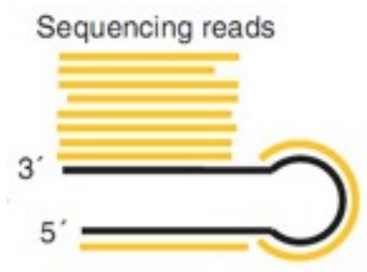
DARIO

A free web server for the analysis of short RNAs
from high throughput sequencing data

David Langenberger & Mario Fasold

Introduction


```
GCGGACGGGAGCTGAGAGCTGGGTCTTTGCGGGCAAGATGAGGGTGTCAGTTCAACTGGCCTACAAAGTCCCAGTCCTCGGCTCCC
......((((((((((..(((((.(((((..(((.((.((((........)))).)).)))..))))))..)))))..))))))))))
                     TGGGTCTTTGCGGGCAA 1                AACTGGCCTACAAAGTC 73
                                                        ACTGGCCTACAAAGTCC 2
                                                        AACTGGCTTACAAAGTC 1
                                                        AACTGGCCTACAGAGTC 1
                                                        AACTGGCCTACAATGTC 1
                                                        AACTGGCCTACAAAGTT 1
                                                        AACTGGCCTACAAATTC 1
```



A novel class of small RNAs: tRNA-derived RNA fragments (tRFs)

Yong
Depar

New
organ
prost
seque
of tR
short
splin
impo
trans
but n
medi
phen
resist
susce
an ab
and s

[Keyw

Suppl

Recei

GCGGACGGGAGCTG
. . . . . ((((((((

Molecular Cell
Article

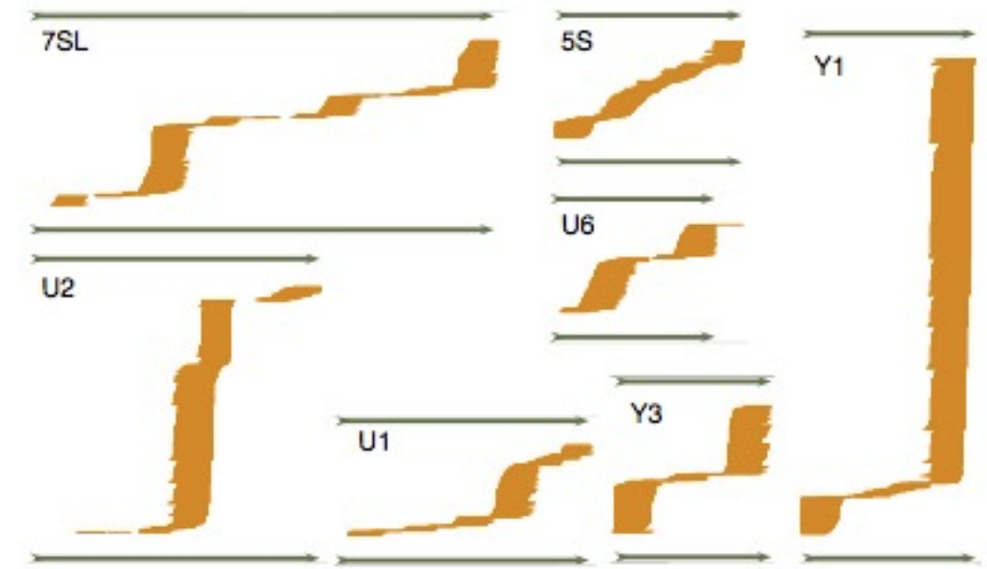
Cell PRESS

A Human snoRNA with MicroRNA-Like Functions

Christine Ender,[1,6] Azra Krek,[2,3,6] Marc R. Friedländer,[2] Mich
Sébastien Pfeffer,[5] Nikolaus Rajewsky,[2,*] and Gunter Meiste
[1]Center for Integrated Protein Science Munich (CIPSM), Laboratory o
Am Klopferspitz 18, 82152 Martinsried, Germany
[2]Max Delbrück Center for Molecular Medicine, Robert-Rössle-Strass
[3]Department of Physics, New York University, 4 Washington Place, N
[4]Department of Human Molecular Genetics, Max Planck Institute of M
[5]IBMP-CNRS, 12 Rue du General Zimmer, 67084 Strasbourg Cedex,
[6]These authors contributed equally to this work
*Correspondence: rajewsky@mdc-berlin.de (N.R.), meister@biochem.
DOI 10.1016/j.molcel.2008.10.017

SUMMARY

Small noncoding RNAs function in concert with Argonaute (Ago) proteins to regulate gene expression at the level of transcription, mRNA stability, or translation. Ago proteins bind small RNAs and form the core of silencing complexes. Here, we report the analysis of small RNAs associated with human Ago1 and Ago2 revealed by immunoprecipitation and deep sequencing. Among the reads, we find small RNAs originating from the small nucleolar RNA (snoRNA)



Grishok et al., 2001; Hutvágner et al., 2001; Lund et al., 2004). Such dsRNA intermediates are subsequently unwound, and the single-stranded mature miRNA is incorporated into effector complexes often referred to as miRNPs (Mourelatos et al., 2002). In the siRNA pathway or RNA interference (RNAi), long

AACTGGCCTACAAAGTT 1
AACTGGCCTACAAATTC 1

High Throughput Sequencing of small RNAs

Different High Throughput Sequencing platforms

Illumina Genome Analyzer IIx (http://www.illumina.com)

Genome Sequencer FLX system (http://www.454.com/)

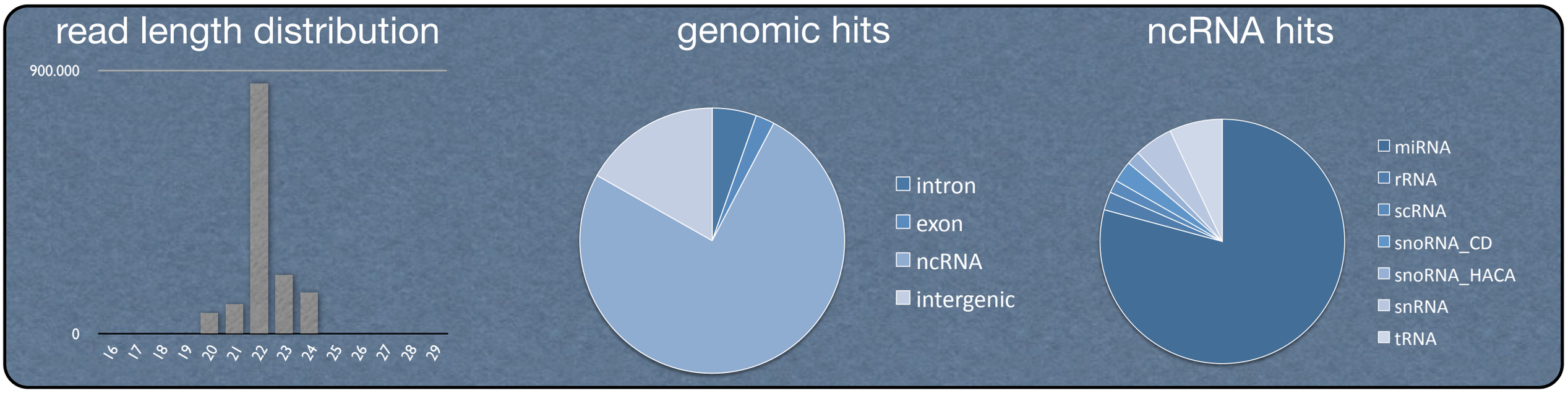
SOLiD™ Sequencers (http://www.appliedbiosystems.com/)

HUGE amount of Data

```
@HWI-EAS385:1:1:151:949#0/1
TAGCTTATCAGACTGATGGTNATAGTNTGACGTCNTCTGCT
+HWI-EAS385:1:1:151:949#0/1
bSLSbaa_OXaSTYX[_^M^DL^N_ND[`FM^[YD^[_QI\
@HWI-EAS385:1:1:154:1082#0/1
TAGCTTATAAGACTGATGTTNATTGANTGACTTCNTTTTCT
+HWI-EAS385:1:1:154:1082#0/1
^N^`_ab]H\``ba]Y]_V\D\`BBBBBBBBBBBBBBBBB
@HWI-EAS385:1:1:156:983#0/1
TAGCTTATAAGACTGATGTTNATTGCNTCGCGTCNTCTGCT
+HWI-EAS385:1:1:156:983#0/1
^NR]ab`^D^`_ba\_^aX]DZ]D[HD^QD]\X[D]bbab^
@HWI-EAS385:1:1:158:1558#0/1
TTCAAGTATTCAAACTAAGANGTCCGNACTACGGNGNAGAC
+HWI-EAS385:1:1:158:1558#0/1
b^^T]`bb[a]DN\YXXaBBBBBBBBBBBBBBBBBBBBB
@HWI-EAS385:1:1:159:837#0/1
TCCTAGTCCGCGTACTGCTANCAAAANAAAACAGNAGAAAC
+HWI-EAS385:1:1:159:837#0/1
aPUGZD\YBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBB
@HWI-EAS385:1:1:159:1132#0/1
TAGCATTCCGGATACTGCCTNAAAAANAAAAGTCNTCTGCT
+HWI-EAS385:1:1:159:1132#0/1
b[[PbXU``\JH]Paa\^DXBBBBBBBBBBBBBBBBBBBBB
@HWI-EAS385:1:1:163:1210#0/1
TCGATTATCAGAATGAAGTTNATACGNTGCAGGCNTCTGAC
+HWI-EAS385:1:1:163:1210#0/1
baZOZ\BBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBB
@HWI-EAS385:1:1:167:928#0/1
TAGCTTATAAGACTGATGTTNATAGTNTGACGTCNTCTGCT
+HWI-EAS385:1:1:167:928#0/1
aX\a`a`b]bb^^aZ\a`Z^D[W\[YD]^[bba[D^aaYa]
@HWI-EAS385:1:1:167:1382#0/1
TCCTACTCCGTGTTCTGCTGNGAAAANAAAACACNCTAATA
+HWI-EAS385:1:1:167:1382#0/1
^DLSbVbb_bXD]Wbb_`SBBBBBBBBBBBBBBBBBBBBB
@HWI-EAS385:1:1:174:342#0/1
ACCAAGTCCGCTTACTGCTANCACAANACAACAGNANAAAC
+HWI-EAS385:1:1:174:342#0/1
JKVDMDQ^BBBBBBBBBBBBBBBBBBBBBBBBBBBBBBB
```

Gigabytes of short read data

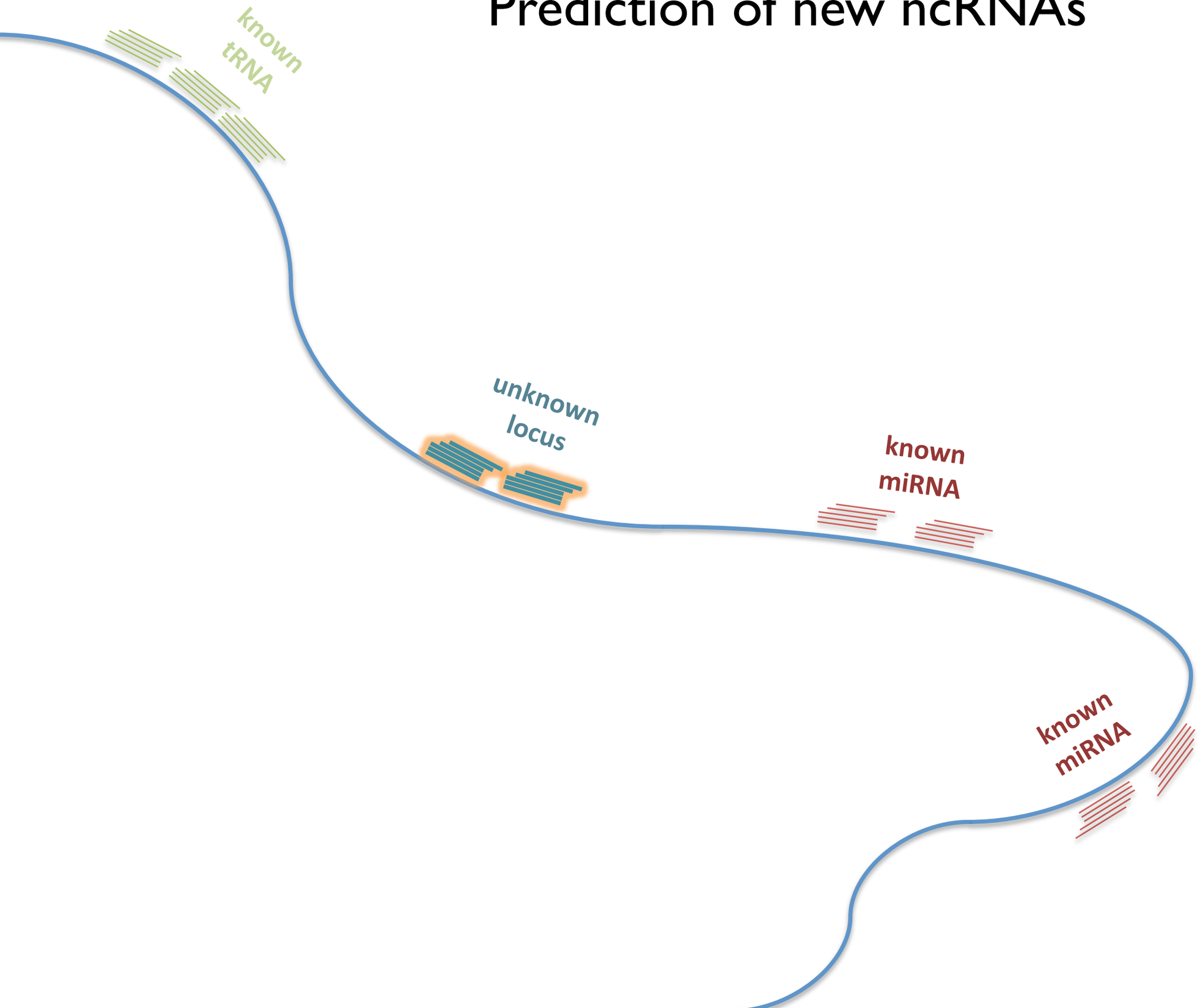
Recurring analyses



expression of ncRNAs

Chromosome	Start Loci	End Loci	Strand	ID	RPM	Reads	Reads (normalized)
chr10	100144965	100145054	-	hsa-mir-1287	2.44e-01	30	3.000
chr10	103351164	103351244	+	hsa-mir-3158-1	1.90e-01	42	2.100
chr10	103351164	103351244	-	hsa-mir-3158-2	1.90e-01	42	2.100
chr10	104186259	104186331	+	hsa-mir-146b	1.20e+01	1.203	120.200
chr10	105144000	105144148	-	hsa-mir-1307	6.16e+00	1.255	125.500
chr10	112738674	112738761	+	hsa-mir-548e	3.86e-01	49	4.650
chr10	115923854	115923928	-	hsa-mir-2110	5.46e-01	56	5.600
chr10	118917179	118917275	-	hsa-mir-3663	7.54e-03	1	100

Prediction of new ncRNAs



Summary:
A Helpfull Tool

- platform independent
- handle huge datasets
- automate recurring analyses
- automatically predict new ncRNAs

Summary:
A Helpfull Tool

- allow user annotations
- prepare data for further analysis (e.g. differential expression)
- easy to use web interface

DARIO

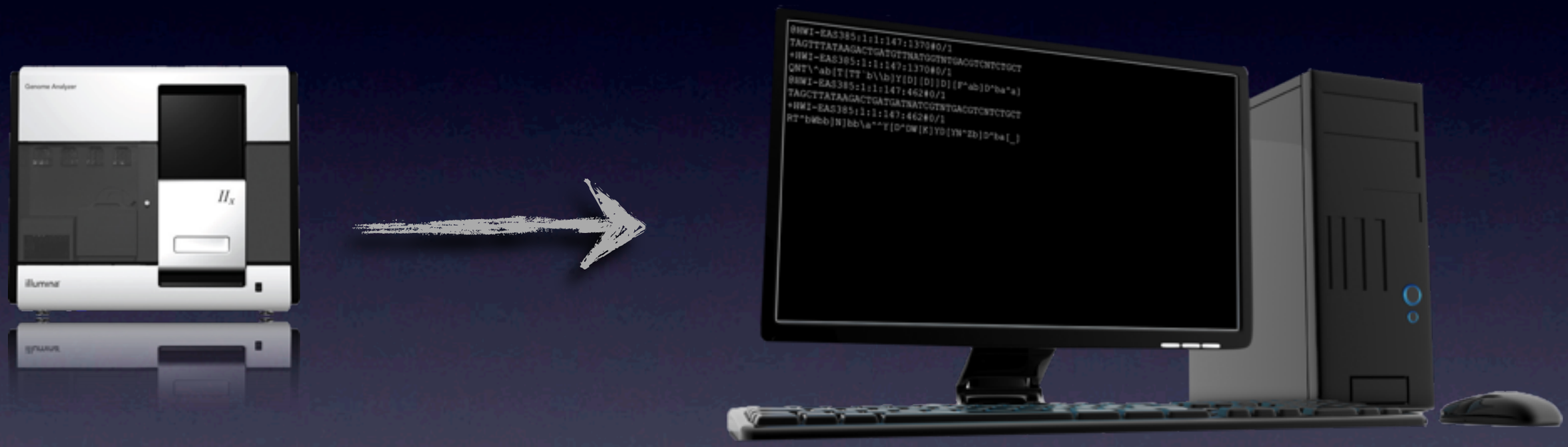
DARIO Web Server

A free web server for the analysis of short RNAs from high throughput sequencing data

DARIO

Deep Analysis of Reads in Interesting Organisms

SEQUENCING EXPERIMENT AND ITS RESULTS



Sequence small RNAs **using a sequencing machine of your choice**.

The result will be a list of sequenced RNA molecules (reads) in a sequencer specific output file (typically in fastq format).

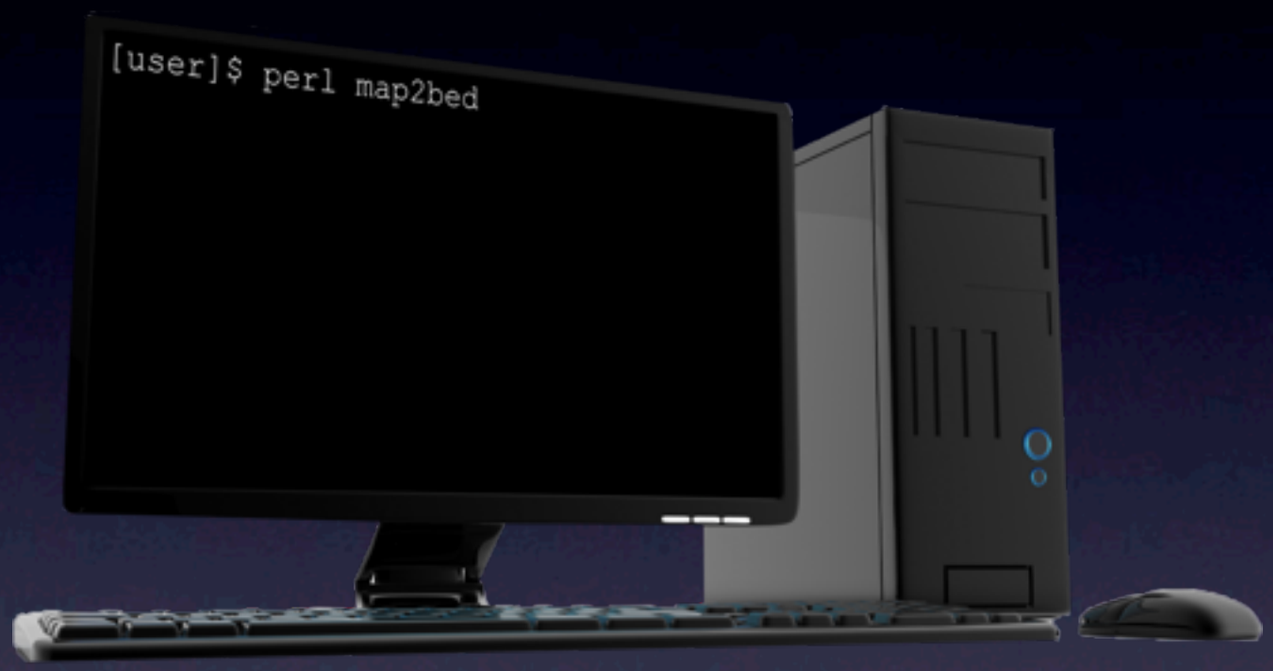
MAP READS TO A REFERENCE GENOME

Freely choose one of your favorite mapping tools to map your data to a reference genome. For example:

- segemehl
- BWA
- Bowtie
- SOAP
- ...

```
@HD VN:1.0
@SQ SN:chrI LN:15072421
@SQ SN:chrII LN:15279323
@SQ SN:chrIII LN:13783681
@SQ SN:chrIV LN:17493785
@SQ SN:chrV LN:20919568
@SQ SN:chrX LN:17718854
@PG ID:segemehl VN:0.9.4-$Rev: 162 $ ($Date: 2010-10-15 12:48:37 +0200 (Fri, 15 Oct 2010) $)
GPL9269_GSM427346_GSE17153_102271 0 chrI   15070506 255  31M1D1M  *   0   0   ATTCTTAGTTGGTTGAGCGAT   *   NM:i:4   MD:Z   XN:i:1
GPL9269_GSM427346_GSE17153_102271 0 chrI   15063309 255  31M1D1M  *   0   0   ATTCTTAGTTGGTTGAGCGAT   *   NM:i:4   MD:Z   XN:i:1
GPL9269_GSM427346_GSE17153_384015 0 chrIV 3233310 255  21M  *   0   0   TGAGATCGTTCAGTACGGCAA   *   NM:i:0   MD:Z:21   XN:i:1
GPL9269_GSM427346_GSE17153_384015 0 chrIV 3233310 255  21M  *   0   0   TGAGATCGTTCAGTACGGCAA   *   NM:i:0   MD:Z:21   XN:i:1
GPL9269_GSM427346_GSE17153_384015 0 chrIV 3233310 255  21M  *   0   0   TGAGATCGTTCAGTACGGCAA   *   NM:i:0   MD:Z:21   XN:i1
...
```

PREPARE YOUR MAPPED READS FOR THE DARIO UPLOAD



Automatically convert the output to the needed bed format using **`map2bed.pl`.**
Optionally you can create the bed file yourself.

`map2bed.pl` will not only create a bed formatted file, but also merge mapped reads to tags and zip the output file. This will **minimize the file size**, resulting in a **short upload time**.

(Note: It might be necessary to explicitly select 'SAM format' as output format when running your mapping tool.)

UR MAPPED READS
DARIO UPLOAD

~1 GB
(SAM output)

↓

398 MB
(BED file with reads)

↓

60 MB
(BED file with tags)

↓

15 MB
(gzipped BED file with tags)

e needed bed format using **`map2bed.pl`**.
ourself.

`map2bed.pl` will not only create a bed formatted file, but also merge mapped reads to tags and zip the output file. This will **minimize the file size**, resulting in a **short upload time.**

(Note: It might be necessary to explicitly select 'SAM format' as output format when running your mapping tool.)

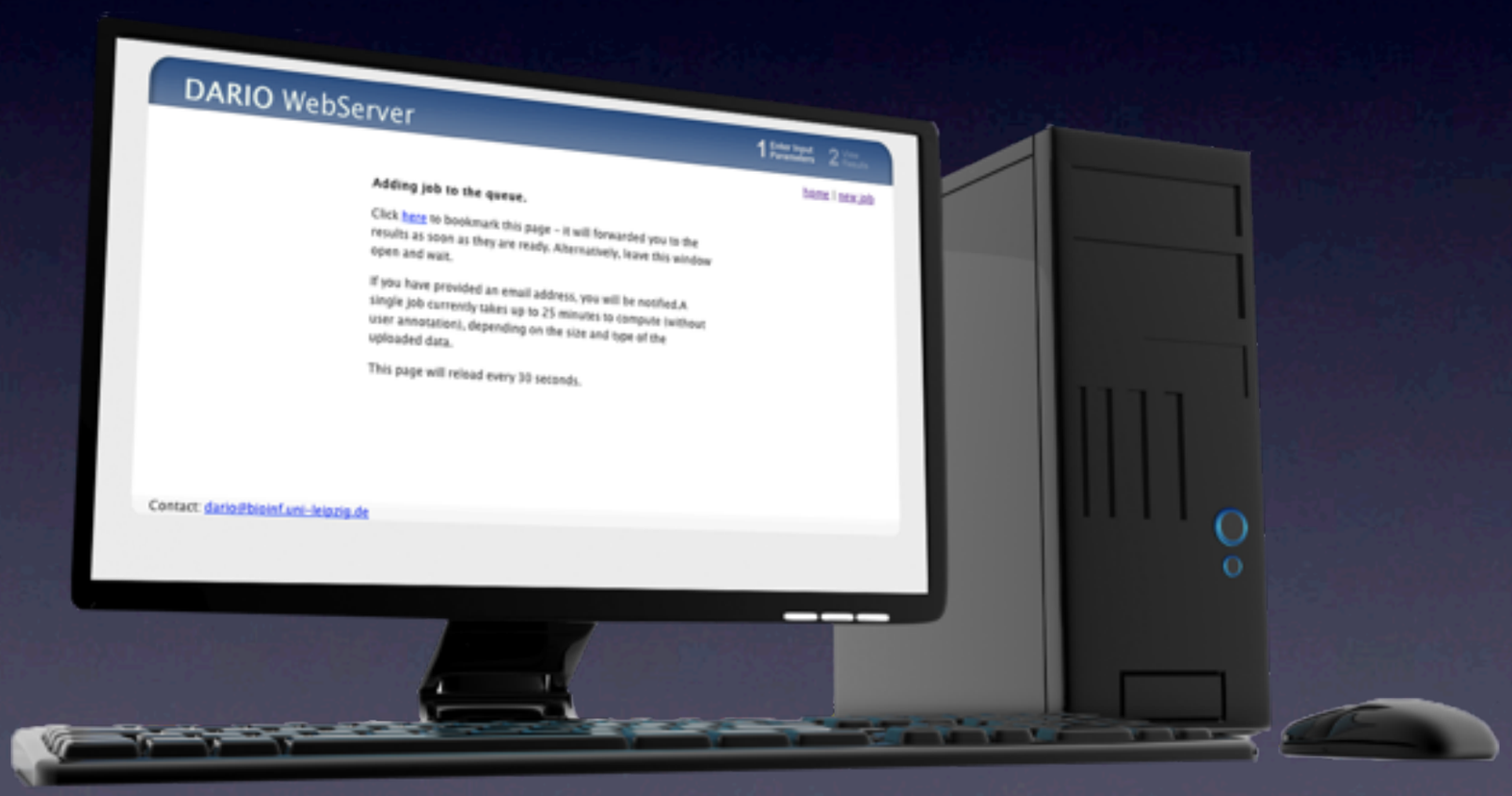
UPLOAD YOUR DATA TO THE DARIO WEBSERVER

Open the DARIO WebServer **http://dario.bioinf.uni-leipzig.de**

- Click on ***CLICK HERE TO START A NEW ANALYSIS JOB***
- Choose your reference species
- Choose upload file
- Optionally: choose a list of your own loci of interest
- Optionally: specify an e-mail address
- Click ***Proceed***

WAIT A MINUTE...

The job is automatically added to a queue and starts as soon as possible. This page reloads every 30 sec and opens the result page when the job is done.



If you specified your e-mail address, you will get an e-mail with a link to the result page, as soon as the job is finished.

...DONE

After **~10 to 25 minutes** running time the result page is generated, containing the following sections:

- **Summary**
- **Quality Control**
- **Analysis**
- **Prediction**
- **User Annotation**
- **Download**

SUMMARY

Summary

Click here to bookmark this page for reference. This result will be available online at least 2 weeks from your jobs finish time.

You will find all downloadable files at the end of this page.

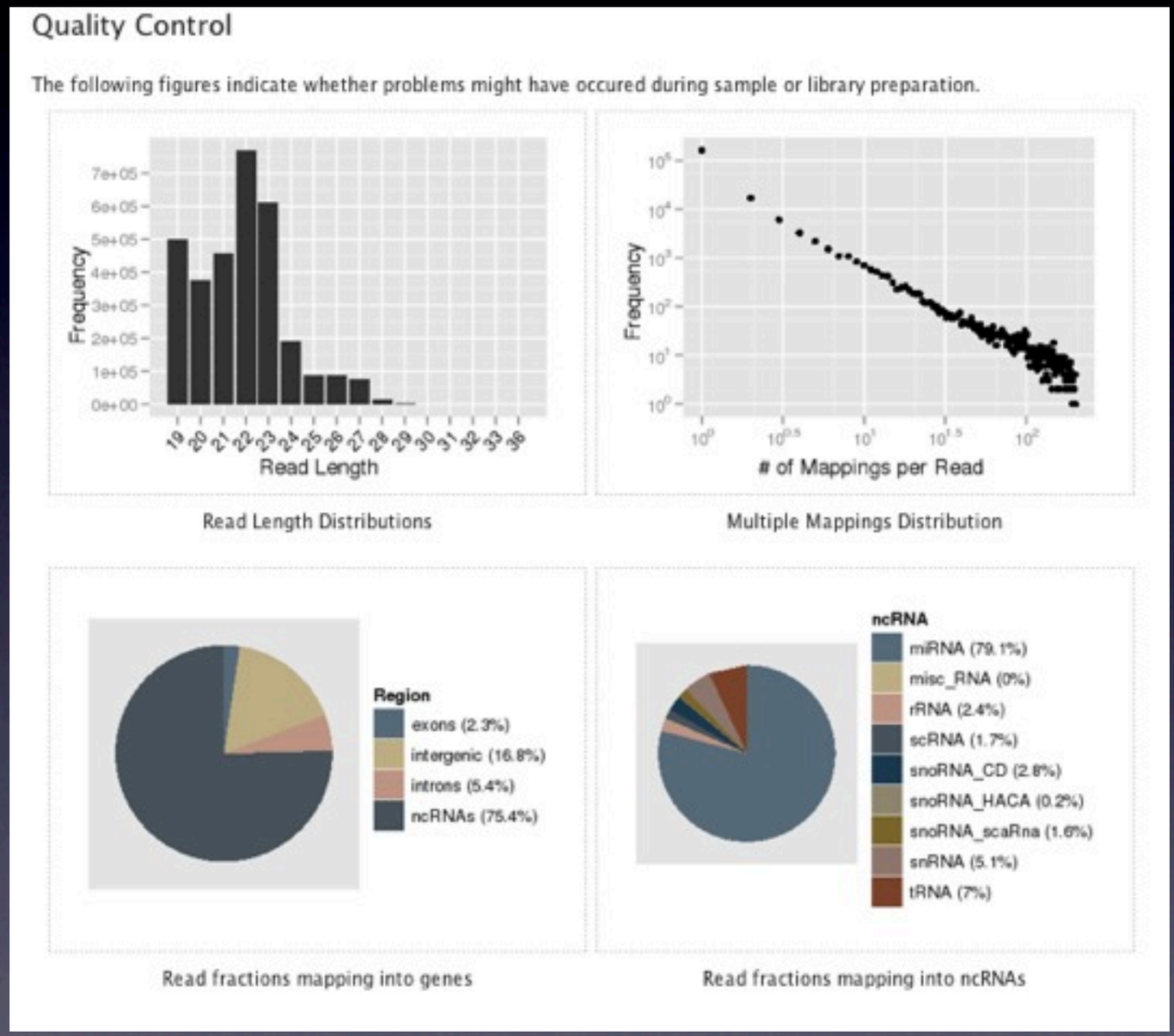
Job received at	2010-12-27 20:45:55
Job finished at	2010-12-27 20:58:32
# of uploaded mapping loci	582,333
Total # of reads	1,367,246
Total # of tags	203,811

The **Summary** contains some basic information about the job, e.g.:

- The date and time you uploaded your file
- The number of mapped loci you uploaded
- The number of reads and the number of tags*

*A tag is defined as a RNA sequence that occurs at least once in a set of sequencer reads. Thus a tag typically corresponds to several identical reads.

QUALITY CONTROL



The **Quality Control** gives a first impression on how good your experiment performed.

ANALYSIS

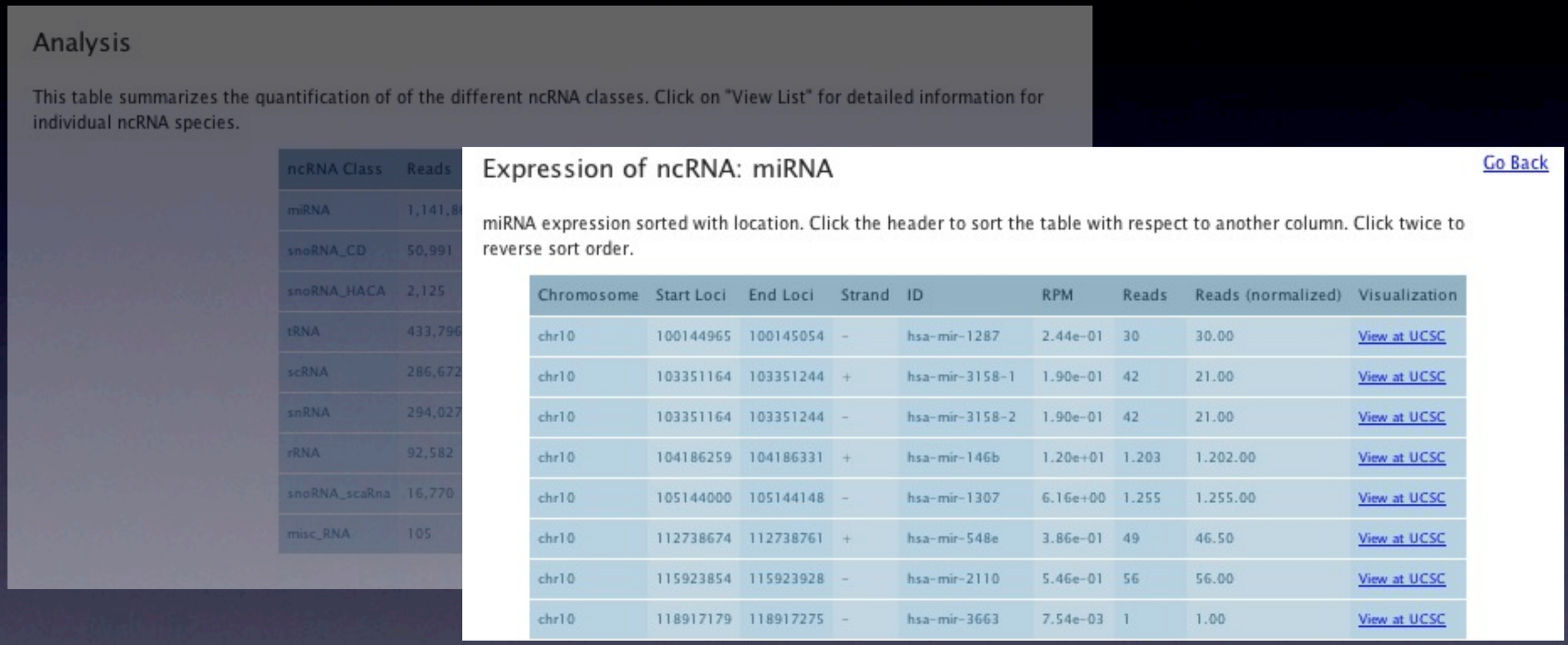
Analysis

This table summarizes the quantification of of the different ncRNA classes. Click on "View List" for detailed information for individual ncRNA species.

ncRNA Class	Reads	Reads (normalized)	# of Genes	Table
miRNA	1,141,865	816,175.5	661	View List
snoRNA_CD	50,991	29,031.09	181	View List
snoRNA_HACA	2,125	2,019.533	85	View List
tRNA	433,796	72,332.05	537	View List
scRNA	286,672	17,815.27	900	View List
snRNA	294,027	52,661.4	712	View List
rRNA	92,582	24,588.18	389	View List
snoRNA_scaRna	16,770	16,752.33	19	View List
misc_RNA	105	95.00559	8	View List

The tables in the **analysis section** are itemized by the types of ncRNAs.

ANALYSIS



Analysis

This table summarizes the quantification of of the different ncRNA classes. Click on "View List" for detailed information for individual ncRNA species.

ncRNA Class	Reads
miRNA	1,141,8
snoRNA_CD	50,991
snoRNA_HACA	2,125
tRNA	433,796
scRNA	286,672
snRNA	294,027
rRNA	92,582
snoRNA_scaRna	16,770
misc_RNA	105

Expression of ncRNA: miRNA

Go Back

miRNA expression sorted with location. Click the header to sort the table with respect to another column. Click twice to reverse sort order.

Chromosome	Start Loci	End Loci	Strand	ID	RPM	Reads	Reads (normalized)	Visualization
chr10	100144965	100145054	-	hsa-mir-1287	2.44e-01	30	30.00	View at UCSC
chr10	103351164	103351244	+	hsa-mir-3158-1	1.90e-01	42	21.00	View at UCSC
chr10	103351164	103351244	-	hsa-mir-3158-2	1.90e-01	42	21.00	View at UCSC
chr10	104186259	104186331	+	hsa-mir-146b	1.20e+01	1.203	1.202.00	View at UCSC
chr10	105144000	105144148	-	hsa-mir-1307	6.16e+00	1.255	1.255.00	View at UCSC
chr10	112738674	112738761	+	hsa-mir-548e	3.86e-01	49	46.50	View at UCSC
chr10	115923854	115923928	-	hsa-mir-2110	5.46e-01	56	56.00	View at UCSC
chr10	118917179	118917275	-	hsa-mir-3663	7.54e-03	1	1.00	View at UCSC

The expression list contains:

- the ncRNA loci (chromosomal position)
- the ncRNA ID
- the number of reads overlap with the loci
- the number of overlapping reads, normalized for multiple mappings
- the RPM (Reads Per Million) normalized expression for each expressed ncR

ANALYSIS

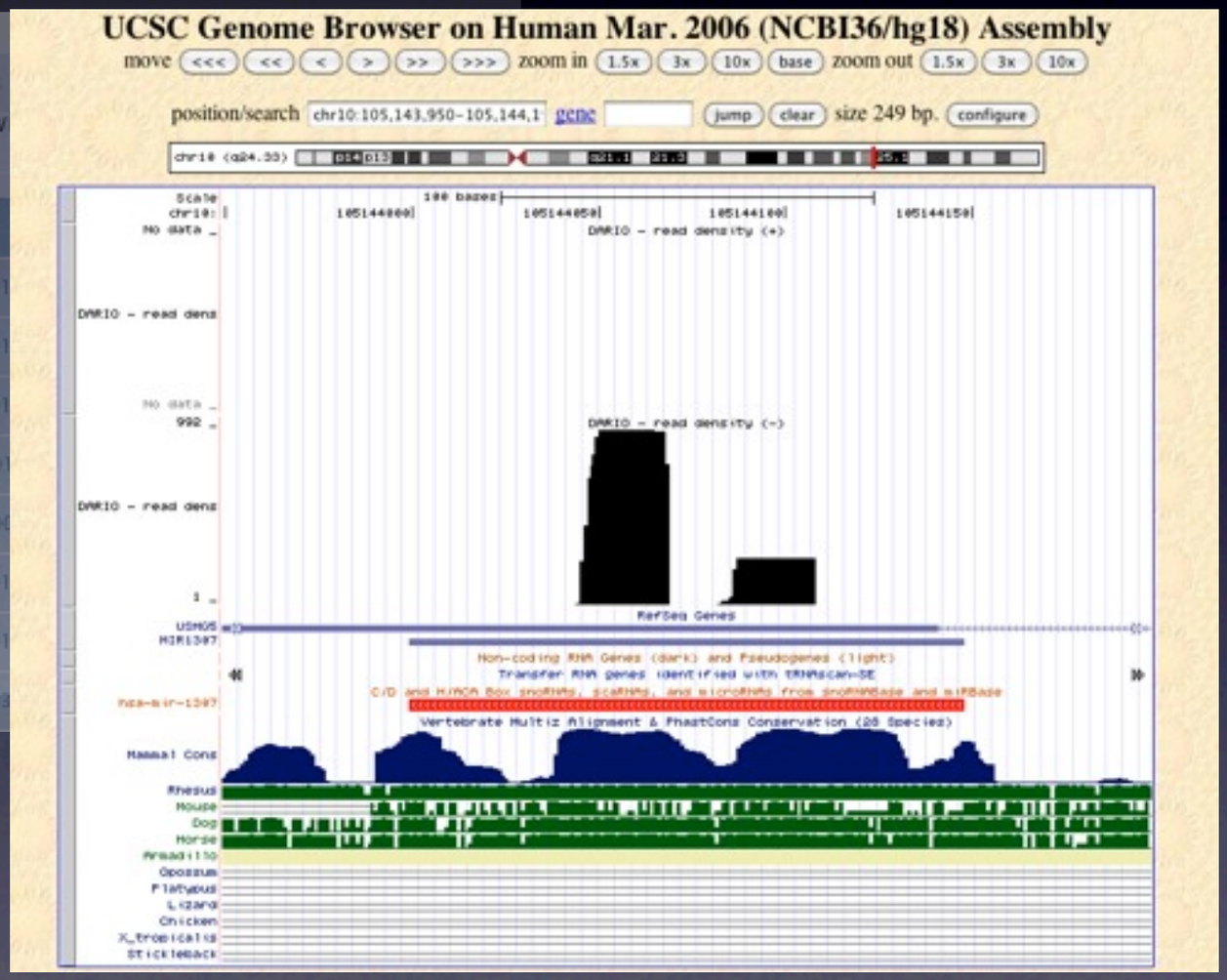
Analysis

This table summarizes the quantification of of the different ncRNA classes. Click on "View List" for detailed information for individual ncRNA species.

Expression of ncRNA: miRNA

miRNA expression sorted with location. Click the header to sort the table w
reverse sort order.

Chromosome	Start Loci	End Loci	Strand	ID	RPM
chr10	100144965	100145054	-	hsa-mir-1287	2.44e-01
chr10	103351164	103351244	+	hsa-mir-3158-1	1.90e-01
chr10	103351164	103351244	-	hsa-mir-3158-2	1.90e-01
chr10	104186259	104186331	+	hsa-mir-146b	1.20e+01
chr10	105144000	105144148	-	hsa-mir-1307	6.16e+00
chr10	112738674	112738761	+	hsa-mir-548e	3.86e-01
chr10	115923854	115923928	-	hsa-mir-2110	5.46e-01
chr10	118917179	118917275	-	hsa-mir-3663	7.54e-03



Take a look on the **expression pattern**.

PREDICTION

Predictions

ncRNAs patterns were identified using blockbuster . Random forest classification was performed using WEKA software. The used machine learning approach is published and can be looked up here.

General classifcation statistics:

	miRNA	snoRNA_HACA	snoRNA_CD	tRNA
miRNA	193	0	12	12
snoRNA_HACA	1	0	2	1
snoRNA_CD	15	0	39	35
tRNA	12	0	12	338

Confusion Matrix

These new putative ncRNAs were predicted within your data.

ncRNA Class	# of Genes	Table
miRNA	19	View List
snoRNA_CD	28	View List
tRNA	141	View List

The upper part gives information about the sensitivity of the predictions on the uploaded dataset. Below the predicted loci are shown.

PREDICTION

New ncRNAs candidates are identified in the uploaded data and ranked by their scores.

Predictions of ncRNA: miRNA

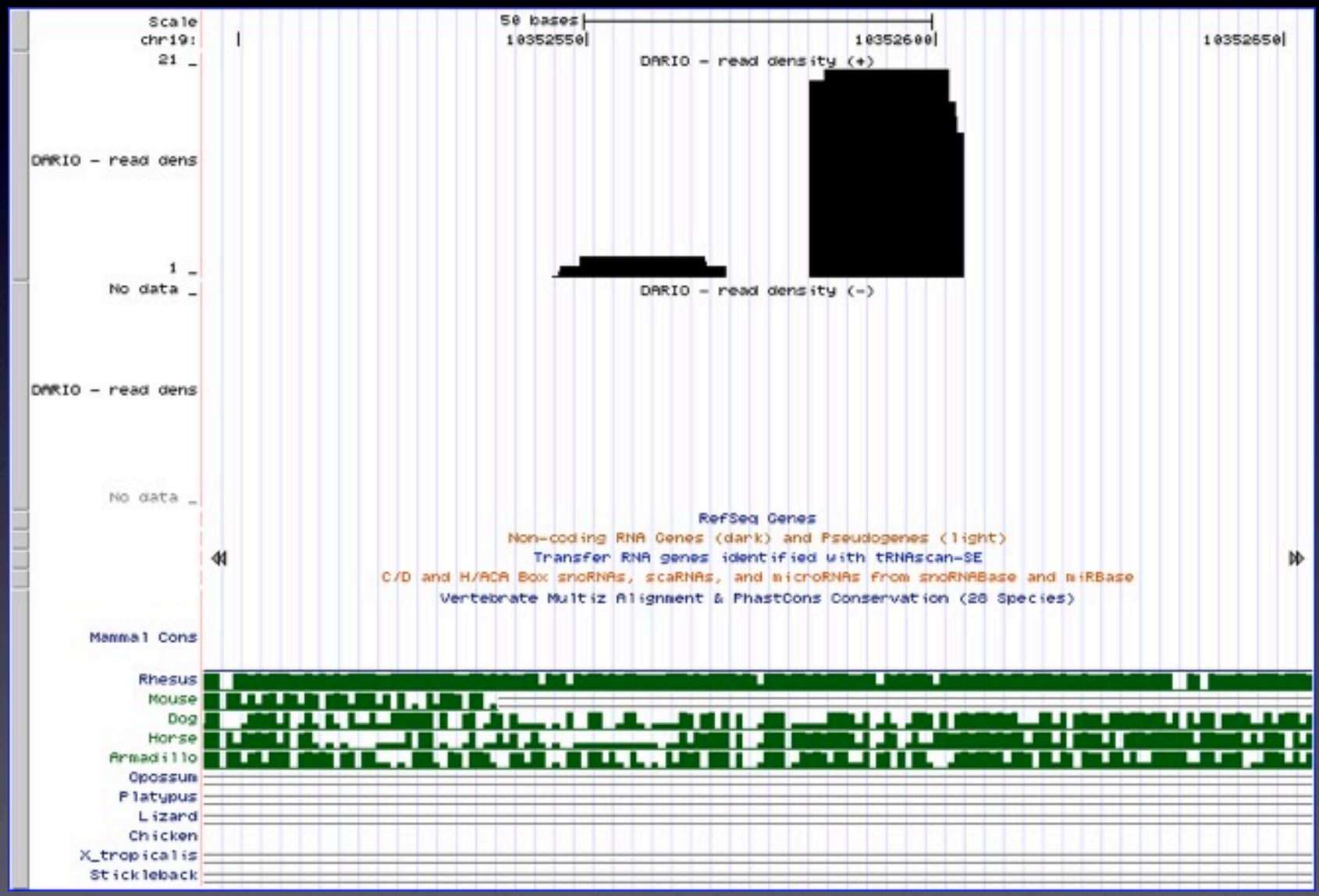
Go Back

Sorted miRNA predictions descending with score. Click the header to sort the table with respect to another column. Click twice to reverse sort order.

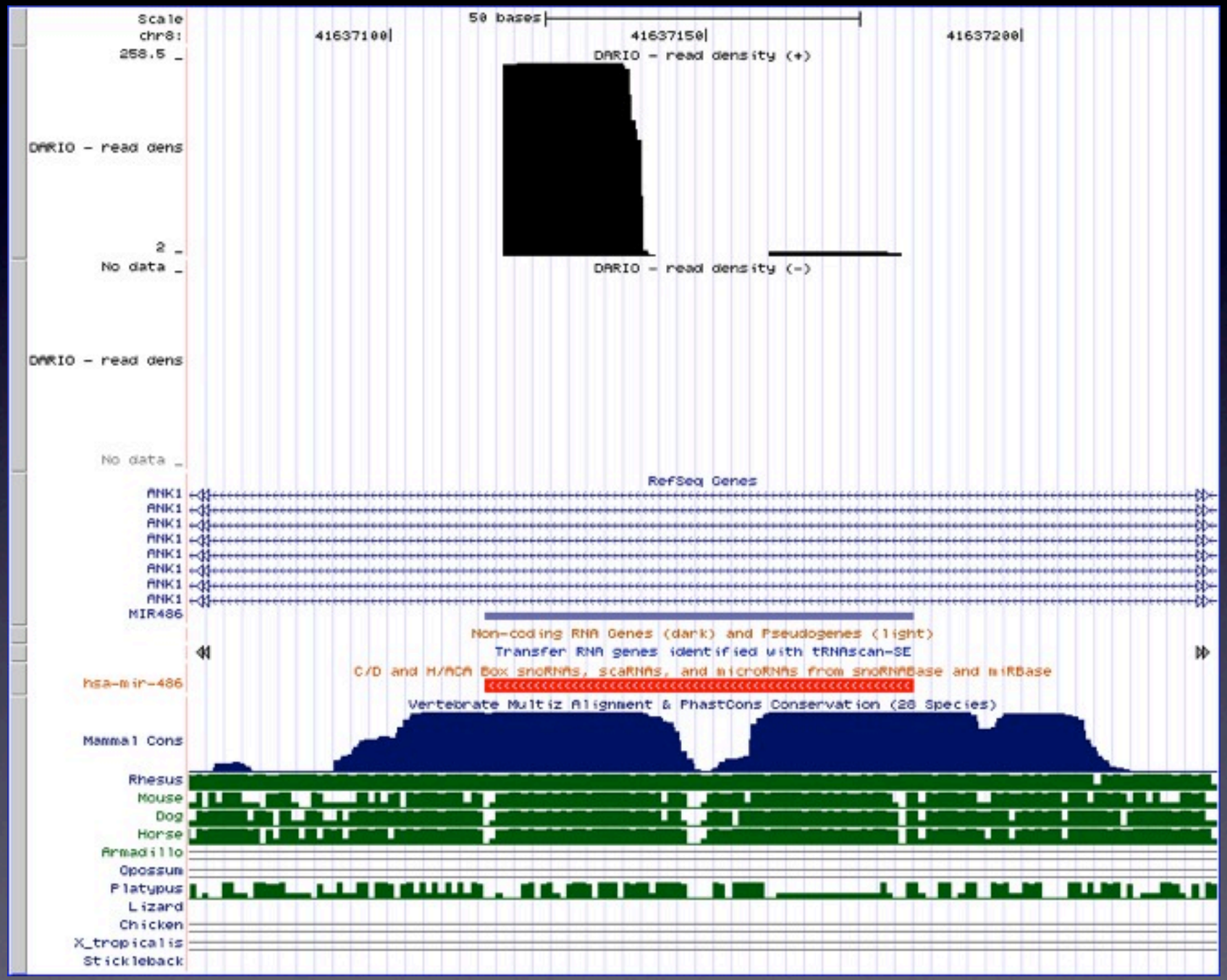
Chromosome	Start Loci	End Loci	Strand	ID	Score	RPM	Reads	Reads (normalized)	RNAz Validation	Visualization
chr11	67984196	67984253	-	miRNA_94	1	1.62e-01	14	14.00		UCSC
chr14	65007583	65007643	-	miRNA_1	1	1.27e+00	211	106.00		UCSC
chr11	62091059	62091114	-	miRNA_76	0.99	1.83e-01	14	14.00		UCSC
chr19	10352546	10352604	+	miRNA_72	0.98	2.98e-01	24	24.00		UCSC

Futhermore, we overlap the predicted loci with RNAz screens. RNAz uses conservation and secondary structure properties to predict regions likely to form functional RNA structures.

PREDICTION


Scale
chr19:
50 bases
10352550
10352600
10352650
21
DARIO - read density (+)
DARIO - read dens
1
No data
DARIO - read density (-)
DARIO - read dens
No data
RefSeq Genes
Non-coding RNA Genes (dark) and Pseudogenes (light)
Transfer RNA genes identified with tRNAscan-SE
C/D and H/ACA Box snoRNAs, scaRNAs, and microRNAs from snoRNABase and miRBase
Vertebrate Multiz Alignment & PhastCons Conservation (28 Species)
Mammal Cons
Rhesus
Mouse
Dog
Horse
Armadillo
Opossum
Platypus
Lizard
Chicken
X_tropicalis
Stickleback

PREDICTION


Scale
chr8:
258.5
DARIO - read dens
2
No data
DARIO - read dens
No data
50 bases
41637100
41637150
41637200
DARIO - read density (+)
DARIO - read density (-)
RefSeq Genes
ANK1
ANK1
ANK1
ANK1
ANK1
ANK1
ANK1
ANK1
MIR486
Non-coding RNA Genes (dark) and Pseudogenes (light)
Transfer RNA genes identified with tRNAscan-SE
C/D and H/ACA Box snoRNAs, scaRNAs, and microRNAs from snoRNABase and miRBase
hsa-mir-486
Vertebrate Multiz Alignment & PhastCons Conservation (28 Species)
Mammal Cons
Rhesus
Mouse
Dog
Horse
Armadillo
Opossum
Platypus
Lizard
Chicken
X_tropicalis
Stickleback

PREDICTION


Scale
chr11:
No data
DARIO - read dens
No data
11
DARIO - read dens
1
50 bases
62091050
62091100
62091150
DARIO - read density (+)
DARIO - read density (-)
RefSeq Genes
EEF1G
Non-coding RNA Genes (dark) and Pseudogenes (light)
Transfer RNA genes identified with tRNAscan-SE
C/D and H/ACA Box snoRNAs, scaRNAs, and microRNAs from snoRNABase and miRBase
Vertebrate Multiz Alignment & PhastCons Conservation (28 Species)
Mammal Cons
Rhesus
Mouse
Dog
Horse
Armadillo
Opossum
Platypus
Lizard
Chicken
X_tropicalis
Stickleback

USER ANNOTATION

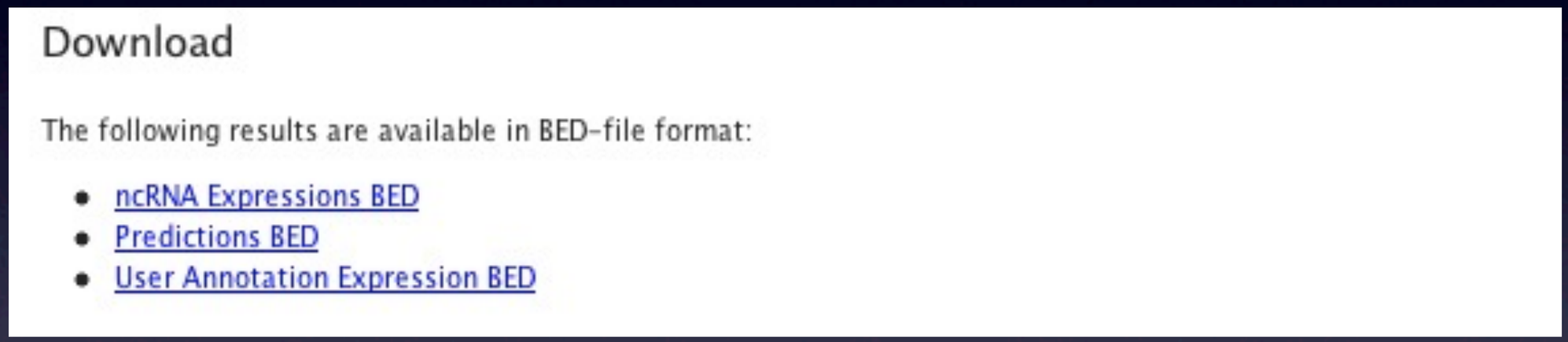
User Annotation

You have provided additional annotation (ncRNAs.bed). The following table shows expression of those.

Class Identifier	# of Genes	Table
snoRNA_CD	181	View List
rRNA	389	View List
snRNA	712	View List
tRNA	537	View List
scRNA	900	View List
snoRNA_HACA	85	View List
miRNA	661	View List
snoRNA_scaRna	19	View List
misc_RNA	8	View List

The output is in the same format as the analysis with the user defined annotations instead of the known ncRNA loci.

DOWNLOAD


Download

The following results are available in BED-file format:

- ncRNA Expressions BED
- Predictions BED
- User Annotation Expression BED

Of course all the expression data and predicted candidates DARIO calculated are downloadable in bed format.

It is possible to use the predicted ncRNA candidates as ***User Annotation*** in upcoming DARIO runs.

We hope that you will enjoy working with DARIO!

Acknowledgements

Steve Hoffmann
Jens Steuck
Christian Otto
Andreas Gruber
Alexander Donath
Fabian ???