Technology development for integrating *in vivo* miRNA-mRNA targeting and miRNA-seq variants into large-scale cancer projects

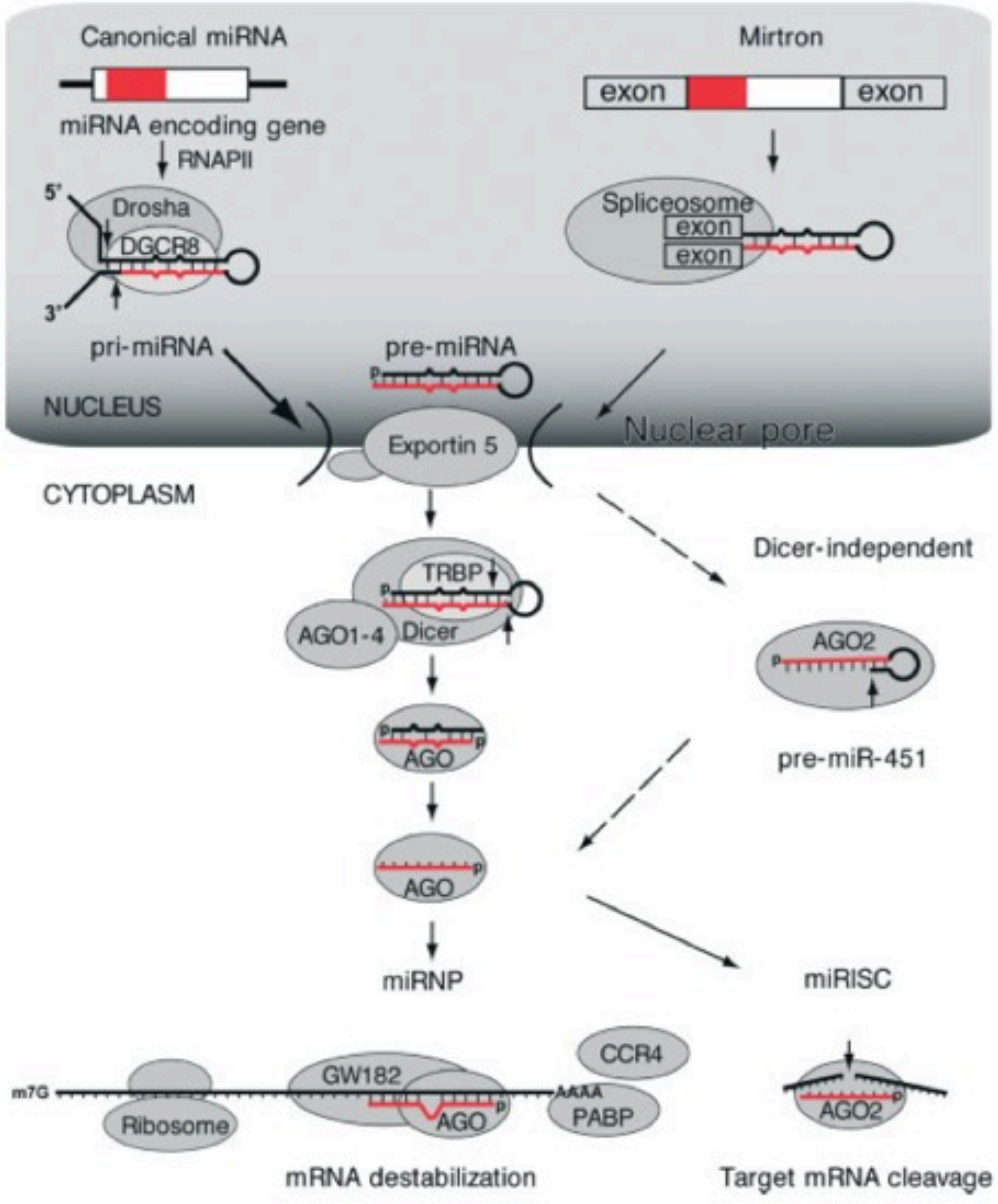
---

Team: Callas

Maria Callas
1923 – 1977
opera singer

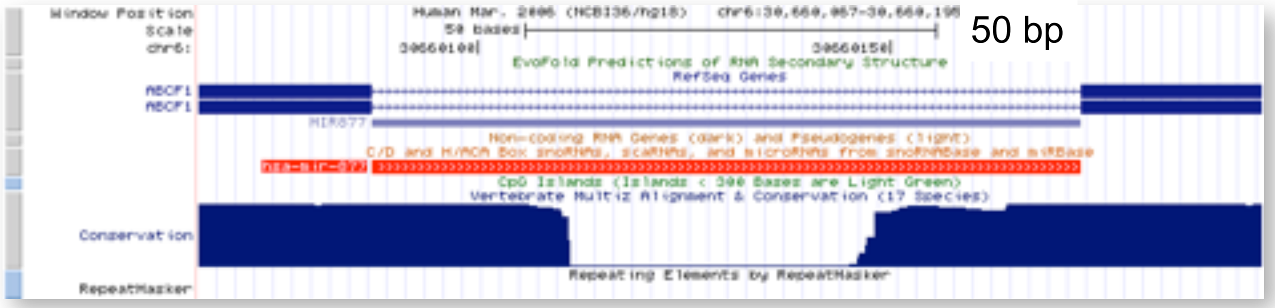
GSC Retreat 2010

miRNAs regulate mRNAs



canonical

mirtron

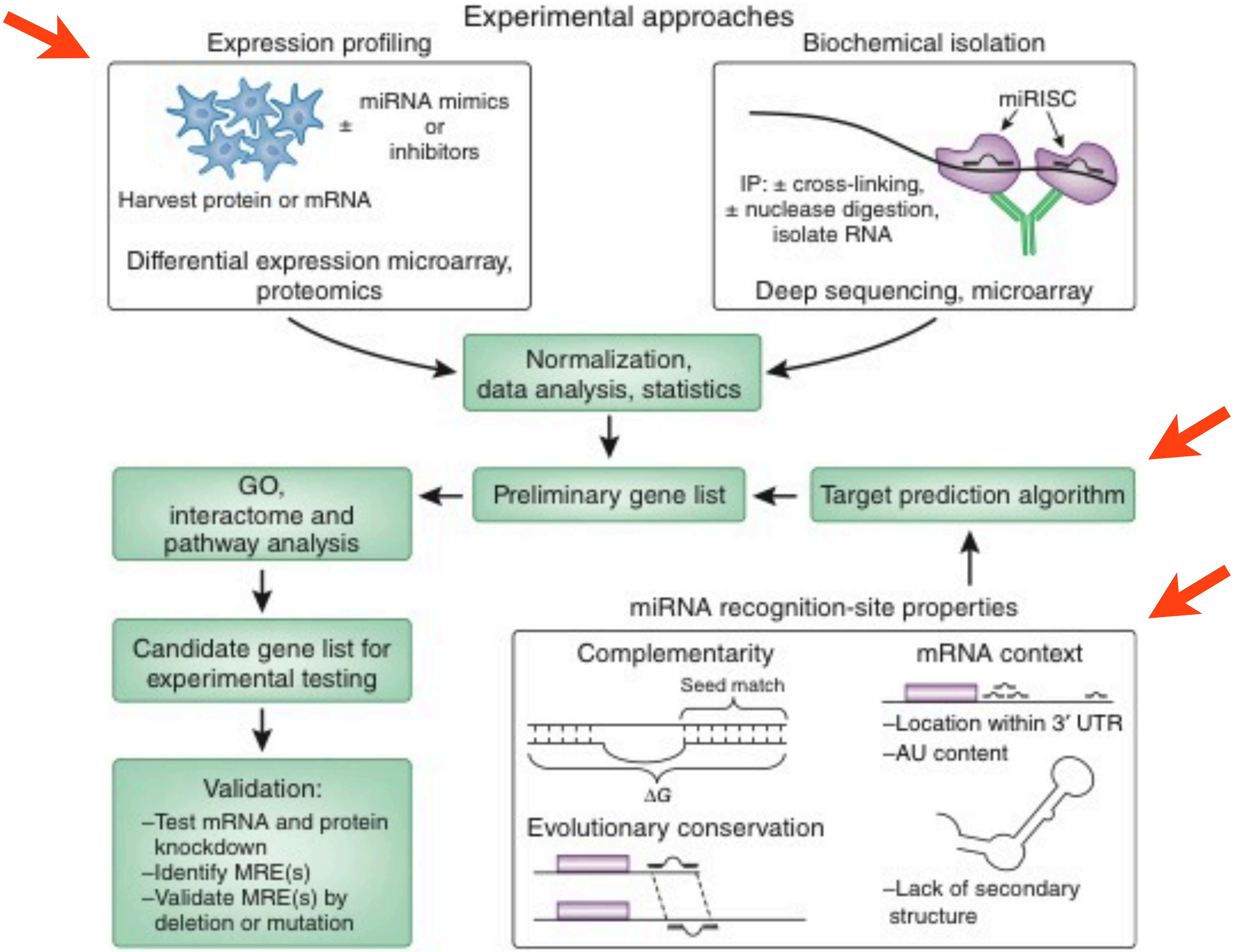


miRBase v16 human mirtrons (14)

hsa-mir-877, hsa-mir-1224, hsa-mir-1225, hsa-mir-1226, hsa-mir-1227, hsa-mir-1228, hsa-mir-1229, hsa-mir-1231, hsa-mir-1233-1, hsa-mir-1233-2, hsa-mir-1234, hsa-mir-1236, hsa-mir-1237, hsa-mir-1238

Farazi et al, miRNAs in human cancer, J Path (2010)

‘Targeting’ is central in work with miRNA & mRNA expression data



Thomas, Nat Struct Mol Biol 2010

Target prediction tools: modest performance, concordance

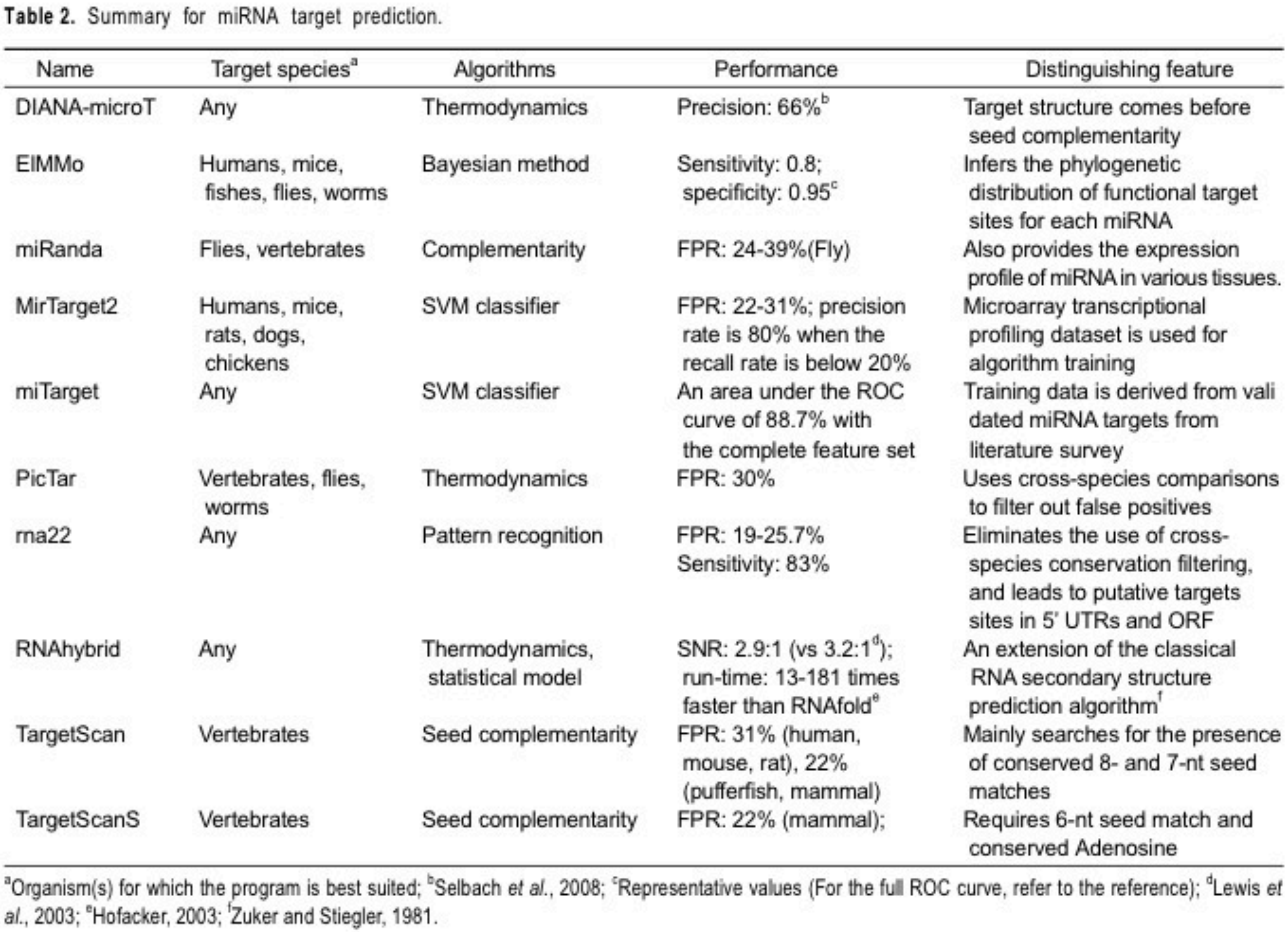


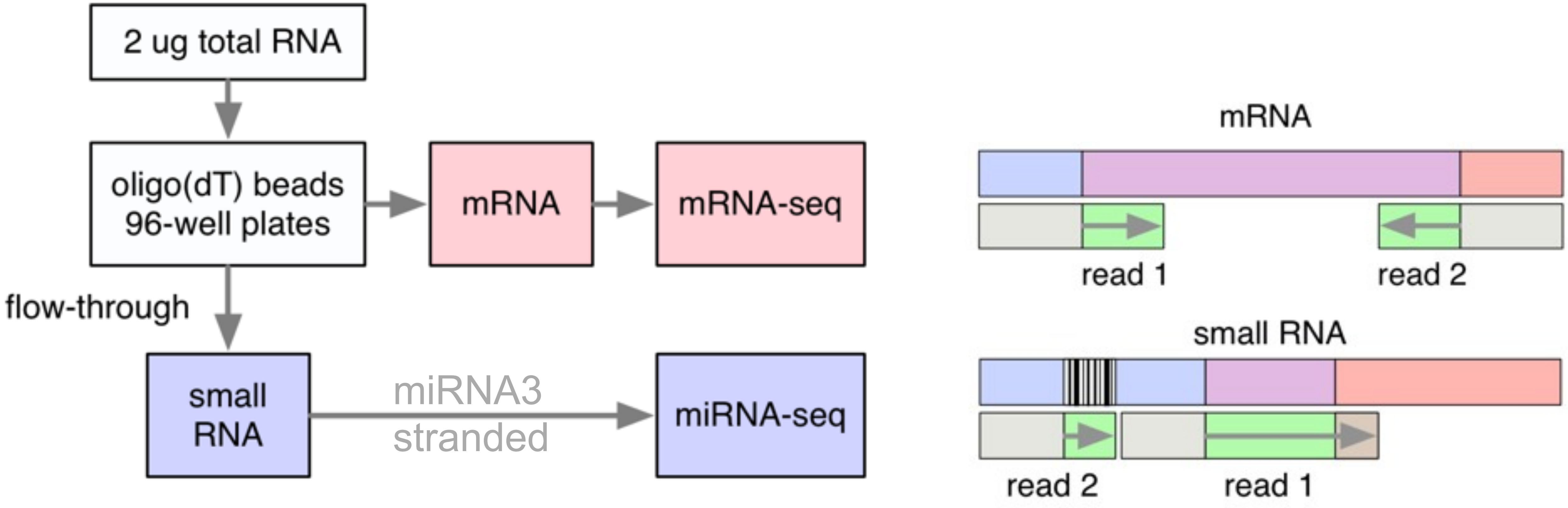
Table 2. Summary for miRNA target prediction.

Name	Target species[a]	Algorithms	Performance	Distinguishing feature
DIANA-microT	Any	Thermodynamics	Precision: 66%[b]	Target structure comes before seed complementarity
ElMMo	Humans, mice, fishes, flies, worms	Bayesian method	Sensitivity: 0.8; specificity: 0.95[c]	Infers the phylogenetic distribution of functional target sites for each miRNA
miRanda	Flies, vertebrates	Complementarity	FPR: 24-39%(Fly)	Also provides the expression profile of miRNA in various tissues.
MirTarget2	Humans, mice, rats, dogs, chickens	SVM classifier	FPR: 22-31%; precision rate is 80% when the recall rate is below 20%	Microarray transcriptional profiling dataset is used for algorithm training
miTarget	Any	SVM classifier	An area under the ROC curve of 88.7% with the complete feature set	Training data is derived from validated miRNA targets from literature survey
PicTar	Vertebrates, flies, worms	Thermodynamics	FPR: 30%	Uses cross-species comparisons to filter out false positives
rna22	Any	Pattern recognition	FPR: 19-25.7% Sensitivity: 83%	Eliminates the use of cross-species conservation filtering, and leads to putative targets sites in 5' UTRs and ORF
RNAhybrid	Any	Thermodynamics, statistical model	SNR: 2.9:1 (vs 3.2:1[d]); run-time: 13-181 times faster than RNAfold[e]	An extension of the classical RNA secondary structure prediction algorithm[f]
TargetScan	Vertebrates	Seed complementarity	FPR: 31% (human, mouse, rat), 22% (pufferfish, mammal)	Mainly searches for the presence of conserved 8- and 7-nt seed matches
TargetScanS	Vertebrates	Seed complementarity	FPR: 22% (mammal);	Requires 6-nt seed match and conserved Adenosine

[a] Organism(s) for which the program is best suited; [b] Selbach *et al.*, 2008; [c] Representative values (For the full ROC curve, refer to the reference); [d] Lewis *et al.*, 2003; [e] Hofacker, 2003; [f] Zuker and Stiegler, 1981.

Min and Yoon, Exptl and Mol Med 2010

Generating miRNA-seq data is inexpensive



2 ug total RNA

oligo(dT) beads
96-well plates

mRNA

mRNA-seq

flow-through

small
RNA

miRNA3
stranded

miRNA-seq

mRNA

read 1

read 2

small RNA

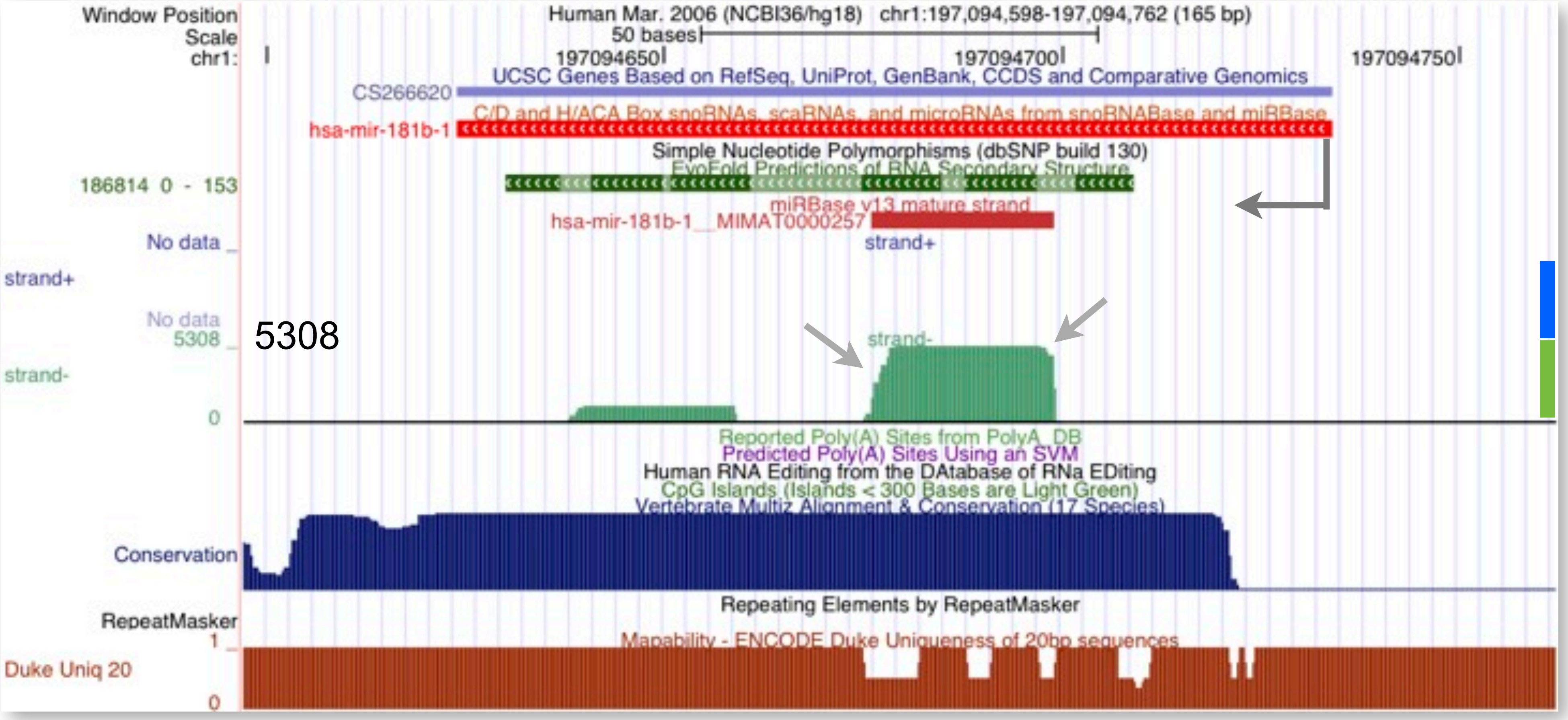
read 2

read 1

Plate-based library construction
Automated gel size selection
Pool 8 indexed samples in one Illumina lane
~$500/sample for library construction, sequencing

Y-J Zhao
Library & engineering teams

miRNA-seq: high spatial resolution, wide dynamic range



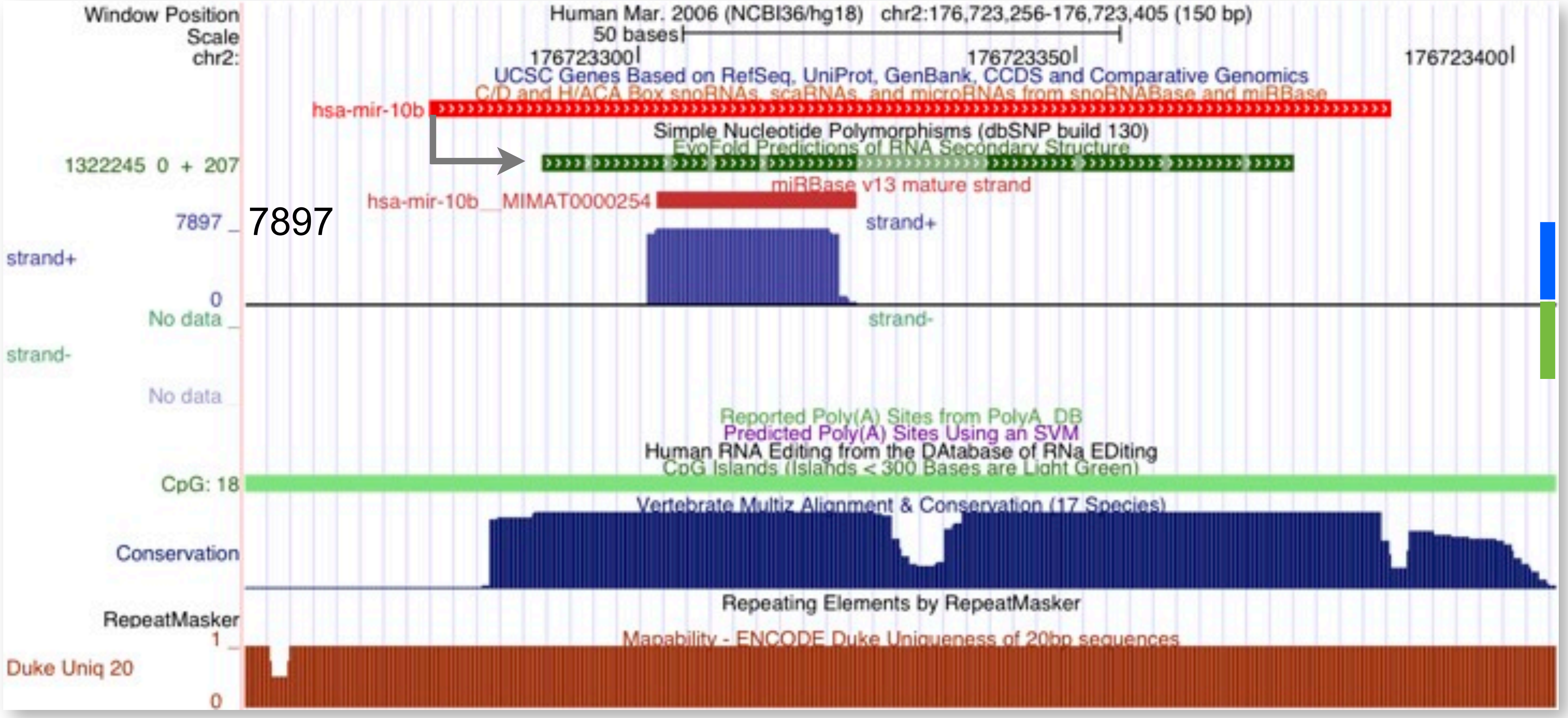
gene
folding
mature strand

pileup (+ strand)

pileup (- strand)

conservation

20-bp mappability



pileup (+ strand)

pileup (- strand)

CpG island

Andy Chu
G Robertson

miRNA-seq reports variants that can affect targeting

Star strands, isomirs

Okamura Nat Struct Mol Biol 2008

Meiri NAR 2010, solid tumors
Starega-Roslan NAR 2010

```
LAML, TCGA-AB-2915-03A-01T-0740-13, hsa-mir-21, 138064 reads
UGUCGGGUAGCUUAUCAGACUGAUGUUGACUGUUGAAUCUCAUGGCAACACCAGUCGAUGGGCUGUCUGACA
       **********************                ********************* 
-------TAGCTTATCAGACTGATGTT----------------------------------------- 10014
-------TAGCTTATCAGACTGATGTTG---------------------------------------- 56040
-------TAGCTTATCAGACTGATGTTGA--------------------------------------- 62921
-------TAGCTTATCAGACTGATGTTGAC-------------------------------------- 3948
---------------------------------------------CAACACCAGTCGATGGGCTGT- 3702

COAD, TCGA-AA-3814-01A-01T-0906-13, hsa-mir-200a, 3038 reads
------CCGGGCCCCUGUGAGCAUCUUACCGGACAGUGCUGGAUUUCCCAGCUUGACUCUAACACUGUCUGGUAACGAUGUUCAAAGGUGACCCGC
               **********************                ********************** 
---------------CATCTTACCGGACAGTGCT------------------------------------------ 125
---------------CATCTTACCGGACAGTGCTG----------------------------------------- 825
---------------CATCTTACCGGACAGTGCTGG---------------------------------------- 1234
---------------CATCTTACCGGACAGTGCTGGA--------------------------------------- 47
----------------ATCTTACCGGACAGTGCTGG---------------------------------------- 37
-----------------------------------------------------TAACACTGTCTGGTAACGAT--- 161
-----------------------------------------------------TAACACTGTCTGGTAACGATG-- 136
-----------------------------------------------------TAACACTGTCTGGTAACGATGT- 129
--------------------------------------------------------CACTGTCTGGTAACGATGT- 81
----------------------------------------------------------CTGTCTGGTAACGATGT- 47
```

AML

colon

mature
star

Andy Chu

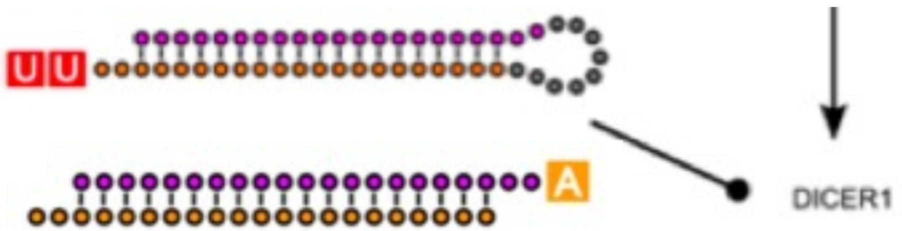
seed

RNA editing

The human pri-miRNA pri-miR-376a-2

Wulff and Nishikura, WIREs RNA 2010
Karahara et al. Science 2007

Sequence additions



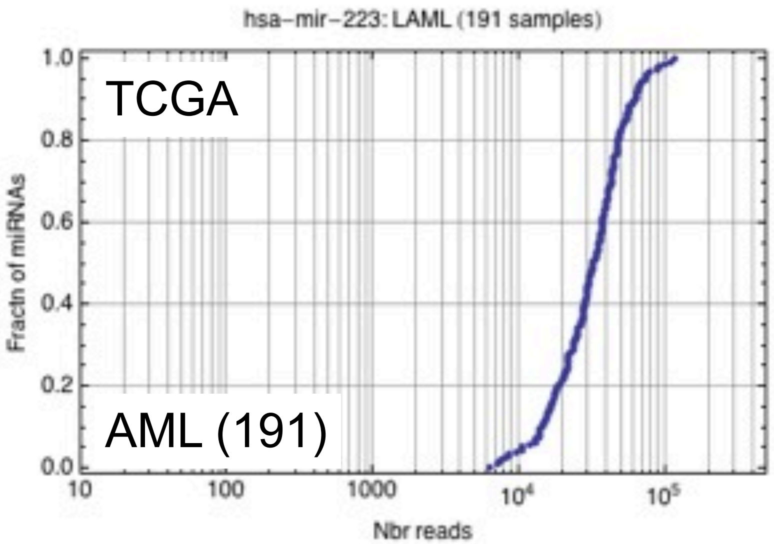
Burroughs et al Gen Res 2010 (AGO IP-seq)
Meiri NAR 2010, solid tumors

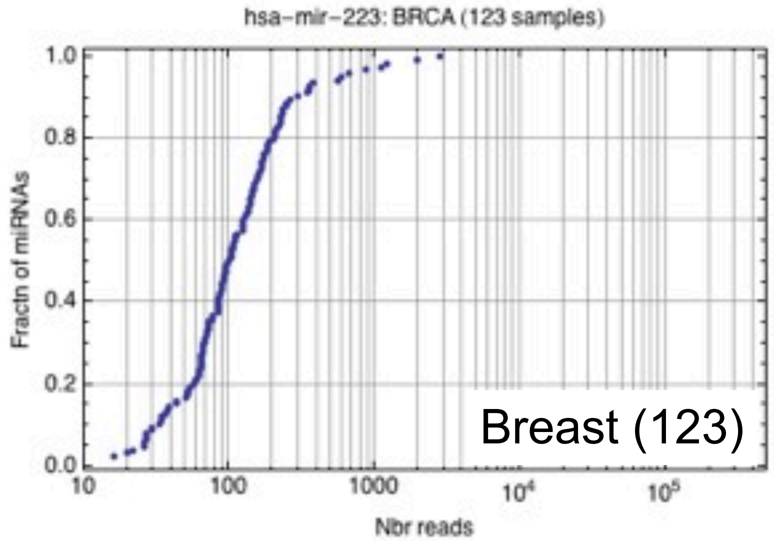
uridylation of precursor
adenylation of DICER1-processed miRNA

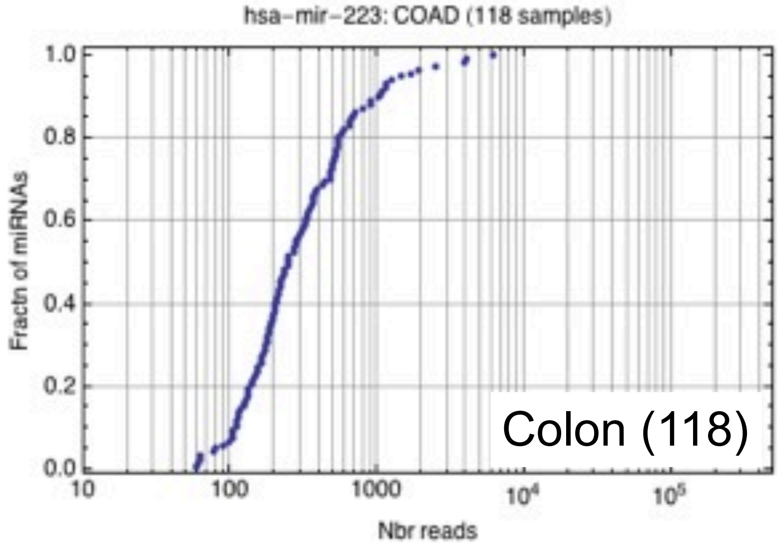
SNVs: biogenesis, targeting

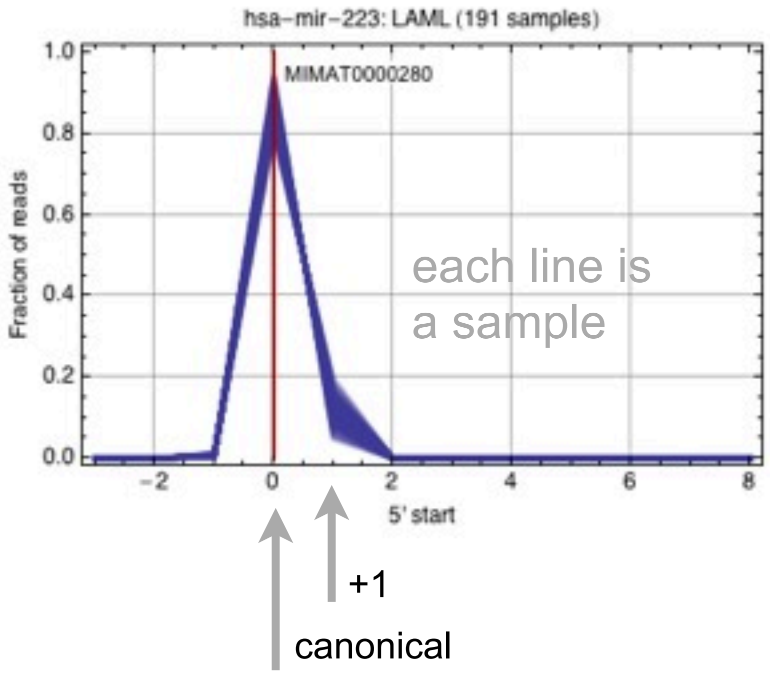
Rotunno M et al. Inherited polymorphisms in the RNA-mediated interference machinery affect microRNA expression and lung cancer survival. Br J Cancer. 2010 Nov 23.

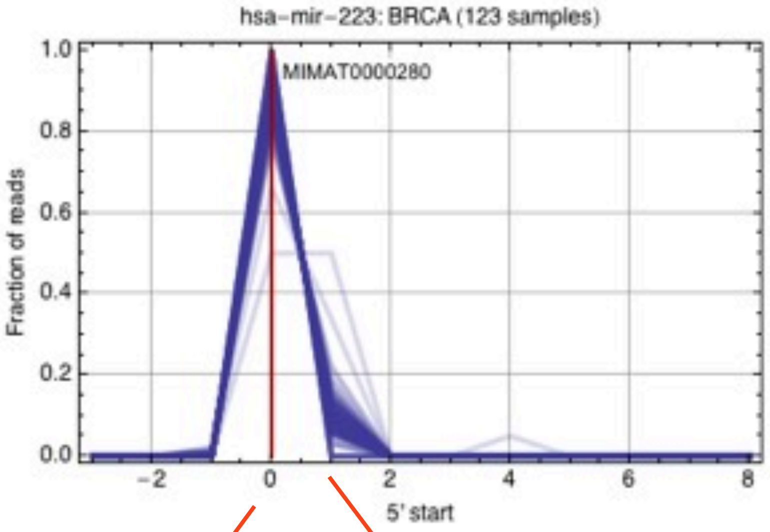
hsa-mir-223 has isomirs with noncanonical 5' starts

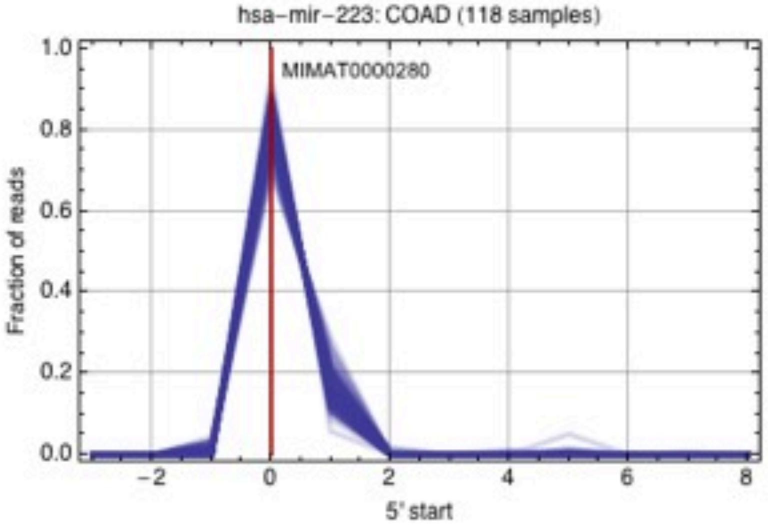






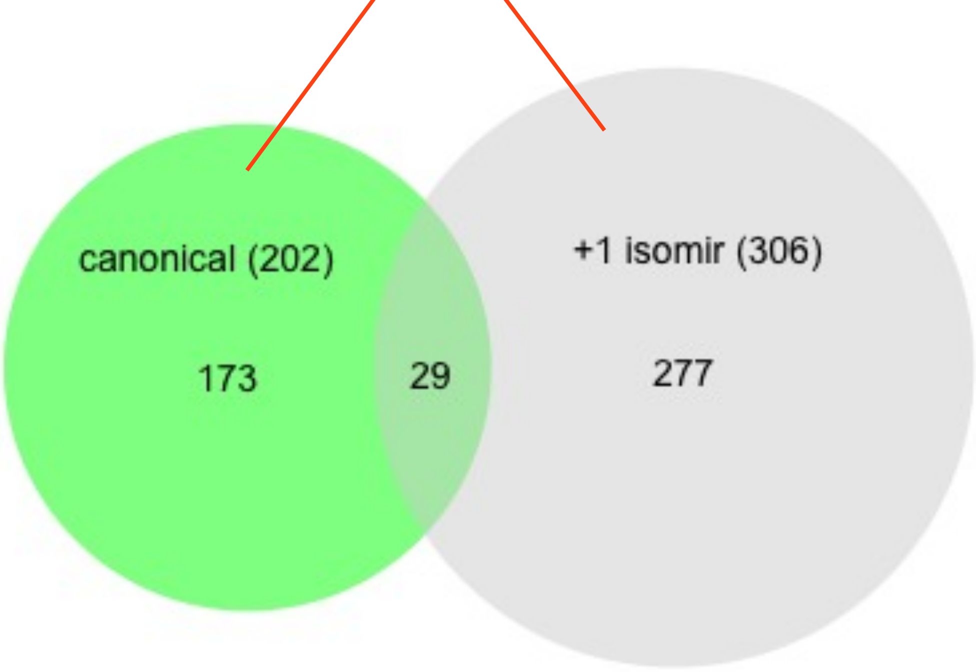






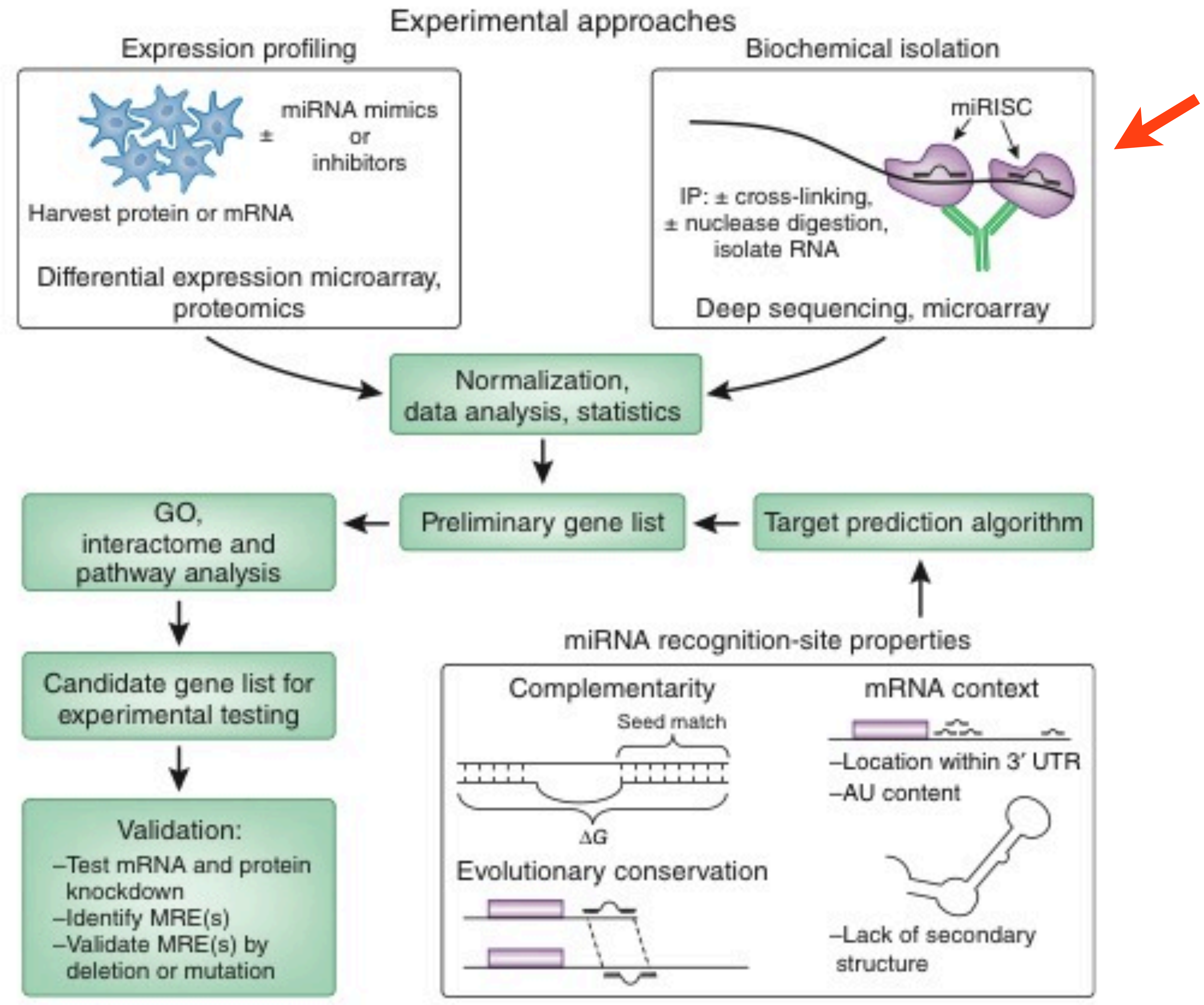
Predicted target genes

TargetScan/Custom



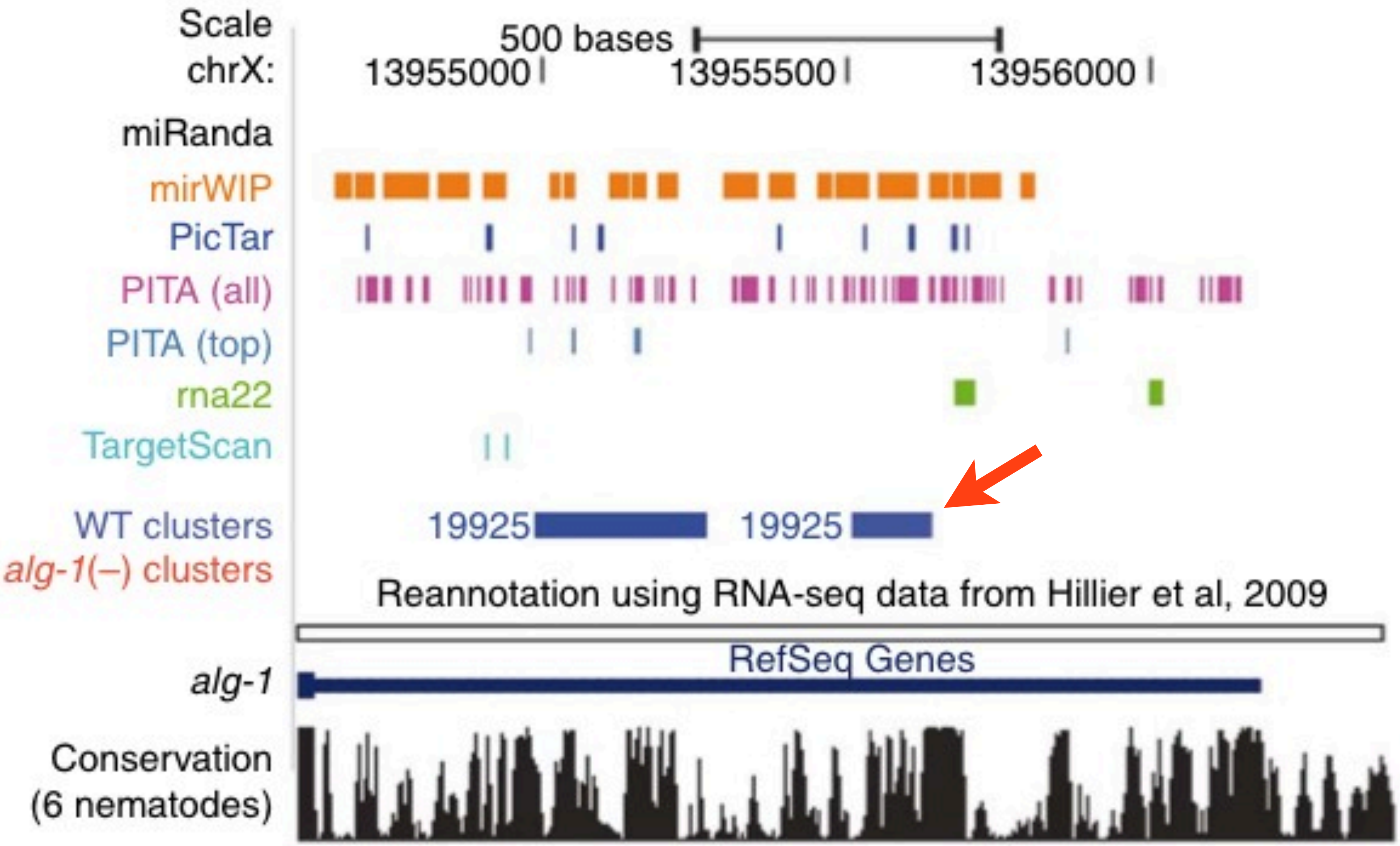
Andy Chu
Gordon Robertson

‘Targeting’ is central in work with miRNA & mRNA expression



Thomas, Nat Struct Mol Biol 2010

RISC-IP returns information on *in vivo* targeting



Reannotation using RNA-seq data from Hillier et al, 2009

Comprehensive discovery of endogenous Argonaute binding sites in *Caenorhabditis elegans*

Dimitrios G Zisoulis[1,3], Michael T Lovci[2,3], Melissa L Wilbert[2], Kasey R Hutt[2], Tiffany Y Liang[2], Amy E Pasquinelli[1] & Gene W Yeo[2]

NATURE STRUCTURAL & MOLECULAR BIOLOGY VOLUME 17 NUMBER 2 FEBRUARY 2010

RISC-IP-seq is stabilizing

Argonaute HITS-CLIP decodes microRNA–mRNA interaction maps

NATURE|Vol 460|23 July 2009

Sung Wook Chi[1], Julie B. Zang[1], Aldo Mele[1] & Robert B. Darnell[1]

Transcriptome-wide Identification of RNA-Binding Protein and MicroRNA Target Sites by PAR-CLIP

Cell *141*, 129–141, April 2, 2010

Markus Hafner,[1,5] Markus Landthaler,[1,4,5] Lukas Burger,[2] Mohsen Khorshid,[2] Jean Hausser,[2] Philipp Berninger,[2] Andrea Rothballer,[1] Manuel Ascano, Jr.,[1] Anna-Carina Jungkamp,[1,4] Mathias Munschauer,[1] Alexander Ulrich,[1] Greg S. Wardle,[1] Scott Dewell,[3] Mihaela Zavolan,[2,*] and Thomas Tuschl[1,*]

The MicroRNA and MessengerRNA Profile of the RNA-Induced Silencing Complex in Human Primary Astrocyte and Astrocytoma Cells

tumor vs. normal

Joanna J. Moser, Marvin J. Fritzler*

Department of Biochemistry and Molecular Biology, Faculty of Medicine, University of Calgary, Calgary, Alberta, Canada

PLoS ONE | www.plosone.org

October 2010 | Volume 5 | Issue 10 | e13445

A comprehensive survey of 3′ animal miRNA modification events and a possible role for 3 ′ adenylation in modulating miRNA targeting effectiveness

1 to $2 \cdot 10^7$ cells

A. Maxwell Burroughs, Yoshinari Ando, Michiel J.L. de Hoon, et al.

Genome Res. 2010 20: 1398-1410 originally published online August 18, 2010

Hypothesis

miRNA variants have noncanonical targets that are important in cancer progression and response to treatment.

Integrating variant data with miRNA-mRNA targeting data will support identifying clinically important patterns.

A White Paper technology development proposal

Aim 1 From well-characterized AML cell lines, identify two lines that have different subtypes and different miRNA isomir profiles.

Aim 2 Identify a RISC-IP approach (from CLIP, PAR-CLIP, RISC-RIP) that is effective in determining *in vivo* targeting in these cell lines, and preferably would be applicable to at least leukemia patient samples, but potentially also to solid tumors.

Aim 3 Validate targeting relationships identified by RISC-IP using miRNA interference and overexpression approaches for specific miRNAs, and, if possible, specific isomirs.

Aim 4 Extend target prediction methods to include isoforms, and optimize the prediction methods to be consistent with the expression and RISC-IP targeting results.

Applying what’s proposed

Apply improved target determination or prediction in large-scale integrated cancer projects.

Replace targeting calculations with RISC-IP-seq.

Use RISC-IP-seq to improve targeting predictions, then use improved targeting predictions.

A bigger picture

stable/underway

ongoing

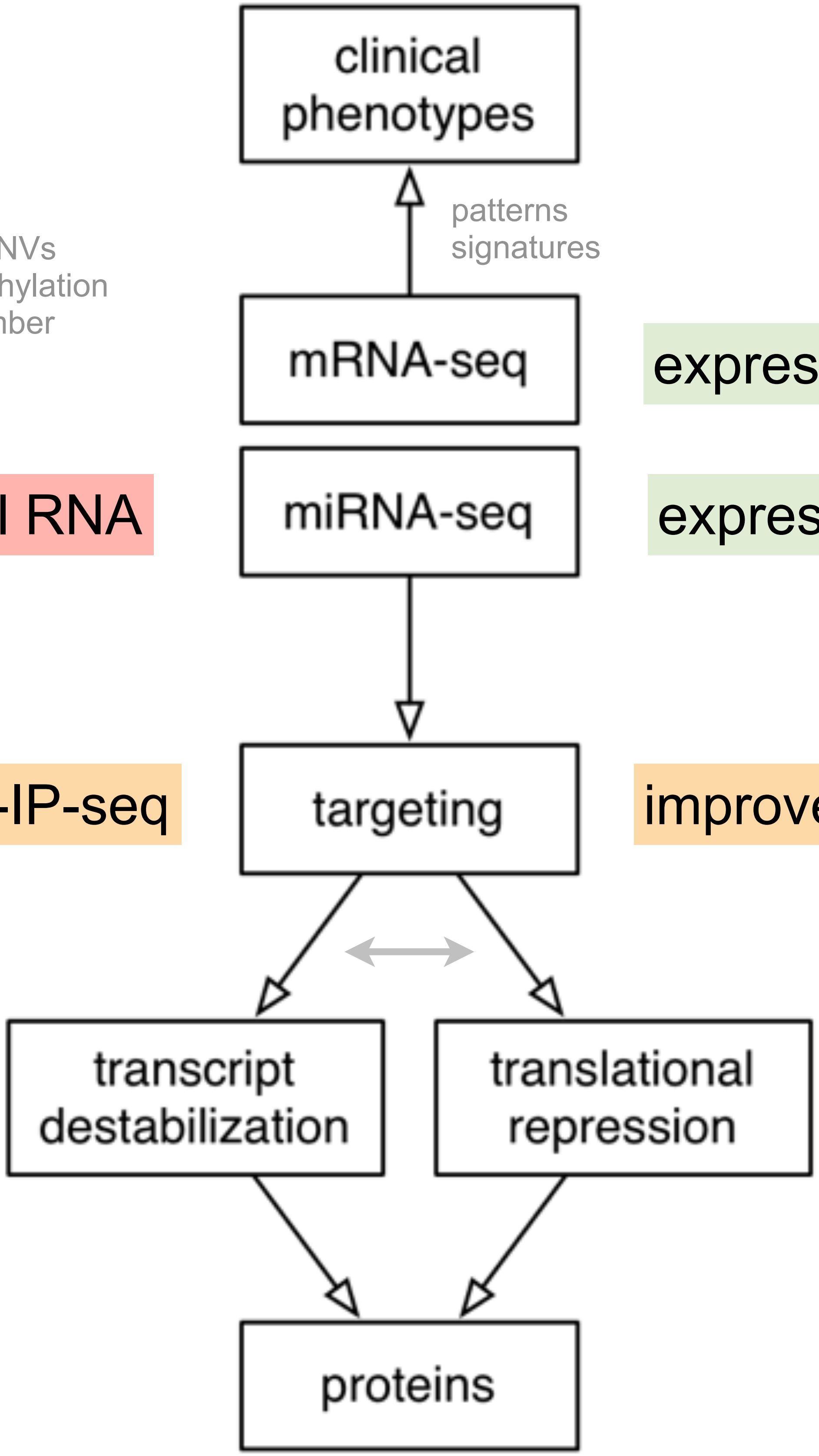
recommended / short term

future

Integrated analyses

Other datatypes

exome/SNVs
DNA methylation
copy number



clinical phenotypes

patterns
signatures

mRNA-seq

expression, variants, alternative 3‘UTRs

other small RNA

miRNA-seq

expression, variants

RISC-IP-seq

targeting

improved targeting

degradome profiling

transcript destabilization

translational repression

ribosome profiling

proteins

SRM/MRM

Acknowledgements

TCGA teams

Andy Chu
Elizabeth Chun
Erin Pleasance
Tom Wang

Trans-ABySS team

Gregg Morin

The many teams that make the GSC what it is