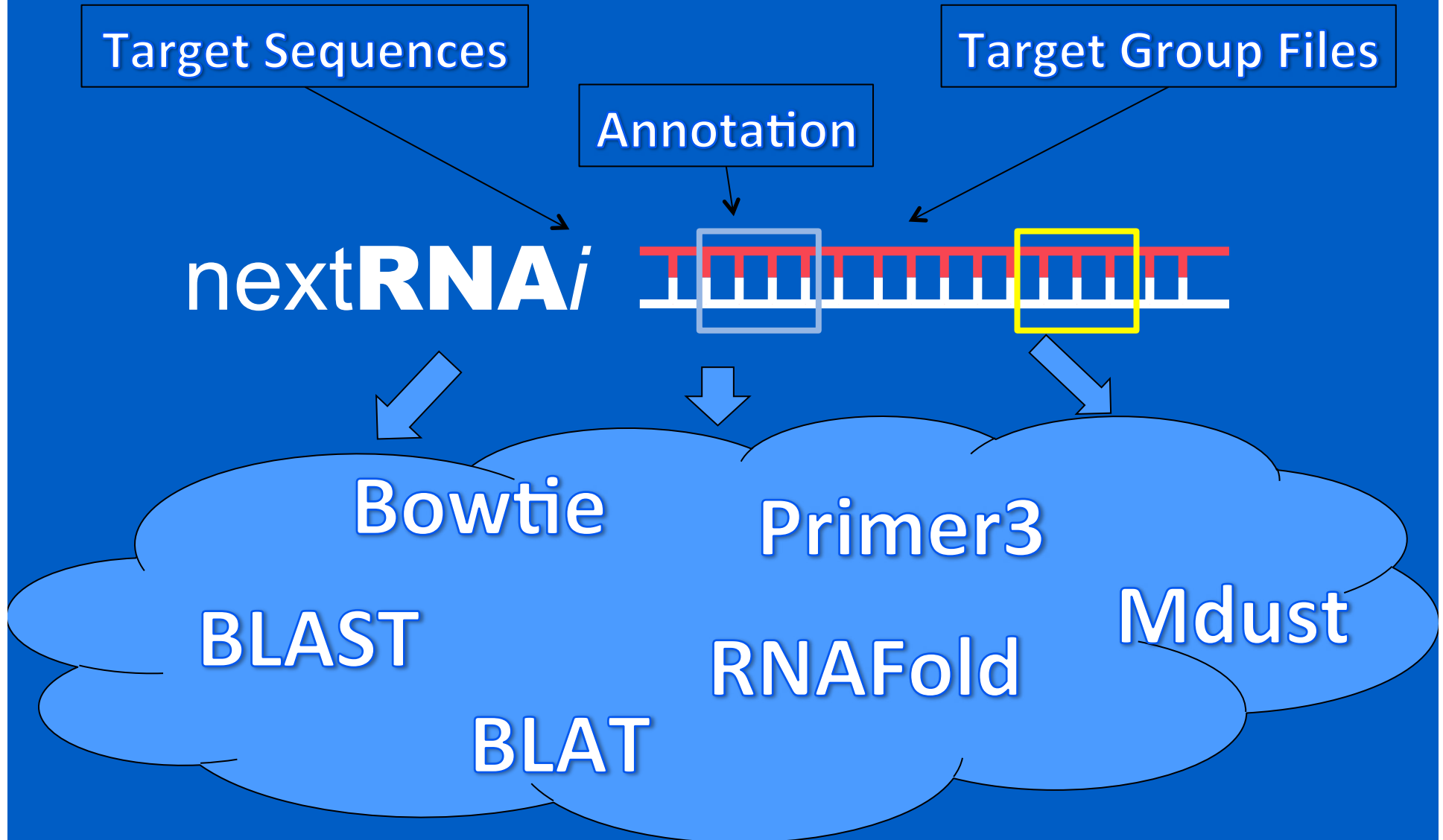
Primer3

BLAST

RNAFold

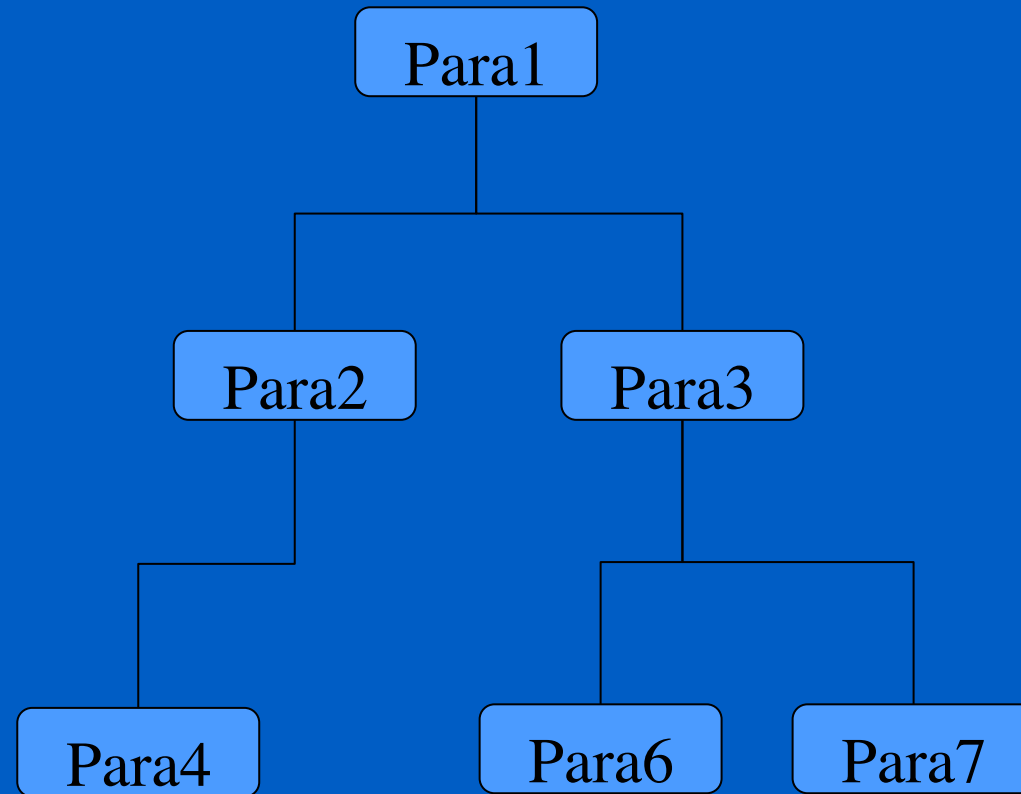
Mdust

BLAT



NEXT RNAi in Galaxy

```
nextrna_i_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 -  
--signalength 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 -  
--gbrowse track xxx -  
--aff xxx -  
--gbrowsebase xxx -
```



NEXT RNAi in Galaxy

NEXT RNAi

Please select type of run:
de novo design of RNAi reagents

de novo design or evaluation

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

Reagent type:
long dsRNA

Run the off-target evaluation:
D. melanogaster r5.25 all transcripts

Mapping reagents to the 'off-target' database

Target groups file:
Selection is Optional

A tab-delimited file defining which sequences in the database file (-d option) belong to one group (e.g. splice variants of a gene). It contains the headers 'Target' (e.g. transcripts) and 'TargetGroup' (e.g. the gene the transcript belongs to) is required. Upload your data using the GET Data link in the tool menu

Name tag:
Probe

For files generated by NEXT-RNAi

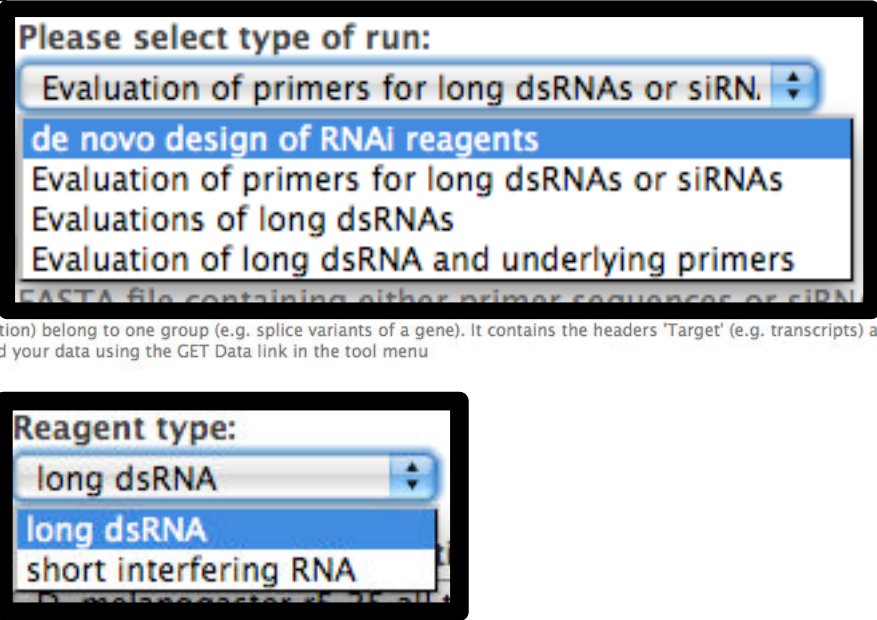
Number of features (FASTA sequences) from input file that are processed:
4000

Do you want to provide more files to the program:
No

Settings to use:
Commonly used

If you want full control and to use full parameter list

Execute



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://b110-wiki.dkfz.de/signaling/wiki/display/nextRNAi/NEXT-RNAi+features>

NEXT-RNAi Output



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 evaluation
- Database file used for mapping of reagents
- Reagent sequence input file (FASTA)
- Validated reagent sequence input file (FASTA)
- 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)

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

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

- [FBgn0038397_cr11](#) [FBgn0038397_cr8](#)
- [FBgn0034138_cr2](#) [FBgn0003388_cr6](#)
- [FBgn0032906_cr4](#) [FBgn0005632_cr5](#)

Query ID: FBgn0038397_cr11

Design 1: FBgn0038397_cr11_1

dsRNA information

Primer forward
 Sequence GATGGGACCGAAGCTTTATCG
 Length [nt] 20
 Tm[°C] 59.387
 GC[%] 50.000
Primer reverse
 Sequence AAAATCGATGAGATGGGCTG
 Length [nt] 20
 Tm[°C] 60.036
 GC[%] 45.000
 Primer pair penalty 0.6494

Amplicon sequence
 GATGGGACCGAAGCTTTATCGTTCTGGGTGAGAAGTACGGCTATCGGC
 CCATTTCCACTTACATAGTGCCTCTGAATGGCCCTAATCGCGAAGAA
 CTCACATCGATGGGATGGAGCCCTGCCATCTCGATCTGGTACAAAA
 GGACAGCAACGAGTGCACCCATCTCGGTGCTGACGCCATCTCATCGA
 TTTT

Amplicon length [nt]
 204

Amplicon location
 3R:11892654..11892857(+)

Target information

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

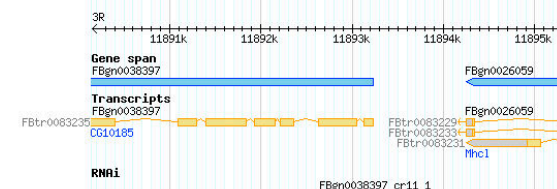
Reagent quality

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\(1e-112\)](#)

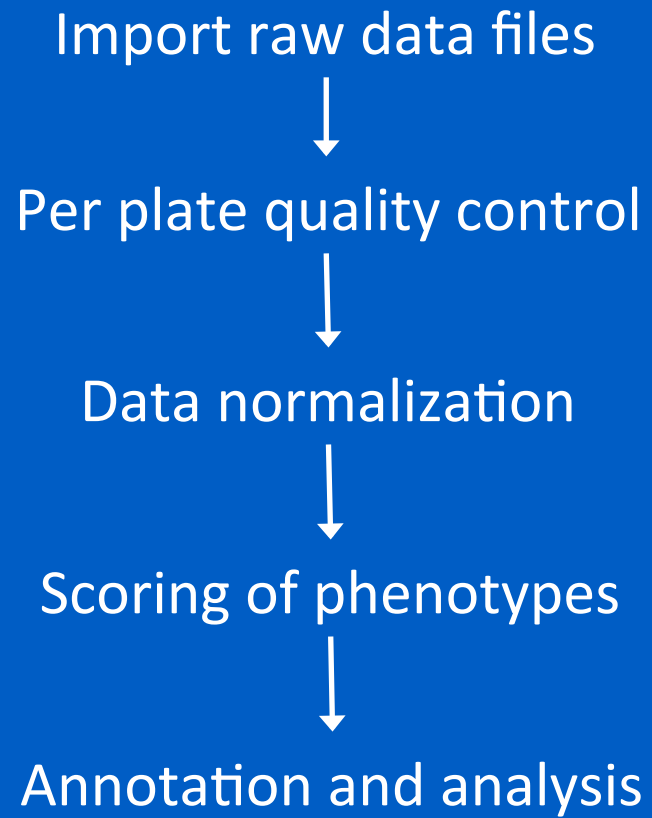
Genome Browser



- Systematic analysis of screens
- 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.



cellHTS2

Please enter the name of your experiment:

Apply variance Adjustment?:

Scale the data?:

Is the data log transformed?:

Apply Normalization?:

Which method to score values?:

Which method to summarise replicates?:

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

Execute

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(row=x)-
out <- writeReport(row=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-calls.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(row=x, normalized=xn, scored=xsc, force=TRUE)-
writeTab(xsc, file="Scores.txt")-
```

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

cellHTS in Galaxy

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

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

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?:

Scale the data?:

Is the data log transformed?:

Apply Normalization?:

Which method to score values?:

Which method to summarise replicates?:

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

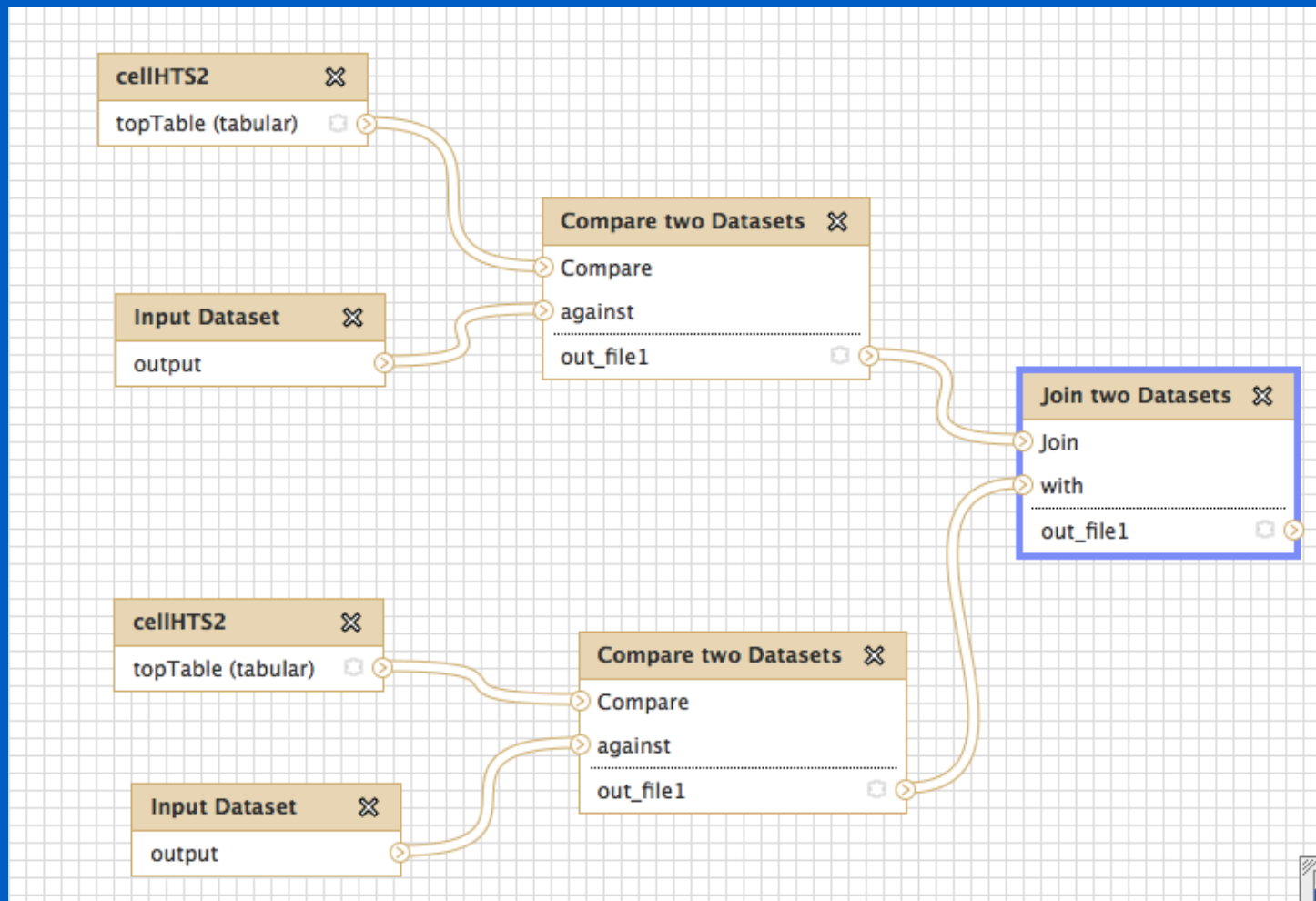
TIP: Upload the data files of your experiment using ftp.

- Web based interactive guide to create input files.

The screenshot displays the webcellHTS2 interface within a Galaxy environment. The main area shows a table with 14 columns: plate, position, well, score, wellAnno, finalWellAnno, raw_r1_ch1, raw_r2_ch1, raw_r1_ch2, and raw_r2_ch2. The table contains 1153 rows of data. On the left, there is a 'Tools' sidebar with various options like 'Get Data', 'Send Data', and 'Text Manipulation'. On the right, a 'History' panel shows the current job '6: webCellHTS' with 1,153 lines of output in a tabular format.

plate	position	well	score	wellAnno	finalWellAnno	raw_r1_ch1	raw_r2_ch1	raw_r1_ch2	raw_r2_ch2
1	26	B02	8.09	pos	2020	1500	1740	1490	1760
3	26	B02	7.68	pos	580	1780	390	1380	1180
2	26	B02	7.61	pos	1970	1400	1560	1650	1685
3	25	B01	6.74	pos	450	1730	320	1990	1090
2	25	B01	6.48	pos	1990	1480	2270	2130	1735
1	25	B01	6.27	pos	1800	1910	1900	3730	1855
1	310	M22	5.29	sample	1180	1170	2090	2270	1175
2	205	I13	4.83	sample	2670	1880	3550	5150	2275
2	86	D14	4.58	sample	1170	1470	1680	4230	1320
3	165	G21	4.51	sample	280	1240	350	2450	760
3	287	L23	4.39	sample	240	1000	330	1920	620
3	45	B21	4.36	sample	260	970	360	1880	615
3	168	G24	4.24	sample	520	1510	1010	2270	1015
2	207	I15	4.02	sample	2730	1850	4030	6770	2290
2	6	A06	3.97	sample	2980	2580	5500	7880	2780
3	20	A20	3.94	sample	270	980	370	2340	625
3	21	A21	3.85	sample	260	1120	400	2510	690
2	39	B15	3.55	sample	180	700	560	2320	440
1	295	M07	3.51	sample	1450	1740	3040	6750	1595
2	187	H19	3.47	sample	1970	1490	4640	4630	1730
3	39	B15	3.37	sample	140	730	920	1780	435
2	331	N19	3.26	sample	910	1260	1910	4790	1085
3	58	C10	3.25	sample	540	1590	990	4040	1065
3	13	A13	3.19	sample	410	1360	490	5230	885
1	332	N20	3.18	sample	1050	1370	2880	4870	1210
2	262	K22	3.18	sample	1380	1310	2820	5320	1345
3	22	A22	3.15	sample	410	1620	680	4730	1015
3	4	A04	2.86	sample	210	770	400	2270	490
1	83	D11	2.8	sample	830	1170	2320	4890	1000
3	10	A10	2.76	sample	350	760	710	2220	555
3	38	B14	2.7	sample	210	890	450	2540	550
1	352	O16	2.63	sample	520	740	2610	3090	630
3	116	E20	2.61	sample	320	1590	660	4910	955
3	94	D22	2.54	sample	120	1070	220	3180	595

Keeping it Simple



Accessing Information

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

Step

Step 1: Input dataset

Input Dataset of Selected KK Lines
select at runtime

Step 2: Input dataset

Input KK Library Human Homolog Dataset
select at runtime

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

Annotation

Input your dataset of selected KK lines

Input the following file from the Shared Data Library In Vivo Drosophila Screens
KK_TriD_BKN_CG_FBgn_Symbol_Hom_Penninger_Ensembl_EnsemblPot.txt

Please specify the transformant ID column (e.g. c1) for the joining

Tools for RNAi are more accessible

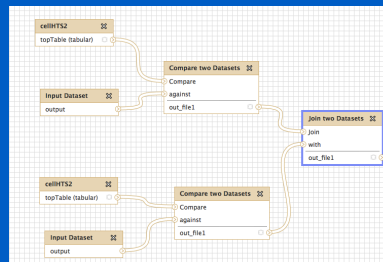
- Rapid design and evaluation of RNAi libraries



- Analysis of screen readout

web cellHTS2

- Day – to – day file manipulation tasks



Acknowledgements



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