SMALL RNA SEQUENCING AND ANALYSIS

PHD PROGRAM   16 JANUARY 2018

Tao YE

Bioinformatics Engineer

GenomEast Platform (IGBMC)

CONTENT

- Introduction of small RNA
- Sample preparation (Illumina TruSeq)
- Sequencing
- Bioinformatics analysis
- Data visualisation
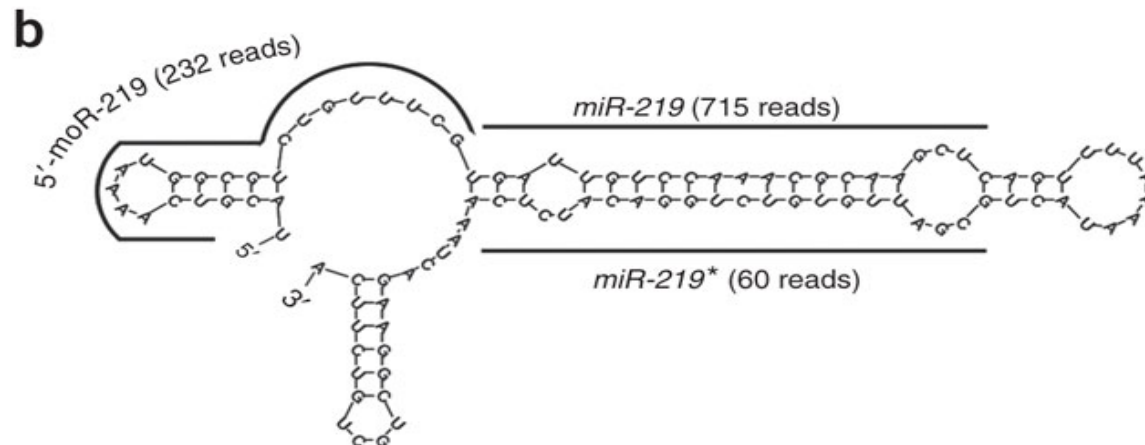- Integration of RNAseq and small RNAseq data
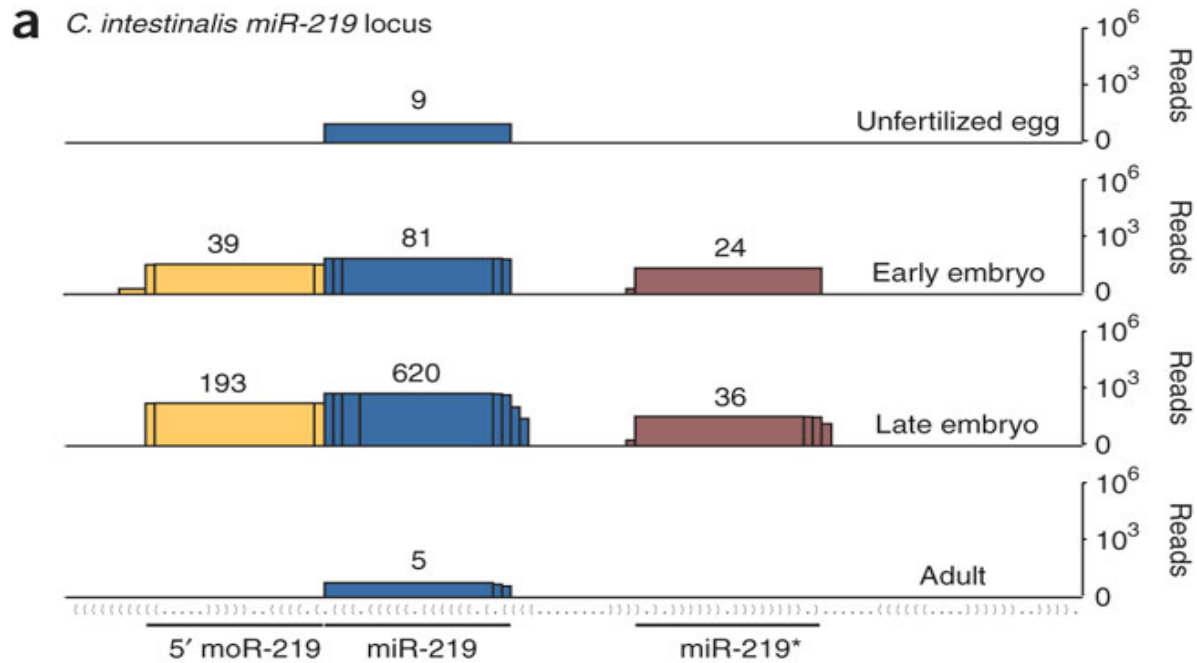
1. Introduction of small RNA

○ **Non-coding small RNA**

- ~ 2% genome: protein coding transcripts
- ?% for non-coding RNA?
- miRNA, piRNA, rRNA, tRNA, snRNA, snoRNA, etc…

○ **MicroRNA:**

- essential roles in many biological processes
- 30% - 60% mRNAs targeted by 1+ miRNA
- Mature micro RNA: ~ 22nt

MICRO RNA SEQUENCES

2. SAMPLE PREPARATION (ILLUMINA TRUSEQ)

Total RNA or purified small RNA Fragments (min: 1μg)

5' p   3' OH

Ligation of PCR adapter

RNA 5' Adapter   RNA 3' Adapter

PCR Amplification

Primer with index sequence

Gel purification
(pooled index option)

~150bp

Sequencing

50bp   +   6bp
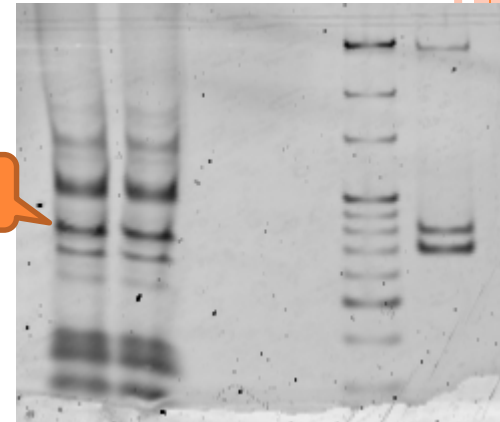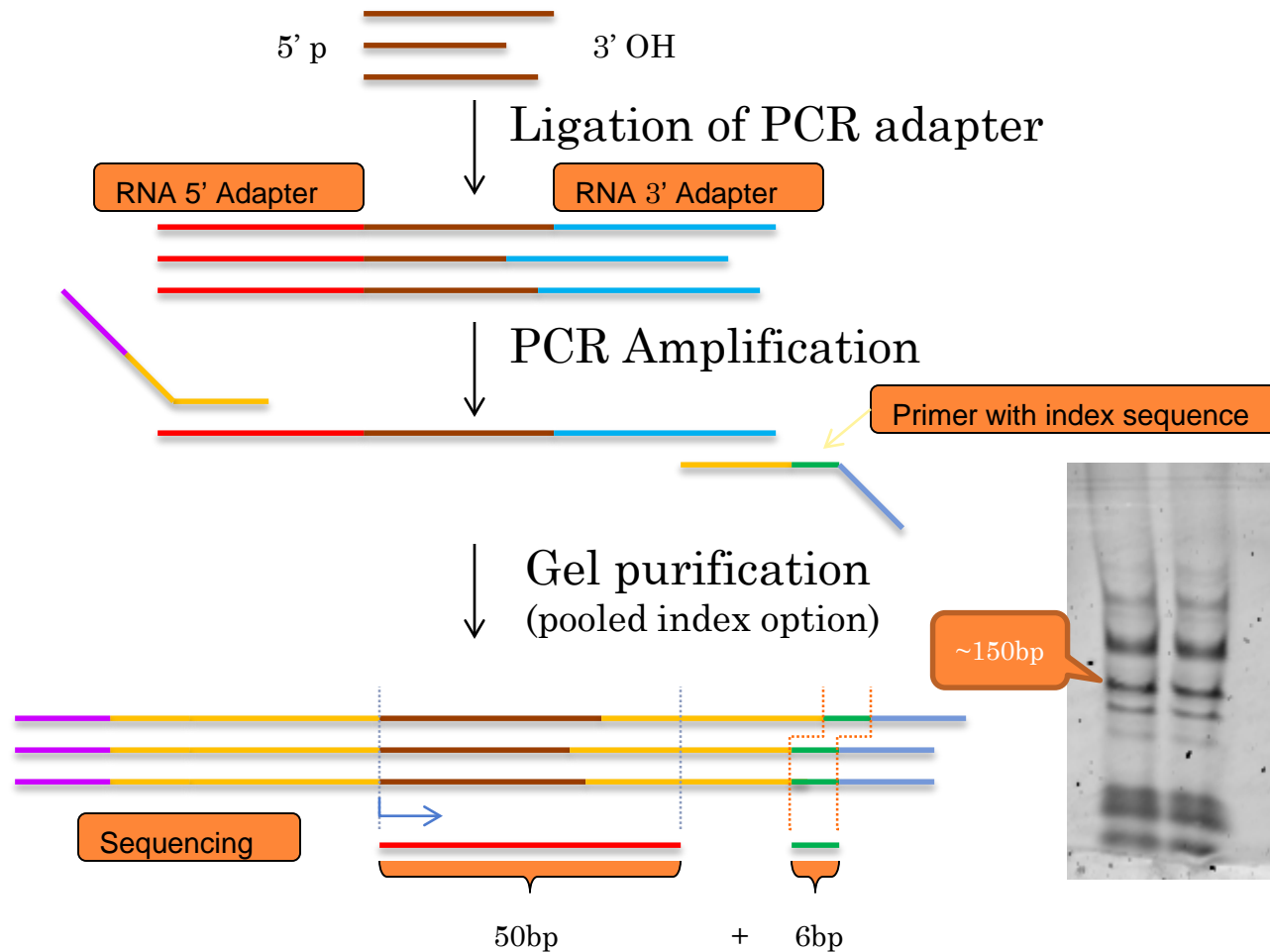
3. SEQUENCING

Hiseq 4000
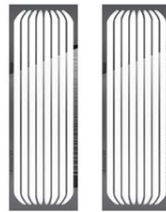
~300 million reads/lane
8 lanes / flowcell

Possiblity to multiplex with indexing ,up to 24 samples/lane, > 10 millions reads/sample

- Basepair resolution, possibility to find nucleotide modification
- Novel small RNA discovery
- ☹ More time for sample preparation, data analysis

QUALITY CONTROL

ChIP-seq

- Per base sequence content

FastQC (Simon Andrews)

QUALITY CONTROL

| 5' PCR primer | sequence | 3' PCR primer |

Sequencing

50bp   +   6bp

- ## Overrepresented sequences

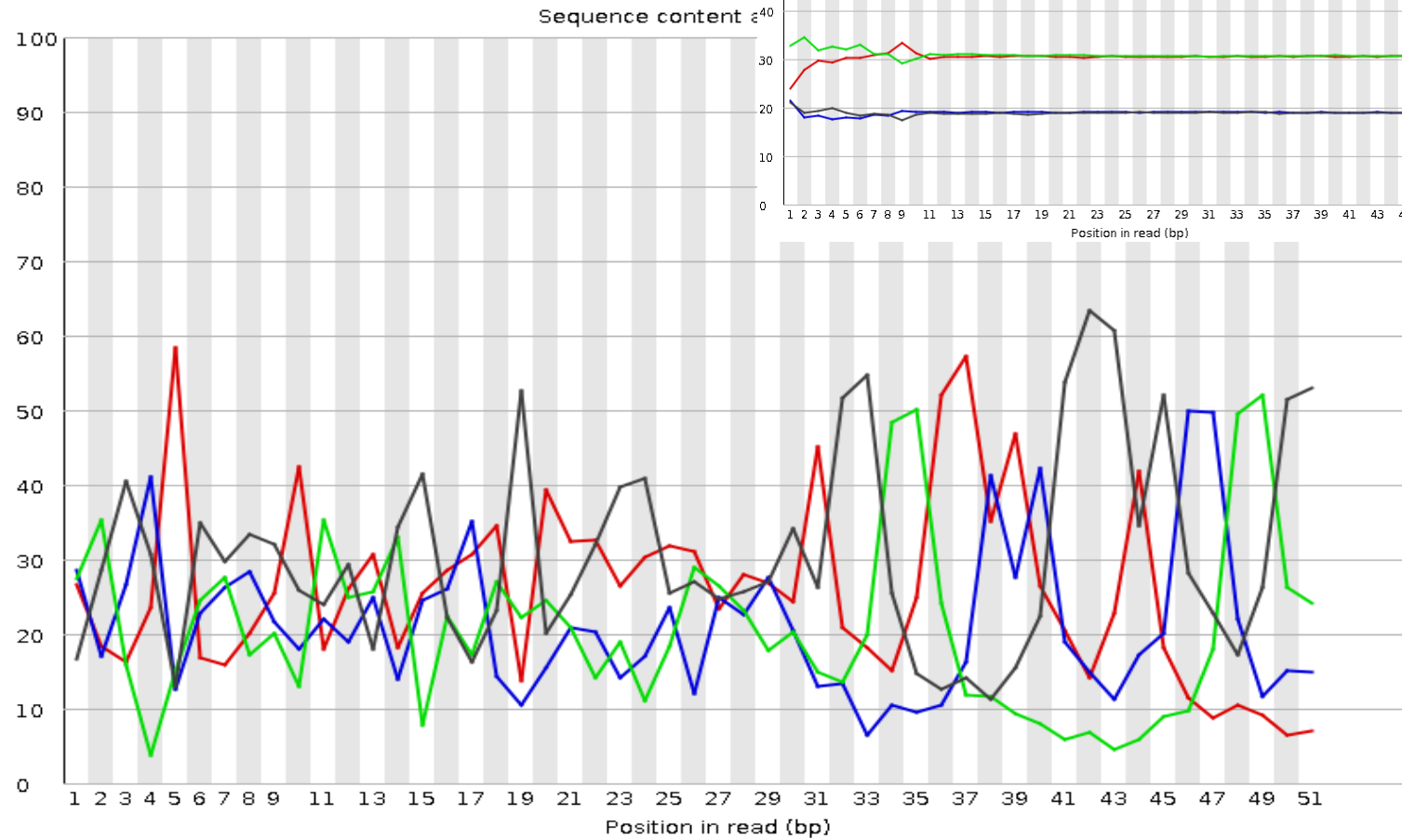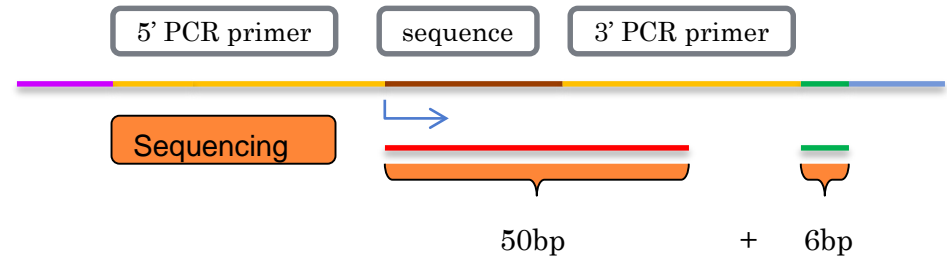| Sequence	Count	Percentage	Possible Source
TCGGATCCGTCTGAGCTTGGCTTGGAATTCTCGGGT	591452	7.308658256228866	No Hit
TCGGATCCGTCTGAGCTTGGCTGGAATTCTCGGGTG	371381	4.589208949934285	No Hit
TCGGATCCGTCTGAGCTTGGTGGAATTCTCGGGTGC	354088	4.375516837598938	No Hit
AACATTCAACGCTGTCGGTGAGTGGAATTCTCGGGT	223282	2.759128099604522	No Hit
TTCAAGTAATCCAGGATAGGCTTGGAATTCTCGGGT	153015	1.8908285762443275	No Hit
TCAGTGCACTACAGAACTTTGTTGGAATTCTCGGGT	118351	1.462480494246266	No Hit
CAACGGAATCCCAAAAGCAGCTGTGGAATTCTCGGG	111818	1.3817512645066707	No Hit
TGTAAACATCCTCGACTGGATGGAATTCTCGGGTGC	110506	1.3655386899745492	No Hit
TGGAATTCTCGGGTGCCAAGGAACTCCAGTCACCGA	107078	1.323178396151293	RNA PCR Primer, Index 2 (100% over 36bp)
AACATTCAACGCTGTCGGTGAGTTGGAATTCTCGGG	92898	1.1479540768940661	No Hit
TGAGATGAAGCACTGTAGCTTGGAATTCTCGGGTGC	72117	0.8911602420221034	No Hit
TCTTTGGTTATCTAGCTGTATGATGGAATTCTCGGG	65899	0.8143235130276439	No Hit
TCGGATCCGTCTGAGCTTGGCATGGAATTCTCGGGT	64730	0.799878010262362	No Hit
TATTGCACTCGTCCCGGCCTCCTGGAATTCTCGGGT	60158	0.7433811423043901	No Hit
TCGGATCCGTCTGAGCTTGGCTATTGGAATTCTCGG	56546	0.6987471337601656	No Hit

4. BIOINFORMATICS ANALYSIS

miRanalyzer: a microRNA detection and analysis tool for next-generation sequencing experiments

Michael Hackenberg[1], Martin Sturm[2], David Langenberger[3,4], Juan Manuel Falcón-Pérez[5] and Ana M. Aransay[1,*]

[1]Functional Genomics Unit, CIC bioGUNE, CIBERehd, Technology Park of Bizkaia, 48160 Derio, Bizkaia, Spain, [2]Institute for Bioinformatics and Systems Biology, German Research Center for Environmental Health, Ingolstädter Landstrasse 1, D-85764 Neuherberg, [3]Department of Genome-Oriented Bioinformatics, Wissenschaftszentrum

BMC Bioinformatics

BioMed Central

Software   **Open Access**

miRExpress: Analyzing high-throughput sequencing data for profiling microRNA expression

Wei-Chi Wang[1], Feng-Mao Lin[1], Wen-Chi Chang[1,5], Kuan-Yu Lin[2,3], Hsien-Da Huang*[1,4] and Na-Sheng Lin*[2,3]

Address: [1]Institute of Bioinforma
of Biotechnology, National Chen
Sinica, Nankang, Taipei 11529, T
Hsin-Chu 300, Taiwan, Republic
of China
Email: Wei-Chi Wang · caaser.hl

DSAP: deep-sequencing small RNA analys

Genome **Biology**

METHOD   **Open Access**

miRTRAP, a computational method for the systematic identification of miRNAs from high throughput sequencing data

miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades

Marc R. Friedländer[1], Sebastian D. Mackow

BIOINFORMATICS   **APPLICATIONS NO**

Discovering microRNAs from deep sequencing data using miRDeep

NOTE   Vol. 26 no. 20 2010, pages 2615–2616
doi:10.1093/bioinformatics/btq493

Sequence analysis

CPSS: a computational platform for the ana deep sequencing data

Yuanwei Zhang[1,†], Bo Xu[1,†], Yifan Yang[2], Rongjun Ban[3], H Howard J. Cooke[1,4], Yu Xue[5,*] and Qinghua Shi[1,*]

[1]Hefei National Laboratory for Physical Sciences at Microscale and School of and Technology of China, Hefei 230027, China, [2]Department of Statistics, Ur 40506, USA, [3]Department of Computer Science & Technology, Nanjing Unive Genetics Unit, IGMM, University of Edinburgh, Edinburgh EH4 2XU, UK, and [5] Huazhong University of Science and Technology, Wuhan 430074, China
Associate Editor: Ivo Hofacker

Marc R Friedländer[1], Wei Chen[2], Catherine Adamidi[1], Jonas Maaskola[1], Ralf Einspanier[3], Signe Knespel[1] & Nikolaus Rajewsky[1]

The capacity of highly parallel sequencing technologies to detect small RNAs at unprecedented depth suggests their value in systematically identifying microRNAs (miRNAs). However, the identification of miRNAs from the large pool of sequenced transcripts from a single deep sequencing run remains a major challenge. Here, we present an algorithm, miRDeep, which uses a probabilistic model of miRNA

and 454 Life Sciences/Roche, can sequence DNA orders of magnitude faster and at lower cost than Sanger sequencing and are evolving so rapidly that increases in sequencing speed by at least another order of magnitude seem likely over the next few years. Although the Solexa/ Illumina system can produce ~32 million sequencing reads in one run, read length is currently limited to 35 bp. In contrast, the current 454 platform yields reads up to 200 bases each, although the number of reads
DOI 10.1097/s11103-012-9988-2

ep sequencing analysis

v[2], Gideon Dror[2], Eran Halperin[3,4]

dicine, Tel Aviv University, [2]The Academic ce Institute, Berkeley, CA, USA and [4]School Biotechnology, George Wise Faculty of Life

shortran: A pipeline for small RNA-seq data analysis

Vikas Gupta[1,2], Katharina Markmann[1], Christian N. S. Pedersen[2], Jens Stougaard[1] a Andersen[1,*]

[1]Centre for Carbohydrate Recognition and Signalling, Department of Molecular Biology and Genetics, Aarh Gustav Wieds Vej 10, 8000 Aarhus C, Denmark and [2]Bioinformatics Research Centre, Aarhus University, C 8, 8000 Aarhus C, Denmark

miRDeepFinder: a miRNA analysis tool for deep sequencing of plant small RNAs

Fuliang Xie · Peng Xiao · Dongliang Chen · Lei Xu · Baohong Zhang

4. Bioinformatics analysis

- Analysis consensus:
  - Sequences cleaning (cut the adapters)

  - Mapping to the target genome

  - Annotation and quantification of small RNAs expression

  - Novel miRNA discovery and quantification

  - Differential expression analysis

CUT THE ADAPTERS

- Example
  - Input reads
  - - adapter reads
  - - too-short reads
  - - non-clipped reads (optional)
  - Output reads: ~90% input
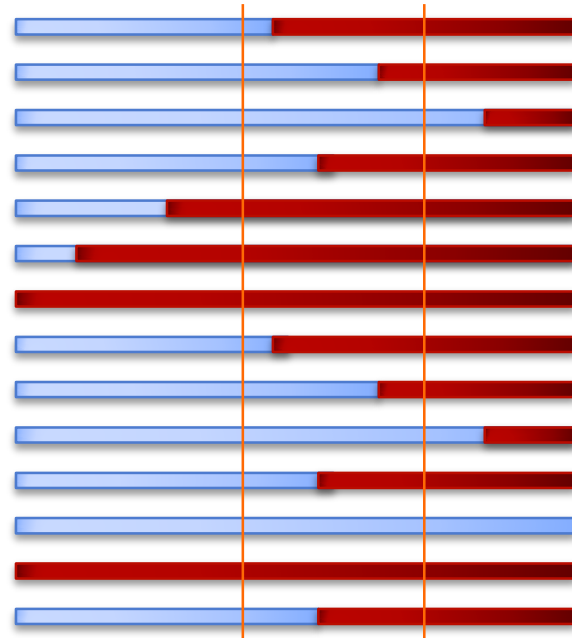
- Tools:
  - Fastx_toolkit: fastx_clipper (http://hannonlab.cshl.edu/fastx_toolkit/)
  - cutadapt(http://code.google.com/p/cutadapt/)

SIZE DISTRIBUTION
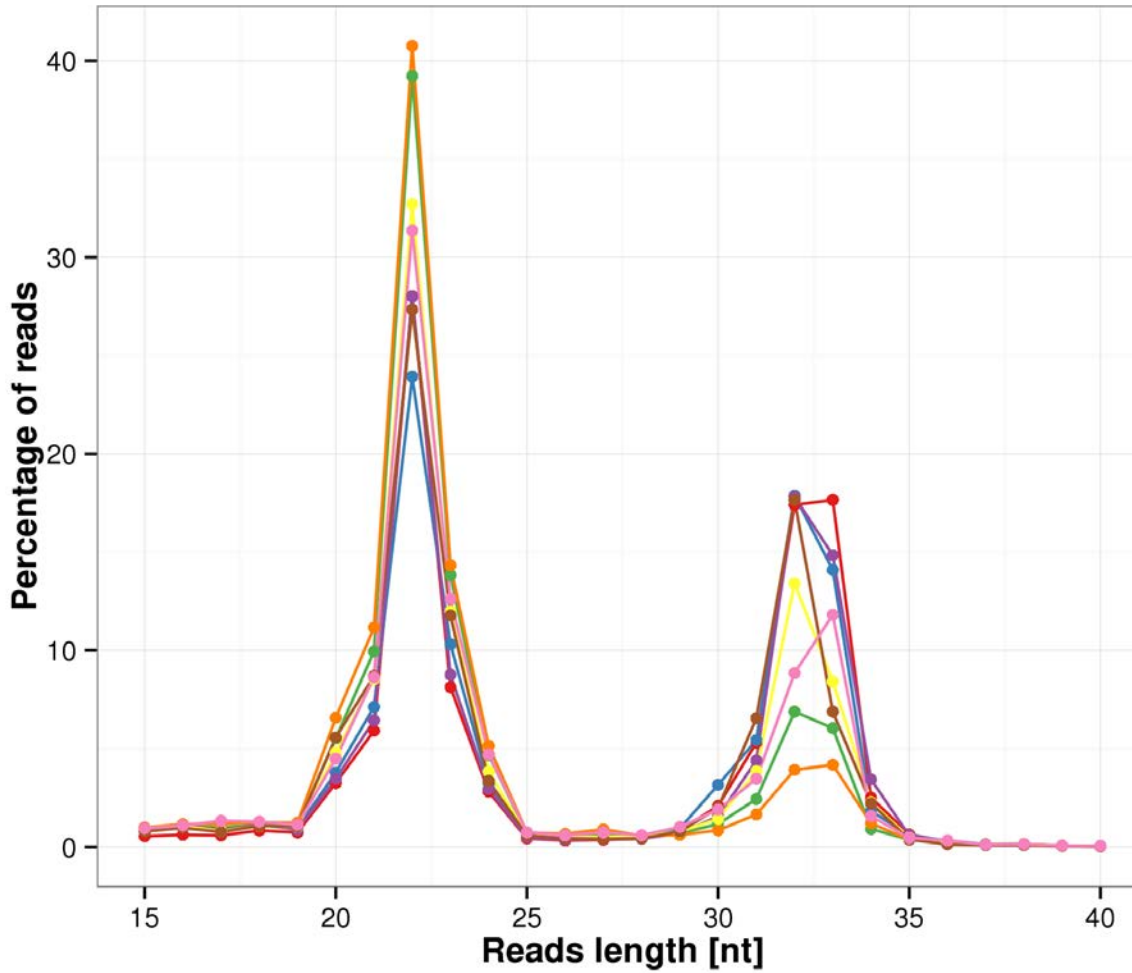
MAPPING

- Small non coding RNA:
  - Small sequence (≈ ChIP-Seq)
  - No splicing (≠ mRNA-Seq)
  - Multiply mappable reads
    - mmu-mir-9: chromosome 3, 7, 13
- Bowtie Aligner:
  - Burrows-Wheeler index
  - ultrafast, memory-efficient short read aligner
  - Used for ChIP-Seq, mRNA-Seq(TopHat, Cufflinks)
  - Output SAM/BAM, BED
  - example

    Input: 11,972,296 reads
    Failed: 904,561 (7.56%)
    Multimap > 15 times: 2,882,000 (31.22%)
    Output: 8,185,735 » » 15,044,571

Bowtie (Langmead B et al. Genome Biol)

Annotation

- miRBase (http://www.mirbase.org/)
  - Release 21 (June 2014)
  - Homo sapiens: 1881 precursors, 2588 mature
- Rfam (http://rfam.xfam.org/)
  - non-coding RNA family database
  - Release 12.0 (July 2014)
  - rRNA, tRNA, snoRNA, miRNA …
- NONCODE (http://www.noncode.org/)
  - a database of all kinds of noncoding RNAs (except tRNAs and rRNAs)
  - Release v4 (January 2014)
  - piRNA, **lncRNA**, etc.

QUANTIFICATION OF EXPRESSION

- ## GFF (General Feature Format)

  | chr1 | Rfam | ncRNA | 566311 | 566376 | 26.31 | - | . | ID=RF00005.7;Alias=tRNA; |
  | chr1 | Rfam | ncRNA | 1102495 | 1102568 | 74.59 | + | . | ID=RF00241.1;Alias=mir-8 |

- ## Overlap reads with annotations

5' mir mature   3' mir mature
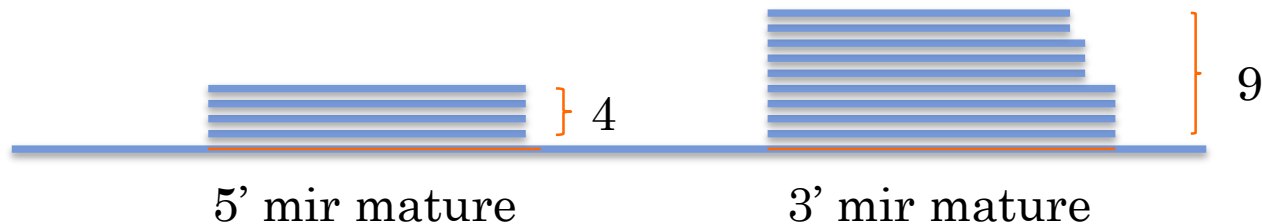
- ## Weight multiple mapping reads to the read number
  - A read aligned to 5 different locations on the genome, count 0.2 read for each location

DIFFERENTIAL EXPRESSION

- Expression, read counts ≈ mRNA-Seq

- Variability of experiment and sequencing are important ⟶ Need replicates

- DESeq2 package (Love MI, Huber W and Anders S (2014))
  - R language

DESᴇǫ2

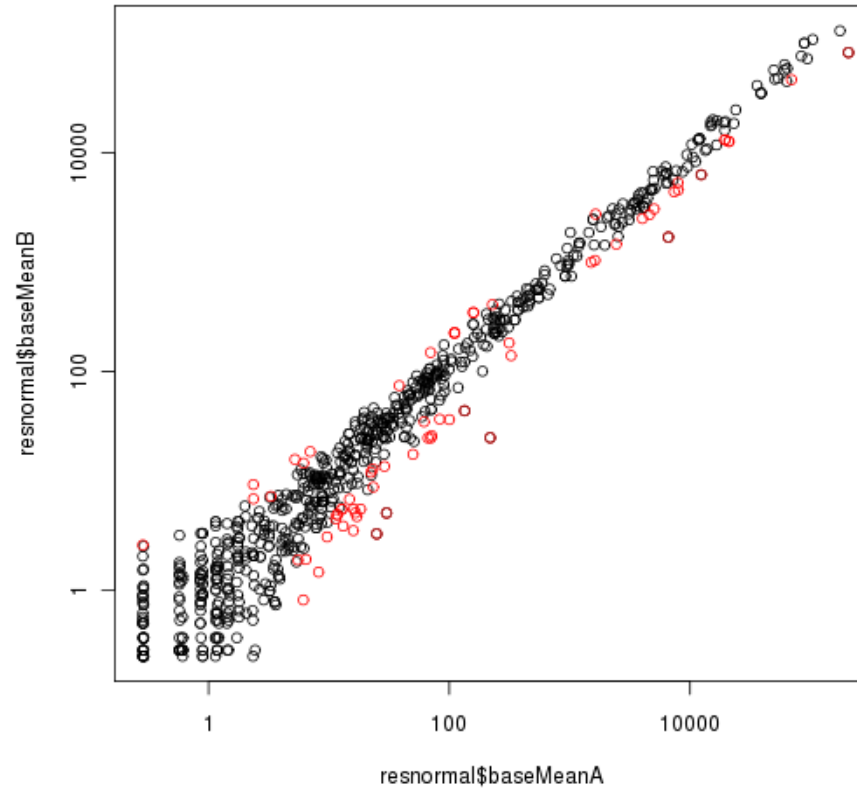| mir	Mean_A	Mean_B	foldChange	Log2 Fold Change	p value	p value adjusted
hsa-mir-1269	416.18	51.48	0.12	-3.015029278	5.95E-15	5.39E-12
hsa-mir-192	7470.37	1907.87	0.25	-1.969211399	1.75E-10	7.91E-08
hsa-mir-891a	35.47	5.40	0.15	-2.713956385	1.38E-07	4.03E-05
hsa-mir-490	29.83	3.80	0.12	-2.969589001	1.78E-07	4.03E-05
hsa-mir-30a	240331.24	93779.96	0.39	-1.357672547	3.97E-07	7.20E-05
hsa-mir-194-2	153.57	49.63	0.32	-1.629491061	9.16E-06	0.001383467
hsa-mir-184	134.27	45.11	0.33	-1.573620723	6.14E-05	0.007945707
hsa-mir-194-1	82.76	28.98	0.35	-1.513708416	0.000101201	0.011461033
hsa-mir-222	14456.12	7372.35	0.50	-0.971484018	0.000234695	0.023625958
hsa-mir-302b	4.09	0	0	=-Inf	0.000416083	0.037535346
hsa-mir-215	88.08	31.72	0.36	-1.473020134	0.000455727	0.037535346
hsa-mir-599	17.79	5.04	0.28	-1.81779395	0.000531559	0.040132729
hsa-mir-876	16.67	3.46	0.20	-2.268892743	0.000820279	0.057167161

AVAILABLE SOFTWARES

- ncPRO-seq (https://ncpro.curie.fr/)
  - Local installation with many dependencies
  - Focus on all ncRNA and the profile of expression
  - Interactive html report and pdf report
  - Galaxy implementation
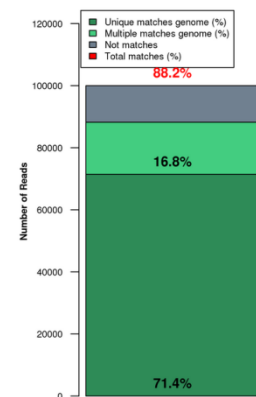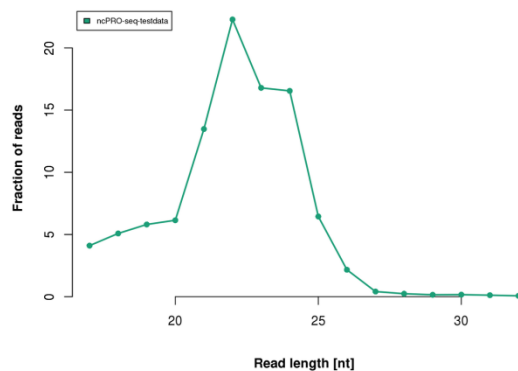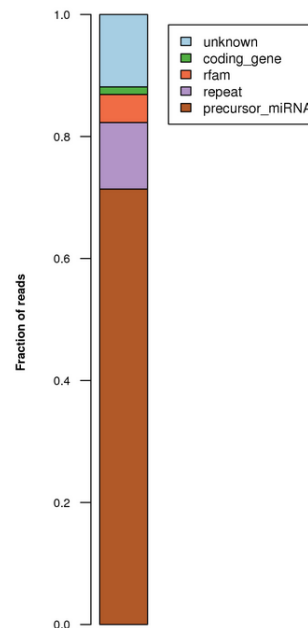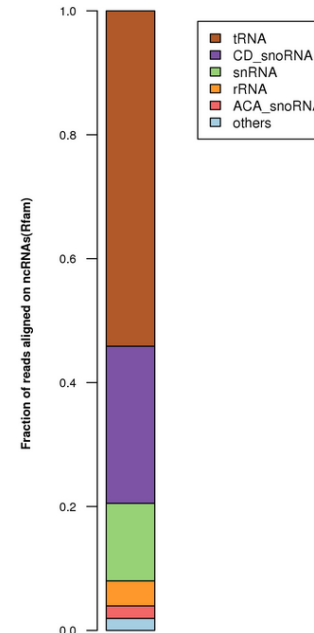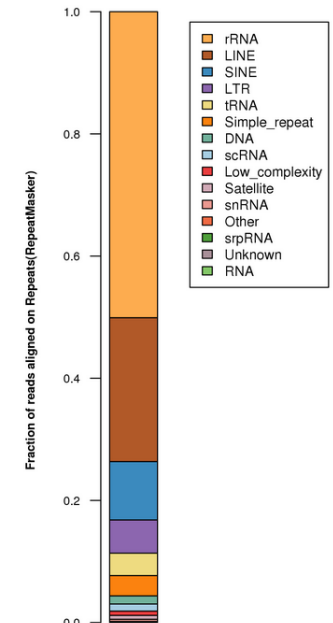
AVAILABLE SOFTWARES
MIRDEEP2

- Command line style, easy installation (external packages included)
- Interactive Html report
- Can do: ⟶
  - **Identify known and novel miRNAs**
  - No species limitation (user have to download a genome and build the genome index himself)

Marc R. FriedIander et al.
Nucleic Acid Research, september 2011

5. Data visualisation

- UCSC Genome browser
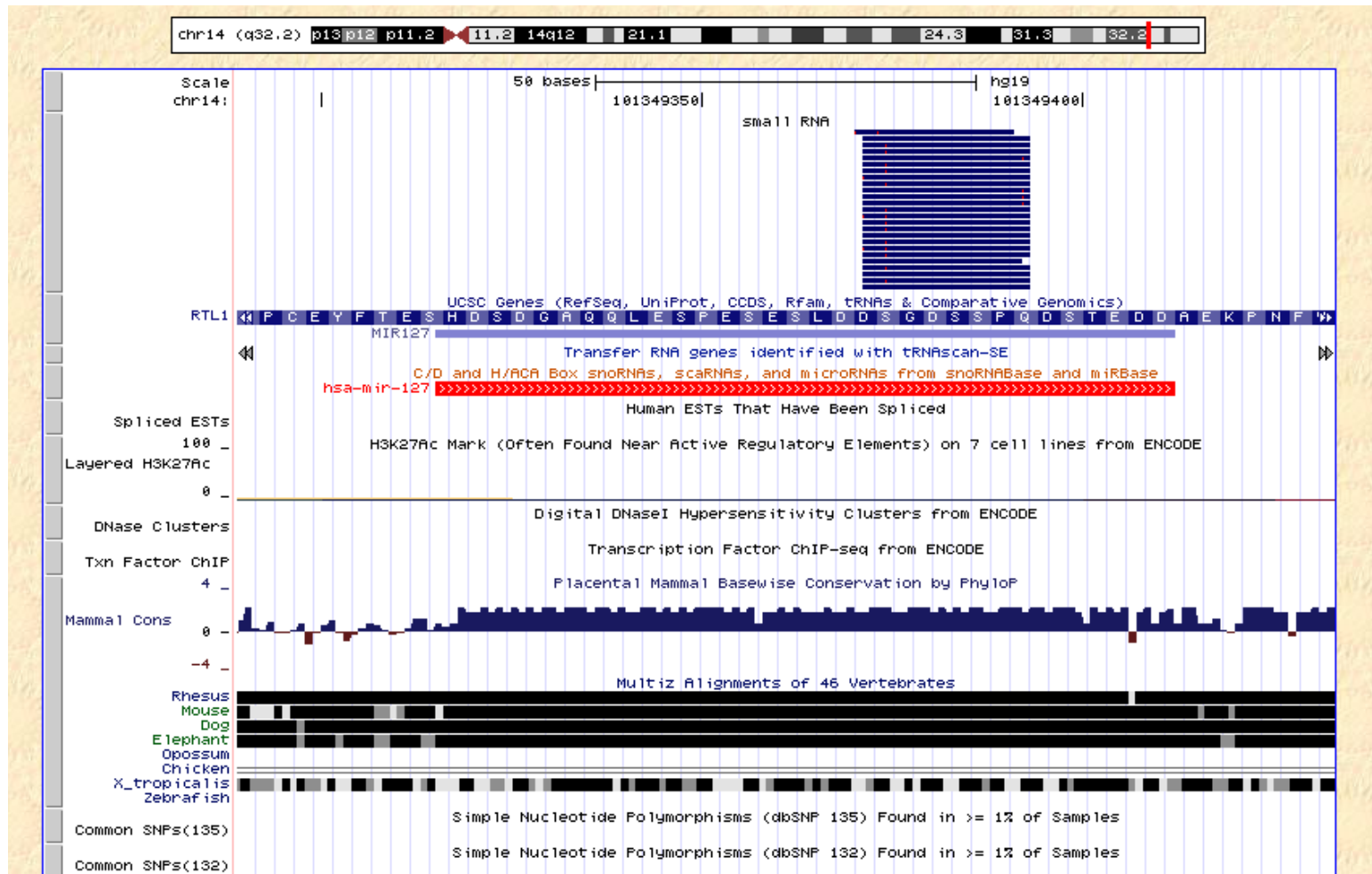
5. DATA VISUALISATION

- IGV Integrative Genomics Viewer

  (James T. et al. Nature Biotech. 2011)

MicroRNA target recognition in animals

Bartel, *Cell* 2009 (review)
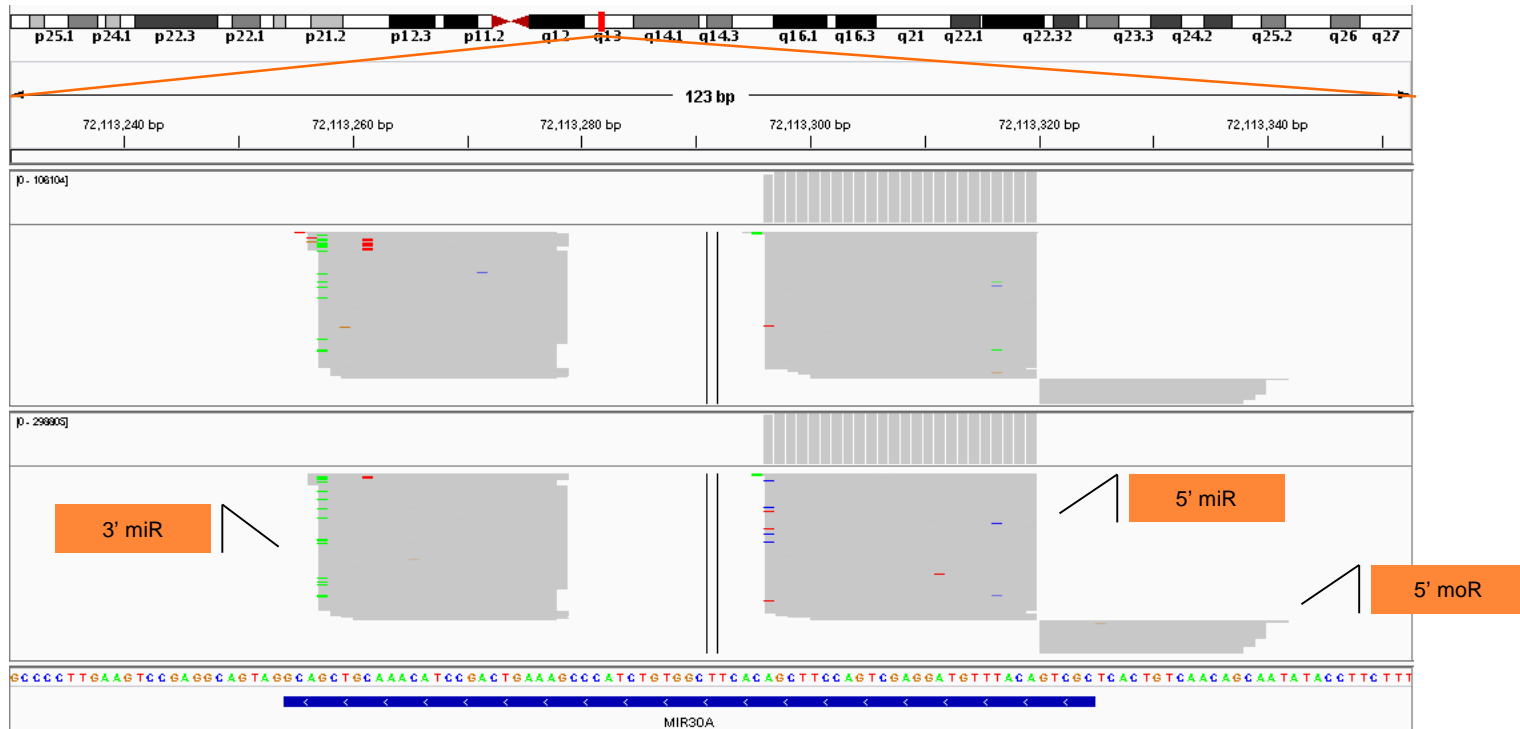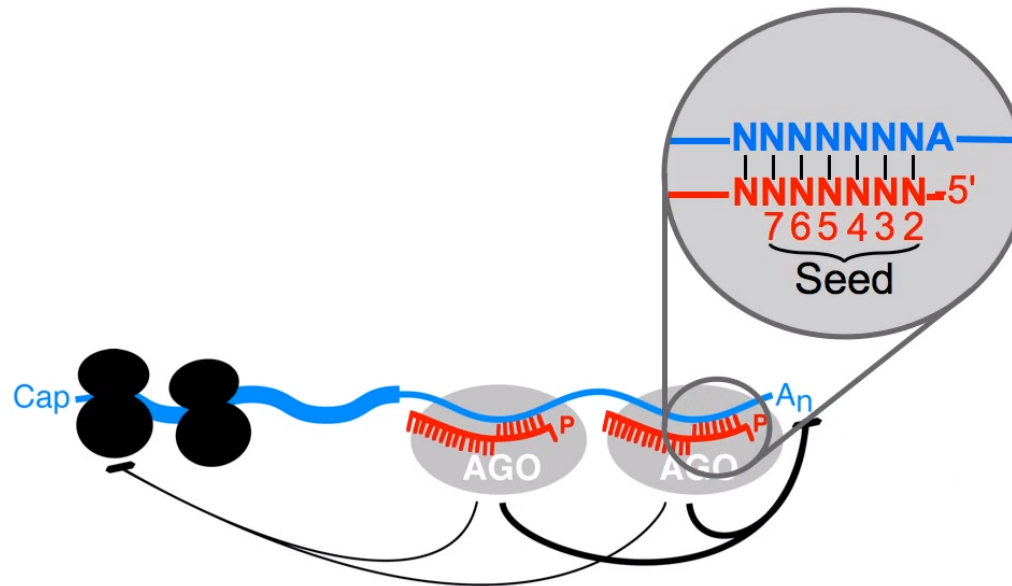
TARGET PREDICTION

- Bioinformatic prediction programs
  - miRanda (John et al., 2004),
  - PicTar (Krek et al., 2005),
  - PITA (Kertesz et al., 2007),
  - RNAhybrid (Rehmsmeier et al., 2004),
  - TargetScan (Lewis et al., 2005)

TargetScanHuman
Prediction of microRNA targets   Release 6.2: June 2012

Search for predicted microRNA targets in mammals   [Go to TargetScanMouse]
                                                        [Go to TargetScanWorm]
                                                        [Go to TargetScanFly]
                                                        [Go to TargetScanFish]

1. Select a species   Human ▼

AND

2. Enter a human Entrez Gene symbol (e.g. "LIN28A")

Many biological targets

>400 conserved targets per miRNA family

4–5 conserved sites per conserved target

TargetScan Web Site
Bartel, *Cell* 2009 (review)

TargetScanHuman
Prediction of microRNA targets Release 7.1: June 2016 Agarwal *et al.*, 2015

Human | let-7-5p/98-5p

1191 transcripts with conserved sites, containing a total of **1365** conserved sites and **242** poorly conserved sites.
Genes with only poorly conserved sites are not shown. [View top predicted targets, irrespective of site conservation]
Table sorted by cumulative weighted context++ score [Sort table by aggregate P_{CT}]
The table shows at most one transcript per gene, selected for being the most prevalent, based on 3P-seq tags (or the one with the longest 3' UTR, in case of a tie). [Download table]

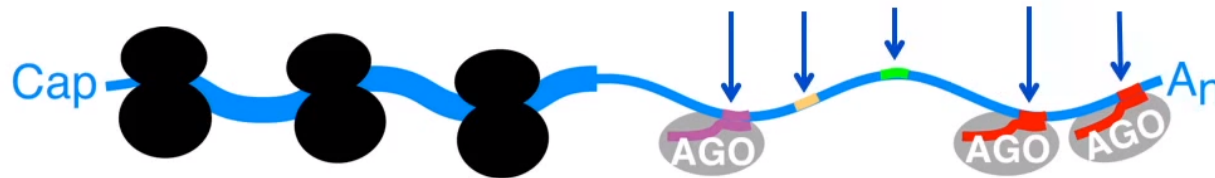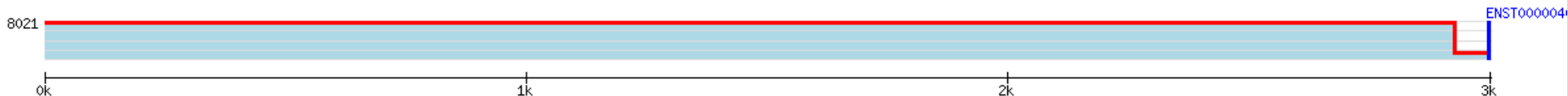| Target gene	Representative transcript	Gene name	Number of 3P-seq tags supporting UTR + 5	Link to sites in UTRs	Conserved sites				Poorly conserved sites				6mer sites	Representative miRNA	Cumulative weighted context++ score
					total	8mer	7mer-m8	7mer-A1	total	8mer	7mer-m8	7mer-A1			
HMGA2	ENST00000403681.2	high mobility group AT-hook 2	8021	Sites in UTR	7	1	3	3	0	0	0	0	1	hsa-let-7d-5p	-2.67
ARID3B	ENST00000346246.5	AT rich interactive domain 3B (BRIGHT-like)	11	Sites in UTR	5	0	1	4	0	0	0	0	0	hsa-let-7d-5p	-1.68
LIN28B	ENST00000345080.4	lin-28 homolog B (C. elegans)	118	Sites in UTR	4	2	2	0	1	0	0	1	0	hsa-let-7d-5p	-1.58
FIGN	ENST00000333129.3	fidgetin	10	Sites in UTR	6	3	0	3	0	0	0	0	0	hsa-let-7d-5p	-1.46
TRIM71	ENST00000383763.5	tripartite motif containing 71, E3 ubiquitin protein ligase	67	Sites in UTR	7	4	1	2	1	0	1	0	0	hsa-let-7d-5p	-1.39
NR6A1	ENST00000487099.2	nuclear receptor subfamily 6, group A, member 1	1082	Sites in UTR	4	2	2	0	1	0	1	0	0	hsa-let-7i-5p	-1.24
THRSP	ENST00000281030.2	thyroid hormone responsive	5	Sites in UTR	1	1	0	0	1	0	0	0	0	hsa-let-7d-5p	-1.04
USP44	ENST00000552440.1	ubiquitin specific peptidase 44	29	Sites in UTR	3	0	0	3	0	0	0	0	2	hsa-let-7d-5p	-1.01
FAM222B	ENST00000582266.1	family with sequence similarity 222, member B	100	Sites in UTR	2	2	0	0	0	0	0	0	2	hsa-let-7d-5p	-1.01
IGDCC3	ENST00000327987.4	immunoglobulin superfamily, DCC subclass, member 3	464	Sites in UTR	3	3	0	0	1	0	1	0	0	hsa-let-7d-5p	-1.01
IGF2BP1	ENST00000290341.3	insulin-like growth factor 2 mRNA binding protein 1	460	Sites in UTR	5	3	1	1	0	0	0	0	1	hsa-let-7d-5p	-0.98

TargetScanHuman
Prediction of microRNA targets

Release 7.1: June 2016   Agarwal *et al.*, 2015

Human HMGA2 ENST00000403681.2   3' UTR length: 3003

8021

ENST000004

0k   1k   2k   3k

Conserved sites for miRNA families broadly conserved among vertebrates

let-7-5p/98-5p   miR-33-5p
miR-9-5p   miR-204-5p/211-5p
miR-15-5p/16-5p/195-5p/424-5p/497-5p
miR-23-3p
miR-196-5p

miR-490-3p   miR-365-3p
miR-190-5p
miR-33-5p
miR-142-3p.1
miR-26-5p
let-7-5p/98-5p

miR-148-3p/152-3p
miR-17-5p/20-5p/93-5p/106-5p/519-3p
miR-132-3p/212-3p

miR-26-5p
miR-196-5p
miR-26-5p   let-7-5p/98-5p
miR-219-5p

let-7-5p/98-5p
let-7-5p/98-5p
let-7-5p/98-5p

miR-103-3p/107
miR-150-5p

let-7-5p/98-5p

miR-25-3p/32-5p/92-3p/363-3p/367-3p

Poorly conserved sites for miRNA families broadly conserved among vertebrates

miR-128-3p
miR-137   miR-129-5p
miR-181-5p
miR-101-3p.2
miR-33-5p
miR-9-5p   miR-142-3p.1   miR-142-5p

miR-202-5p
miR-182-5p
miR-194-5p
miR-22-3p
miR-7-5p   miR-1-3p/206   miR-22-3p   miR-194-5p

miR-302-3p/372-3p/373-3p/520-3p
miR-125-5p

miR-132-3p/212-3p

[Hide poorly conserved sites for miRNA families conserved among vertebrates]
[Show conserved sites for miRNA families conserved only among mammals]
[Show poorly conserved sites for miRNA families conserved among mammals]
[Show sites for poorly conserved but confidently annotated miRNA families]
[Show sites for other miRBase annotations, most of which are miRNA* sequences or RNA fragments misannotated as miRNAs]

[Download SVG image of miRNA sites]
[View table of miRNA sites]
[View human genome browser (hg19)]

[Show all species]

Key:
Sites with higher probability of preferential conservation
■ 8mer   ■ 7mer-m8   ■ 7mer-A1   ■ non-canonical
Sites with lower probability of preferential conservation
■ 8mer   ■ 7mer-m8   ■ 7mer-A1   ■ non-canonical

```
.....1220......1230.............. 1240......1250.......1260........1270.....................1280........1290.......1300........ 1310......1320.........
Human   :AC-UUUUAGCCAGAGAUGC-──AAU--AU-A-UCCCCACUACUC-AAUACUACCUCUGA--AUGUU─────────-ACAACGAAU-U-UA-CAGUCUAGUACUUAUU-ACAUGCUGC──UAUACACAAGCAAUGC───AA──-
Chimp   :AC-UUUUAGCCAGAGAUGC-──AAU--AU-A-UCCCCACUACUC-AAUACUACCUCUGA--AUGUU─────────-ACAACGAAU-U-UG-CAGUCUAGUACUUAUU-ACAUGCUGC──UAUACACAAGCAAUGC───AA──-
Rhesus  :AC-UUUUAGCCAGAGAUGC-──AAU--AU-A-UCCCCACUACUC-AAUACUACCUCUGA--AUGUU─────────-ACAAUGAAU-U-UA-CAGUCUAGUACUUAUU-ACAUGCUGC──UAUACACAAGCAAUGC───AA──-
Squirrel:AC-UUUUAACCAGAGAUGC-──AAU--A──UCCUCACUACUC-AAUACUACCUCUGA--AUGUU─────────-ACAUUGAAU-U-UA-CAGU────CUUAUU-ACAUGCUGC──UAUACACAAGCAAUGC───AA──-
Mouse   :AC-UUUUAGCCAGAGAUGC-──AAU--AA──UCCCCACUCCUC-AAUACUACCUCUGA--AUGCU─────────-ACAGUGAAU-U-UA-CAGCCCUGCACUUGUU-ACACGCUGC──UAGACACAAGCCCUGC───AA──-
Rat     :AC-UUUUAGCCAGAGAUGC-──AAU--AA──CCCCACUACUC-AAUACUACCUCUGA--AUGCU─────────-ACCGUGAAU-U-UA-CAGUCCUGCACUUGUU-ACAGCGUGC──UAGACACAAGCCCUGC───GA──-
Rabbit  :AC-UUUUAGCCAGAGAUGC-──AAU--AU-A-UCCCCACUACUC-AAUACUACCUCUGA--AUGUU─────────-ACAAUGAAU-U-UA-CAGUCUAGUACUUAUU-ACAUGCUGC──UAUACACAAGCAAUGC───AA──-
Pig     :AC-UUUUAGCCAGAGAUGC-──AAU--AC-A-UCCCCACUGCUC-AAUACUACCUCUGA--AUGUU─────────-ACAAUGAAU-U-UA-CAGUUUAGUACUUAUU-ACAUGCUGC──UAUACCCAAGCAAUGC───AA──-
Cow     :AC-UUUUAGCCAGAGAUGC-──AAU--AC-A-AACCCACUGCUC-AAUACUACCUCUGA--AUGUU─────────-ACAAUUAAU-U-UA-CAGUUUAGUACUUAUU-ACAUGCUGC──UAUACACAAGCAAUGC───AA──-
```

CLASH SEQ

- CLASH=Cross-linking ligation and sequencing of hybrids
- Purpose : Identification of sites of RNA-RNA interaction
- Protocol optimized for studying miRNA targets bound by Argonaute proteins
  - « but should be easily adapted for other RNA-binding proteins and class of RNA »
- Hyb: A bioinformatics pipeline for the analysis of CLASH data
  - Travis et al. Methods 2014

CLASH SEQ

Helwak et . Cell 2013 & Nature Protocols 2014
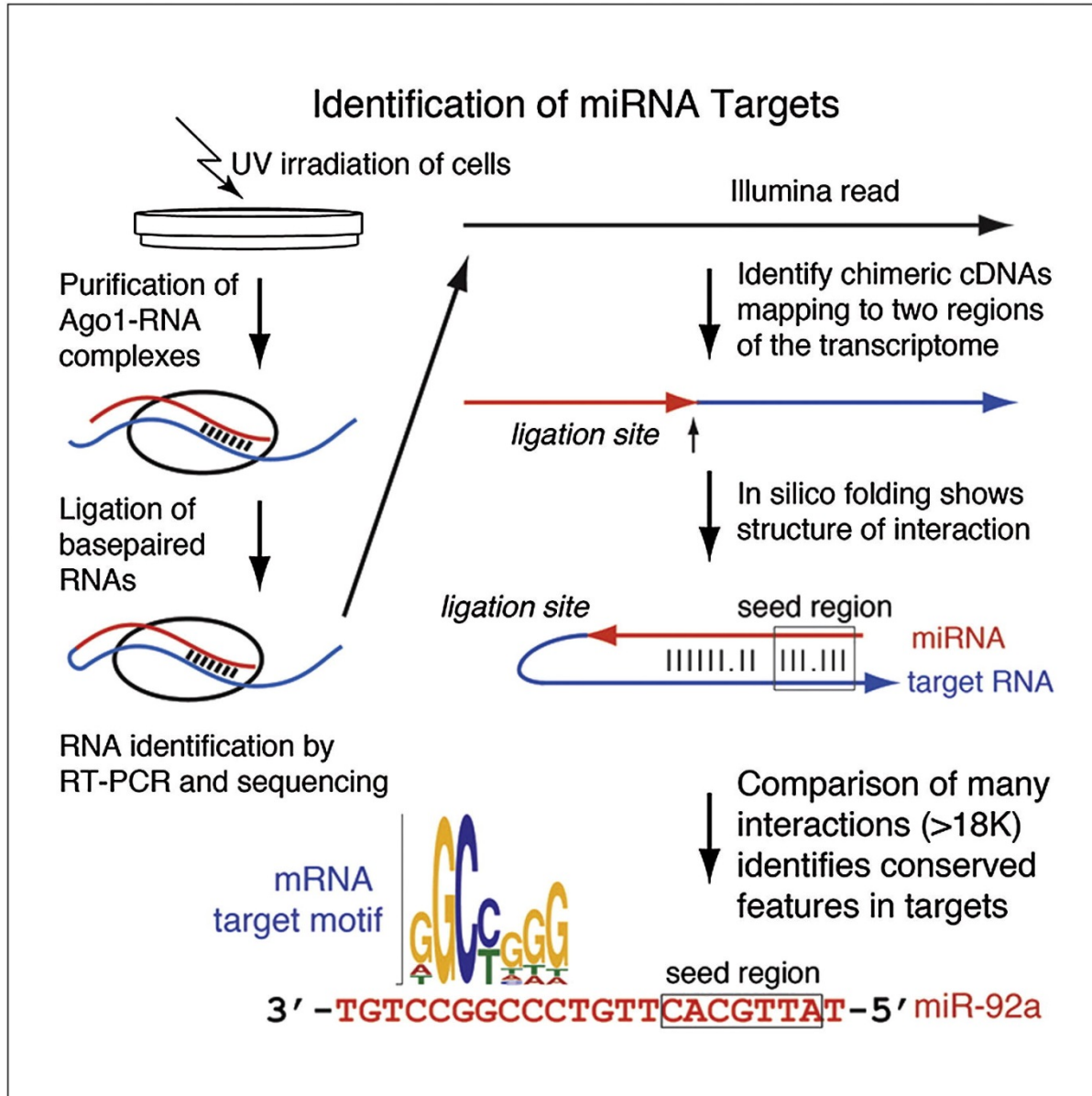
CLASH SEQ RESULT

- ~2% were chimeric reads reflecting intermolecular stem structures present in the AGO1-associated RNAs

- Only around 37% of seed interactions are uninterrupted Watson-Crick base pairing

- Different miRNAs vary in the relative proportions of targets in 5′ UTRs, coding sequences, and 3′ UTRs.
  - miR-100: 4% 5′ UTR: 23% CDS: 73% 3′ UTR
  - miR-149: 8% 5′ UTR: 72% CDS: 19% 3′ UTR

MIRNA-TARGET INTERACTION DATABASES

- miRTarBase, LT, HT (Hsu et al. nar 2011)
  - miRNA-target interactions: 366,181
- TarBase, LT, HT, Pre (Vlachos et al. nar 2014)
  - indexing more than 65,000 miRNA-gene interactions
- starBase 2.0, HT (Li et al. nar 2014)
  - **108 CLIP-Seq (HITS-CLIP, PAR-CLIP, iCLIP, CLASH)**
- miRecords, LT, HT (Xiao et al. nar 2009)
  - 2705 records of interactions between 644 miRNAs and 1901 target genes in 9 animal species.
- …

Ingenuity IPA

- using experimentally validated interactions from **TarBase, miRecords**,
- the peer-reviewed biomedical literature,
- predicted microRNA-mRNA interactions from **TargetScan**
- **Friendly user interface**

IPA RESULT

- Input DE small RNA-seq and RNA-seq data

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	© 2000-2014 QIAGEN. All rights reserved.												
2				micro RNA seq						target mRNAseq			
3	ID	Symbol	Intensity/R	Log Ratio	False Discovery Rate	Source	Confidence	ID	Symbol	Intensity/R	Log Ratio	False Disc	Pathway
4	hsa-miR-100-5p	miR-100-5p (:	23.944	1.046	8.12E-03	TargetScar	Moderate (	ENSG000(	ABO	0.076	-0.079	9.09E-01	
5	hsa-miR-100-5p	miR-100-5p (:	23.944	1.046	8.12E-03	TargetScar	High (predi	ENSG000(	ADCY1	76.988	-0.407	6.33E-05	Breast Cancer
6	hsa-miR-100-5p	miR-100-5p (:	23.944	1.046	8.12E-03	TargetScar	High (predi	ENSG000(	AGO2	578.852	0.162	2.60E-01	EIF2 Signaling
7	hsa-miR-100-5p	miR-100-5p (:	23.944	1.046	8.12E-03	TargetScar	Moderate (	ENSG000(	AIPL1	3.535	0.007	9.97E-01	
8	hsa-miR-100-5p	miR-100-5p (:	23.944	1.046	8.12E-03	TargetScar	Moderate (	ENSG000(	ALG13	320.962	-0.089	5.43E-01	Dolichyl-diphos
9	hsa-miR-100-5p	miR-100-5p (:	23.944	1.046	8.12E-03	TargetScar	Moderate (	ENSG000(	AMMECR	322.052	-0.159	2.28E-01	
10	hsa-miR-100-5p	miR-100-5p (:	23.944	1.046	8.12E-03	TargetScar	Moderate (	ENSG000(	ANKRD28	486.805	-0.058	6.83E-01	
11	hsa-miR-100-5p	miR-100-5p (:	23.944	1.046	8.12E-03	TargetScar	Moderate (	ENSG000(	ANTXR1	11.544	-1.025	1.60E-02	
12	hsa-miR-100-5p	miR-100-5p (:	23.944	1.046	8.12E-03	TargetScar	High (predi	ENSG000(	AP1AR	667.220	-0.118	6.93E-01	
13	hsa-miR-100-5p	miR-100-5p (:	23.944	1.046	8.12E-03	TargetScar	Moderate (	ENSG000(	ARID3A	153.419	0.555	1.35E-02	
14	hsa-miR-100-5p	miR-100-5p (:	23.944	1.046	8.12E-03	TargetScar	High (predi	ENSG000(	ATP11C	176.055	-0.393	1.96E-04	
15	hsa-miR-100-5p	miR-100-5p (:	23.944	1.046	8.12E-03	TargetScar	High (predi	ENSG000(	BAZ2A	754.460	0.162	7.71E-02	
16	hsa-miR-100-5p	miR-100-5p (:	23.944	1.046	8.12E-03	TargetScar	High (predi	ENSG000(	BMPR2	476.280	-0.031	8.47E-01	BMP signaling
17	hsa-miR-100-5p	miR-100-5p (:	23.944	1.046	8.12E-03	TargetScar	Moderate (	ENSG000(	C1orf109	508.515	-0.181	4.94E-02	
18	hsa-miR-100-5p	miR-100-5p (:	23.944	1.046	8.12E-03	TargetScar	Moderate (	ENSG000(	C1orf115	408.227	-0.027	9.08E-01	

IPA RESULT

Pair number	Source
26969	TargetScan Human
92	miRecords
69	"TargetScan Human,miRecords"
57	TarBase
56	"Ingenuity Expert Findings,TargetScan Human"
51	Ingenuity Expert Findings
43	"TarBase,TargetScan Human"
16	"Ingenuity Expert Findings,TargetScan Human,miRecords"
13	"TarBase,TargetScan Human,miRecords"
12	"TarBase,miRecords"
7	"Ingenuity Expert Findings,TarBase,TargetScan Human,miRecords"
1	"Ingenuity Expert Findings,Ingenuity ExpertAssist Findings,TargetScan Human"
1	"Ingenuity Expert Findings,miRecords"
1	"Ingenuity Expert Findings,TarBase,TargetScan Human"

# FUNCTIONS AND REGULATORY MECHANISMS ANALYSES

- Gene Ontology, pathways analyses
  - Need accurate miRNA-RNA interaction data
- Tools (miRNet, 2016 nar)

Table 1. Comparison with other web-based applications (except multimiR) for miRNA target identification and functional analysis

Servers	miRNet	miRPath	miRTar	miRFunction	miRSystem	multimiR
Species #	8	7	4	2	2	2
Target genes						
Experimental	+++	++	-	+	+	++
Predicted	+	+++	+++	+++	+++	+++
Other interactions						
Disease	++	-	-	-	-	+
Small molecule	++	-	-	-	-	+
lncRNA	+	-	-	+	-	-
Epigenetic modifier	+	-	-	-	-	-
Input multiple items						
miRNAs	+	+	+	-	+	+
Targets	+	-	+	-	+	+
Statistical data analysis						
qPCR	+	-	-	-	-	-
Microarray	+	-	-	-	-	-
RNA-seq	+	-	-	-	-	-
Enrichment analysis						
Hypergeometric tests	+	+	+	+	+	-
Empirical sampling	+	+	-	-	-	-
Network visualization	+++	-	-	-	-	-

THANK YOU

- ## GENOMEAST PLATFORM
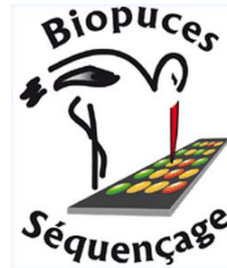
Référent Scientifique
Christelle THIBAULT-CARPENTIER

Ingénieurs & Techniciens
Violaine ALUNNI
Cathy HEROUARD
Marie CERCIAT
Bernard JOST
Gregory AMANN
Romain KAISER
Matthieu JUNG
Céline KEIME
Stéphanie LE GRAS
Damien PLASSARD
Serge VICAIRE
Tao YE

Ingenieur Statisticien
Doulaye DEMBELE