# The roles of miRNA in cancer

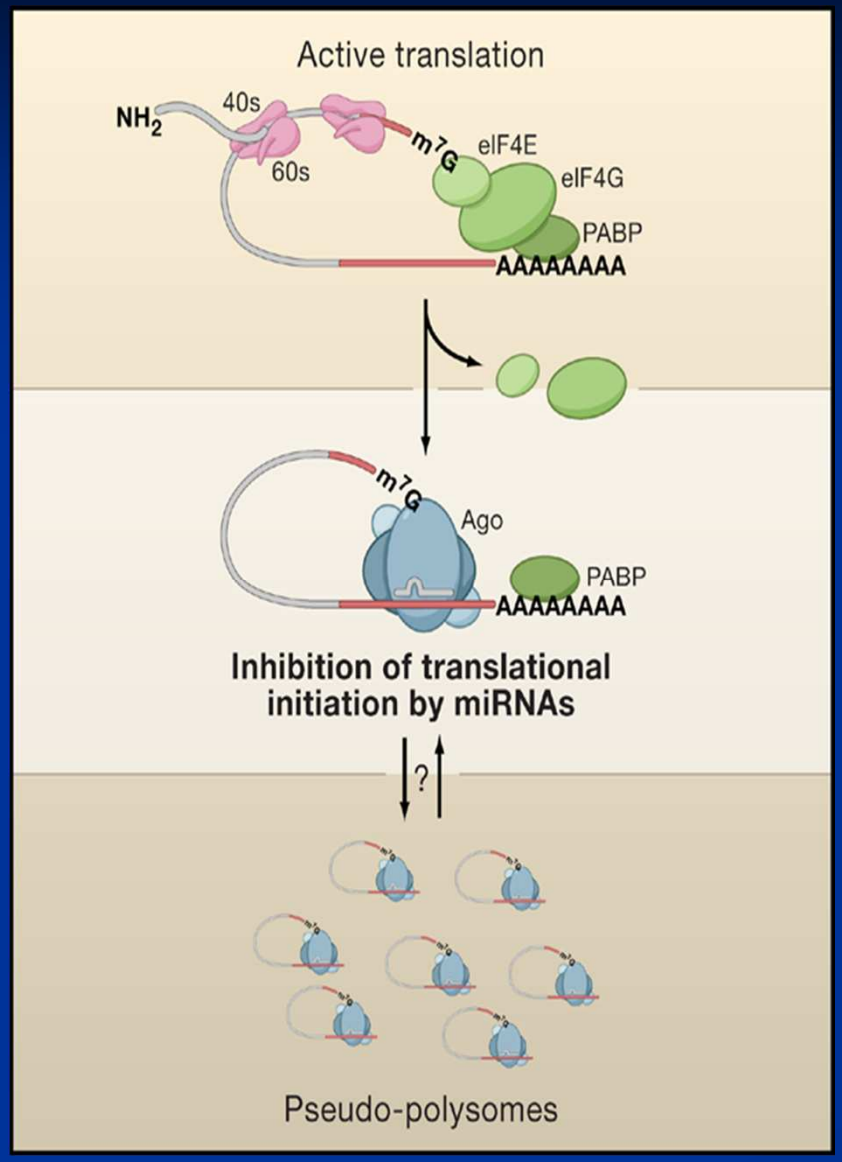
**CAMDA, Vienna**
**July 16th, 2011**

Stefano Volinia,
Università di Ferrara

# Many thanks to all the people involved!

Maria Elena Sana, Marco Galasso, Marco Catozzi, Luca Tagliavini, Giacomo Gamberoni, Simona Rossi, Alessandra Drusco, Jlenia Marchesini, Nicoletta Mascellani, Ramzey Abu Jarour, Caroline Desponts, Mike Teitell, Raffaele Baffa, Rami Aqeilan, Cristian Taccioli, Ramiro Garzon, Gianpiero Di Leva2 Muller Fabbri, Marco Catozzi, Maurizio Previati, Stefan Ambs, Tiziana Palumbo, Curtis C Harris, Rosa Visone, Yuri Pekarsky, Flavia Pichiorri, Nicola Zanesi, George Calin, Chang-Gong Liu, Jeff Palatini, Massimo Negrini, Andrea Vecchione, Anne Rosenberg & Carlo M Croce.

*DAMA - DAta Mining for Analysis of DNA, Dept. Of Morphology and Embryology, Università degli Studi, Ferrara, Italy*
*Comprehensive Cancer Center, Ohio State University, Columbus, OH.*

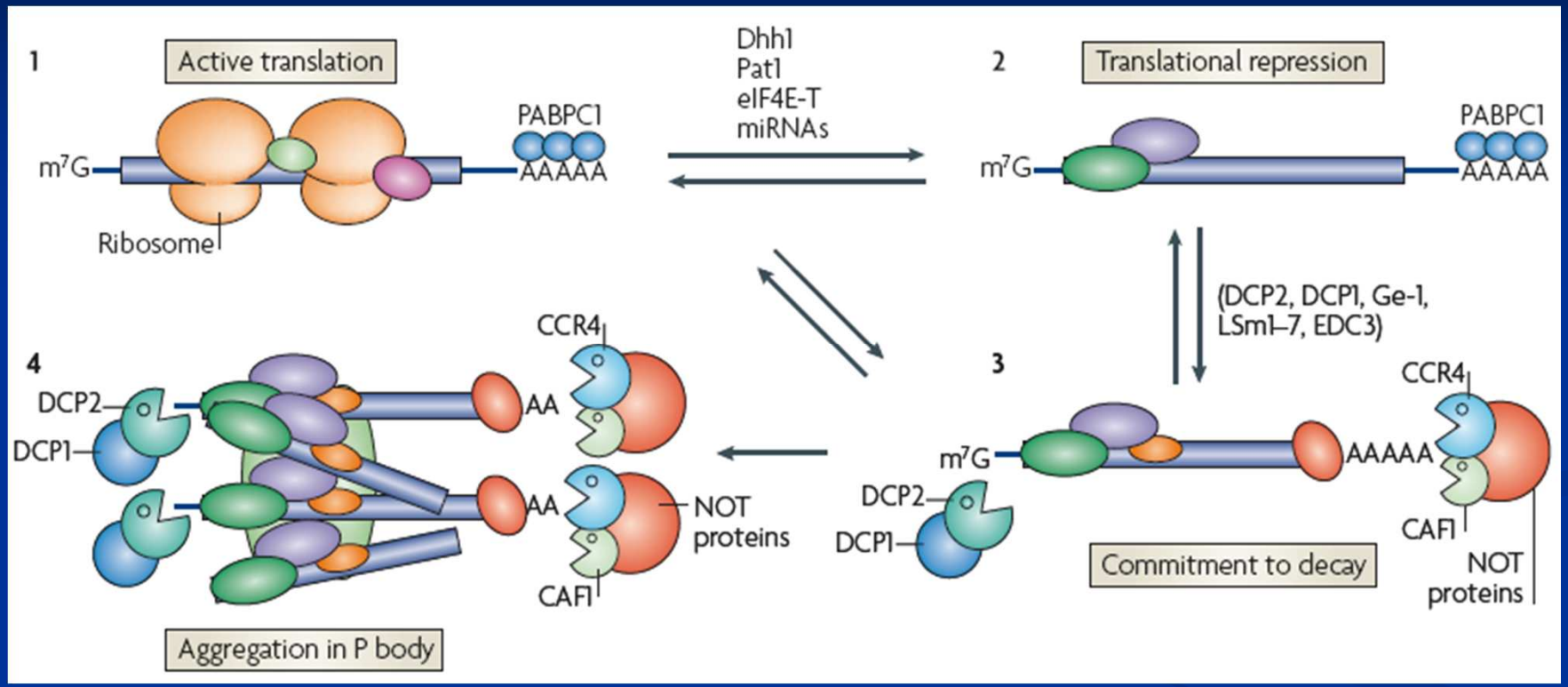



# miRNA-Guided Repression of Translational Initiation

In the absence of miRNAs, the translation initiation factor eIF4E binds to the 7-methyl-guanine (m7G) cap. eIF4G binds to both eIF4E and the poly(A)-binding protein (PABP) and allows for the establishment of a loop, required for translation initiation.

Upon miRNP binding to the 3′ UTR, Ago complex competes with eIF4E for cap binding. The interaction of Ago with the cap releases eIF4E/G and inhibits initiation. (Meister, 2007)

# Translating protein or dumping mRNA

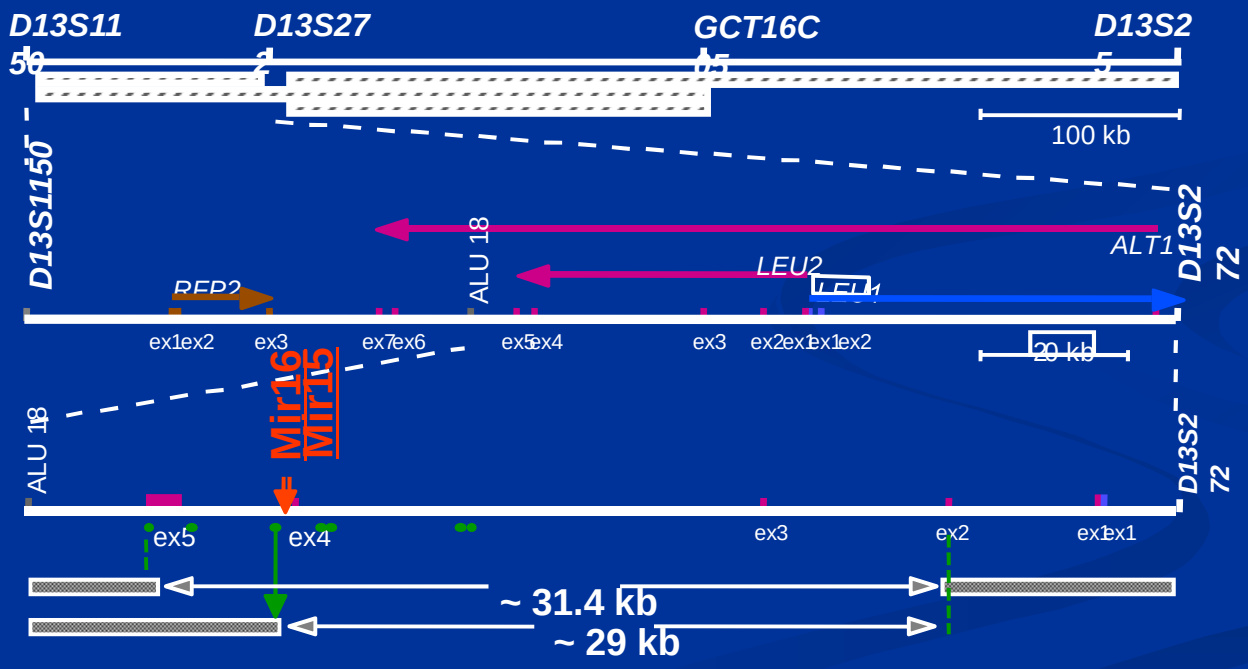



- Messenger ribonucleoproteins (mRNPs) can be in a translationally active state that is associated with polysomes (**1**) or in an inactive state (**2–4**) as those bound to miRNAs.

# Implications in cancer

The prototype: Chronic lymphocytic leukemia (CLL), the most frequent leukemia in adults in the Western world, is characterized by ***predominantly non-dividing*** malignant CD5(+) B cells over-expressing the anti-apoptotic Bcl2 protein

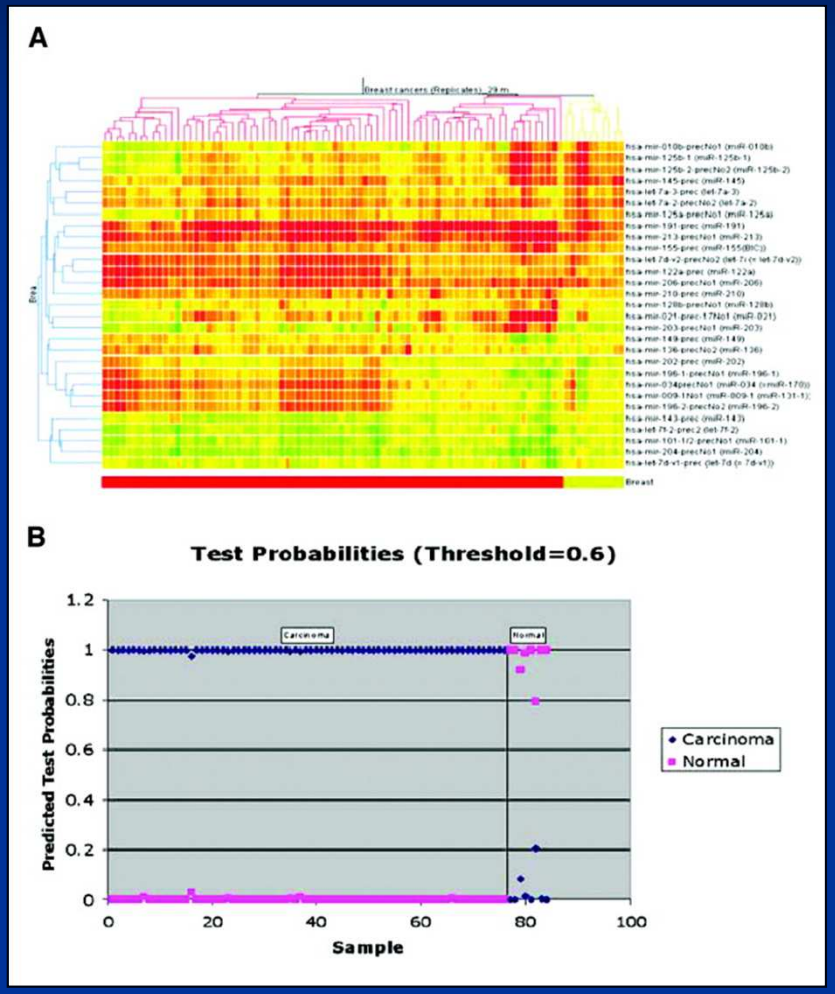
**Discovery:** 13q14.3 deletions in CLL span miR-15/16 locus

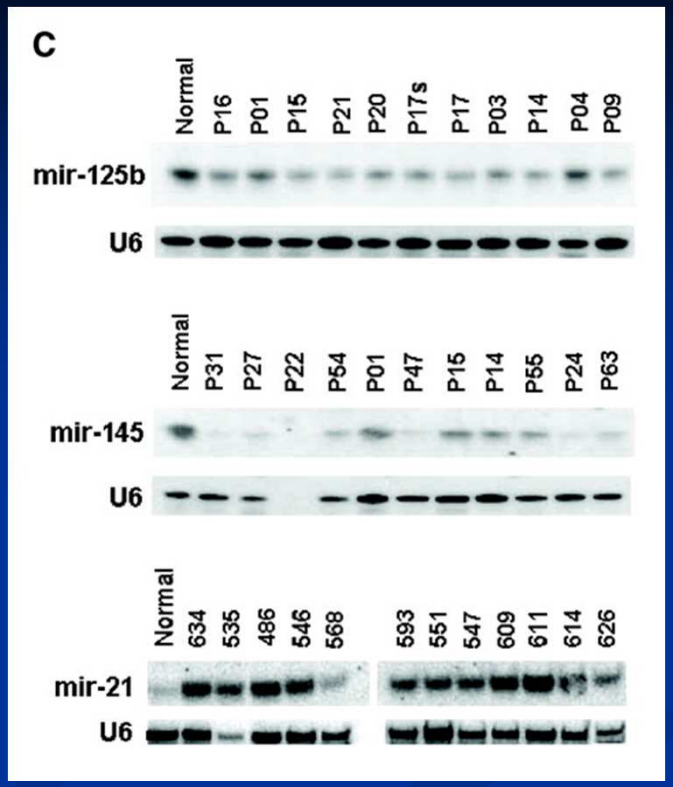


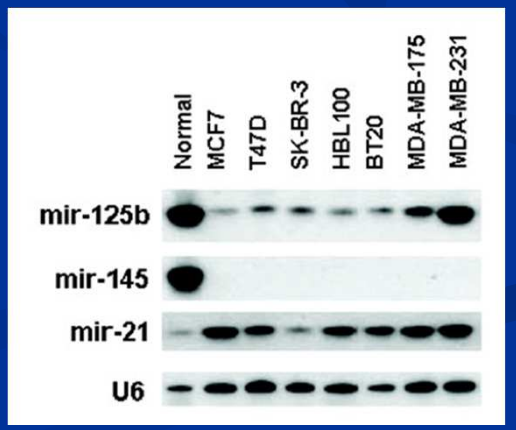
# miRNA microarrays

- Prof. Croce's lab developed the first microarray based platform for assay of miRNAs.
- Platform: one-colour chip (similar to Codelink system), spotted oligonucleotides, precursor and mature miRNAs.
- Versions 1 (2004) , 2 (2005), 3 (2006), 4 (2007) with increasing numbers of miRNAs.
- Chip ->Image analysis (Axon) – data extraction (Gene Pix Pro)
- Parsing of data in microRNA database (over 20,000 chips)
- Statistical analysis/Visualization
- Publication /Submission to public repository

# Breast Cancer



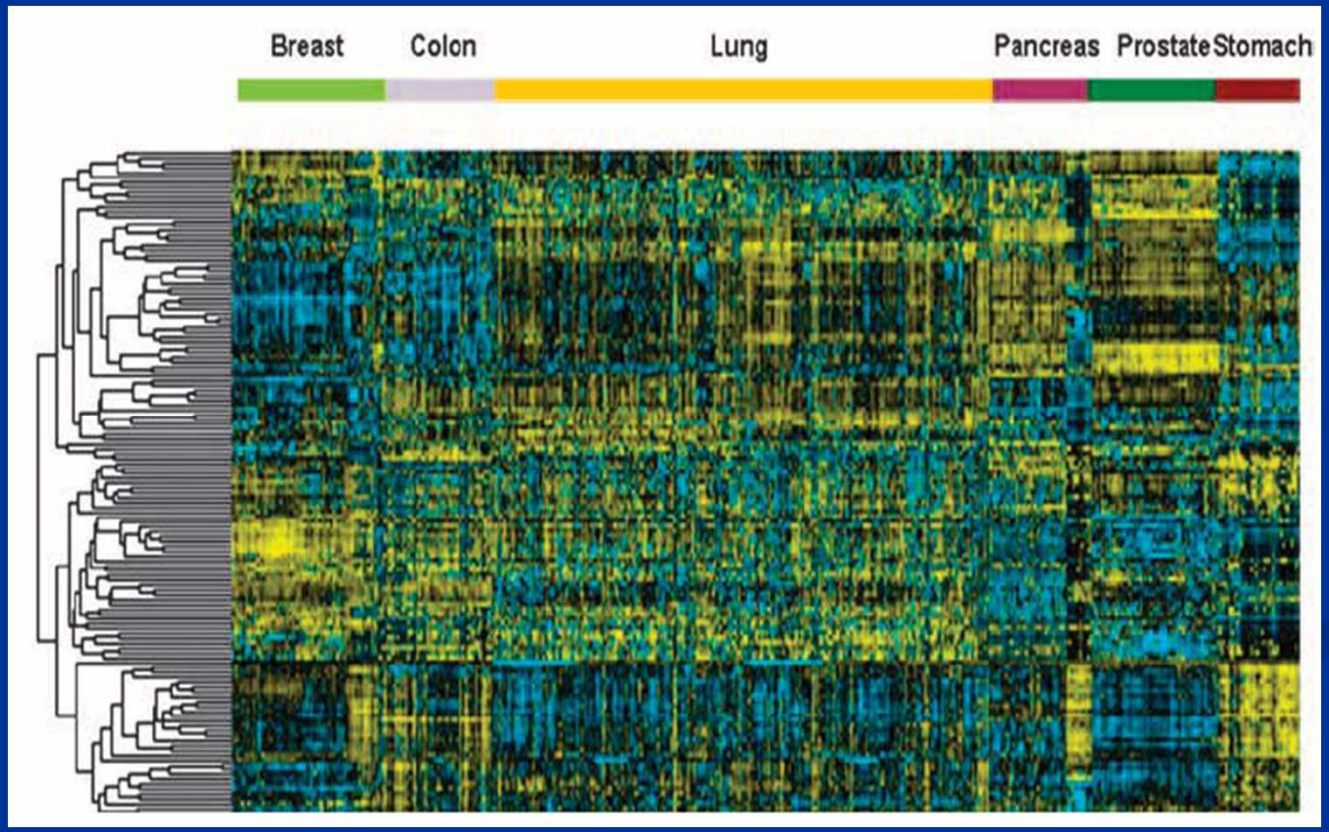



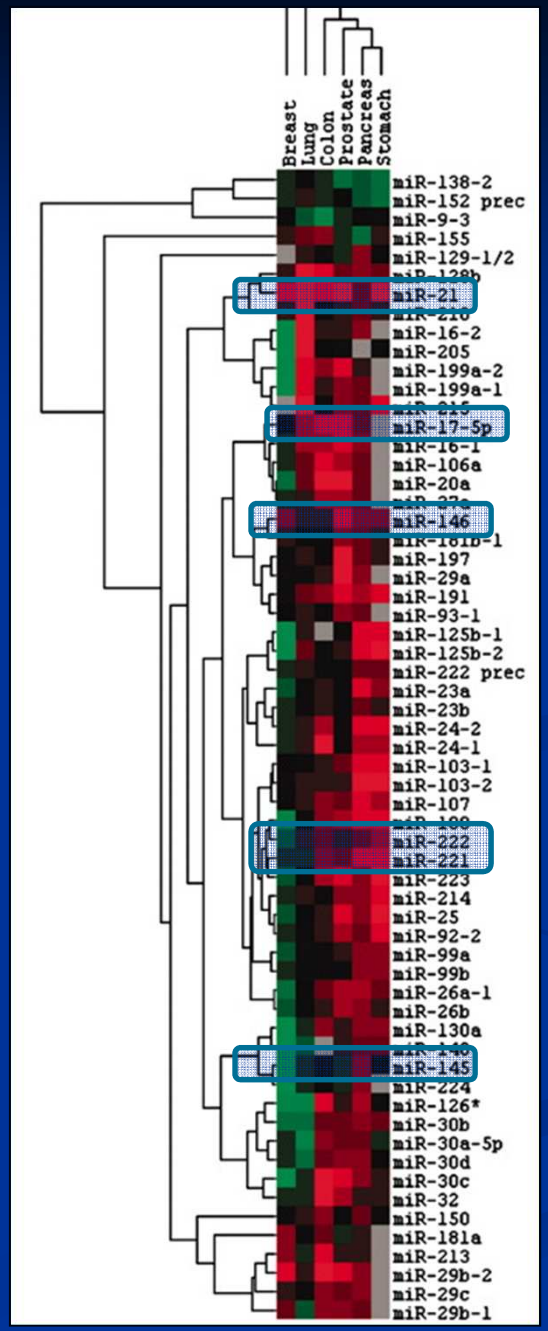

Iorio, M. V. et al. Cancer Res 2005

# Comprehensive analysis of 6 solid cancers

- Five hundred and forty samples, including a total of 363 primary tumors and 177 normal tissues, were used.
- The solid cancers represented were lung carcinoma, breast carcinoma, prostate carcinoma, stomach carcinoma, colon cancer, and pancreatic endocrine tumor.



**Clustering of six solid cancers by miRNA expression (year 2005)**



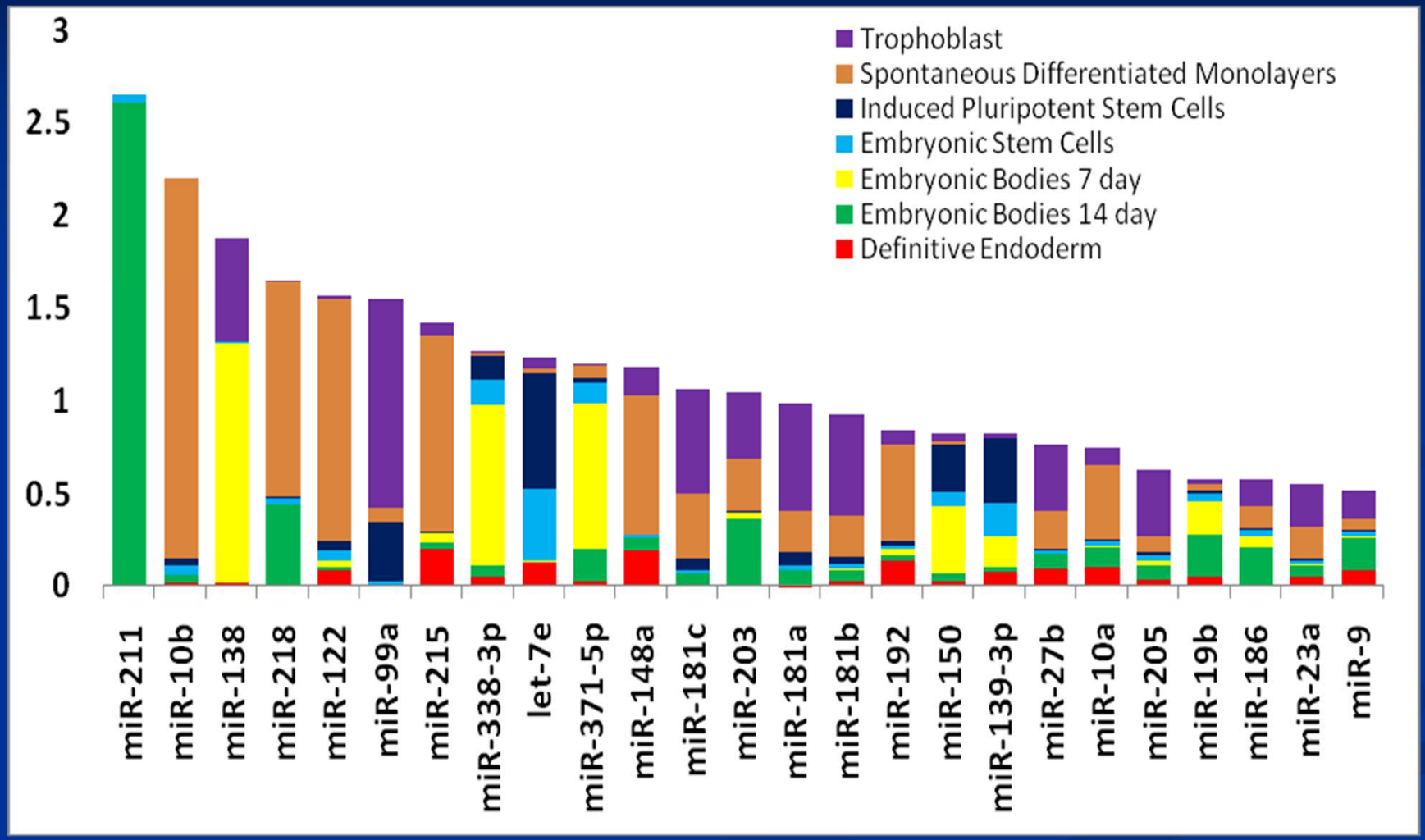

Average
Fold change:

Cancer/Normal

# Differential Expression of microRNA in solid cancers
## -year 2009-
## (2505 solid cancer samples vs. 806 normal samples)

| FDR | Geom mean of intensities in Solid Cancers | Geom mean of intensities in Normal Tissues | Fold-change | microRNA |
|---|---|---|---|---|
| < 1e-07 | 967.0 | 617.9 | 1.57 | miR-21 |
| < 1e-07 | 1378.2 | 917.7 | 1.50 | miR-25 |
| < 1e-07 | 902.7 | 626.3 | 1.44 | miR-20a |
| < 1e-07 | 925.7 | 646.9 | 1.43 | miR-17 |
| < 1e-07 | 652.3 | 469.0 | 1.39 | miR-106a |
| < 1e-07 | 410.0 | 297.8 | 1.38 | miR-106b |
| 1.08e-05 | 918.8 | 697.9 | 1.32 | miR-146a |
| < 1e-07 | 11893.3 | 9370.9 | 1.27 | miR-92a |
| 1.29e-05 | 2354.9 | 1919.4 | 1.23 | miR-103 |
| 0.0003869 | 750.4 | 615.7 | 1.22 | miR-30c |
| < 1e-07 | 289.4 | 237.8 | 1.22 | miR-130b |
| < 1e-07 | 452.7 | 372.0 | 1.22 | miR-93 |
| 2.74e-05 | 2116.4 | 1743.6 | 1.21 | miR-107 |
| < 1e-07 | 297.7 | 248.0 | 1.20 | miR-30e |
| < 1e-07 | 227.0 | 191.9 | 1.18 | miR-15a |
| 0.0020002 | 1209.2 | 1031.6 | 1.17 | miR-181b |
| 0.002002 | 333.8 | 287.0 | 1.16 | miR-32 |
| < 1e-07 | 191.3 | 164.8 | 1.16 | miR-15b |
| 0.0026477 | 508.2 | 438.5 | 1.16 | miR-181a |
| 3.4e-05 | 160.0 | 147.3 | 1.09 | miR-345 |
| 0.0022001 | 208.1 | 193.5 | 1.08 | miR-34a |
| 0.0037199 | 154.1 | 146.2 | 1.05 | miR-374a |
| 5.46e-05 | 153.4 | 165.0 | 0.93 | miR-339-5p |
| < 1e-07 | 129.8 | 141.1 | 0.92 | miR-139-5p |
| 1.8e-06 | 139.6 | 152.7 | 0.91 | miR-133a |
| < 1e-07 | 140.7 | 154.8 | 0.91 | miR-129-3p |
| < 1e-07 | 139.0 | 153.9 | 0.90 | miR-132 |
| < 1e-07 | 140.7 | 156.7 | 0.90 | miR-299-5p |
| 0.003905 | 617.0 | 694.2 | 0.89 | miR-214 |
| 1.49e-05 | 188.0 | 212.4 | 0.89 | miR-338-3p |
| 0.0019696 | 162.9 | 184.9 | 0.88 | miR-193a-3p |
| 0.0004584 | 529.1 | 603.1 | 0.88 | miR-128a |
| 0.0020002 | 160.4 | 183.8 | 0.87 | miR-183 |
| 4.82e-05 | 225.6 | 259.4 | 0.87 | miR-138 |
| < 1e-07 | 150.4 | 172.9 | 0.87 | miR-202 |
| 0.0017889 | 190.6 | 220.4 | 0.86 | miR-95 |
| < 1e-07 | 173.9 | 204.2 | 0.85 | miR-143 |
| < 1e-07 | 164.8 | 193.6 | 0.85 | miR-9 |
| < 1e-07 | 150.2 | 176.9 | 0.85 | miR-338-5p |
| 1e-06 | 220.5 | 260.2 | 0.85 | miR-326 |
| 0.0002492 | 175.1 | 208.5 | 0.84 | miR-204 |
| 2.67e-05 | 419.9 | 507.8 | 0.83 | miR-193a-5p |
| < 1e-07 | 204.0 | 247.9 | 0.82 | miR-33b |
| 1.8e-06 | 382.5 | 471.2 | 0.81 | miR-206 |
| 2.23e-05 | 273.7 | 349.6 | 0.78 | miR-205 |
| < 1e-07 | 714.9 | 1015.1 | 0.70 | miR-145 |
| < 1e-07 | 275.7 | 403.0 | 0.68 | miR-203 |

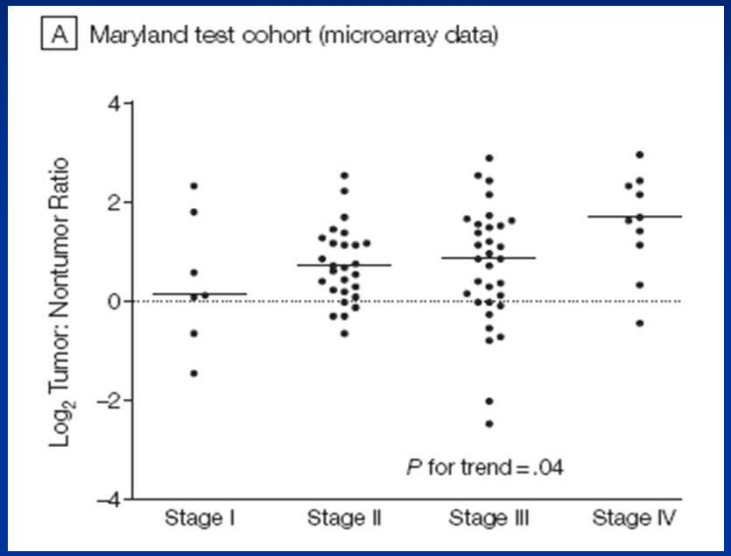
# miRNA specificity through ES differentiation

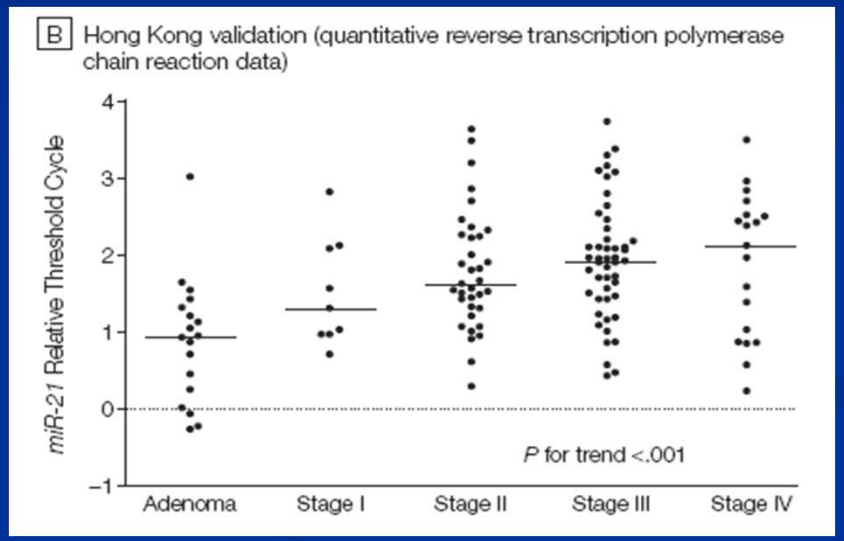


# Question

- Can we replicate and confirm microarray results?
- On different patient cohorts?
- With different techniques?

# *miR-21* Expressed at Higher Levels in Colon Adenocarcinomas With Increasing Expression in Advanced Tumors

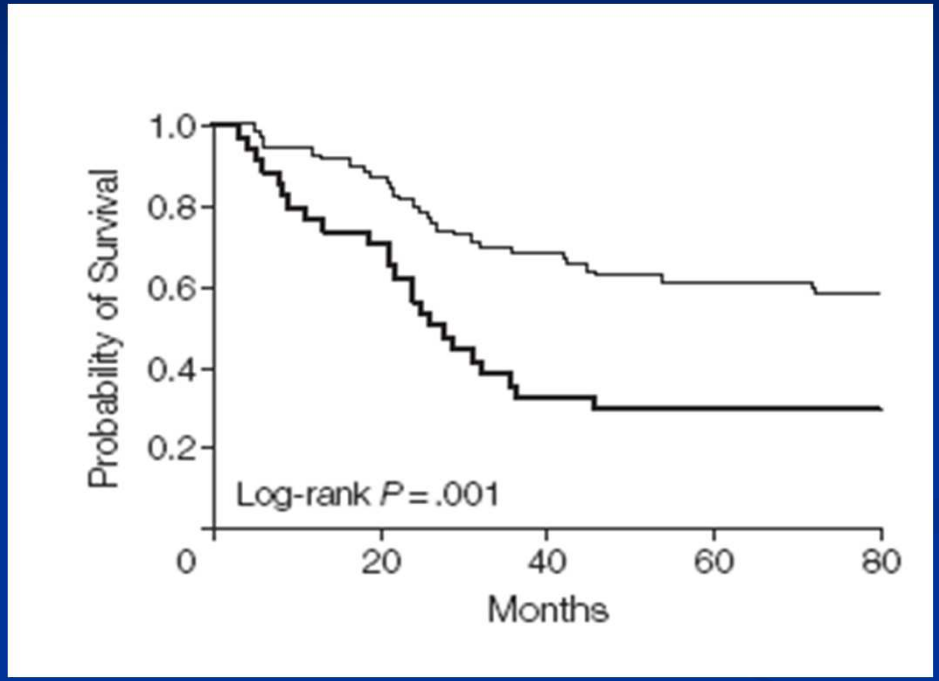



MicroRNA microarrays were used to measure *miR-21* expression levels in the Maryland test cohort. Dot plots represent *miR-21* log2 (tumor:nontumor ratios) for paired tissues as calculated from microRNA microarrays from the original cohort. Values greater than 0 indicate tumors with expression values higher than nontumorous tissue.

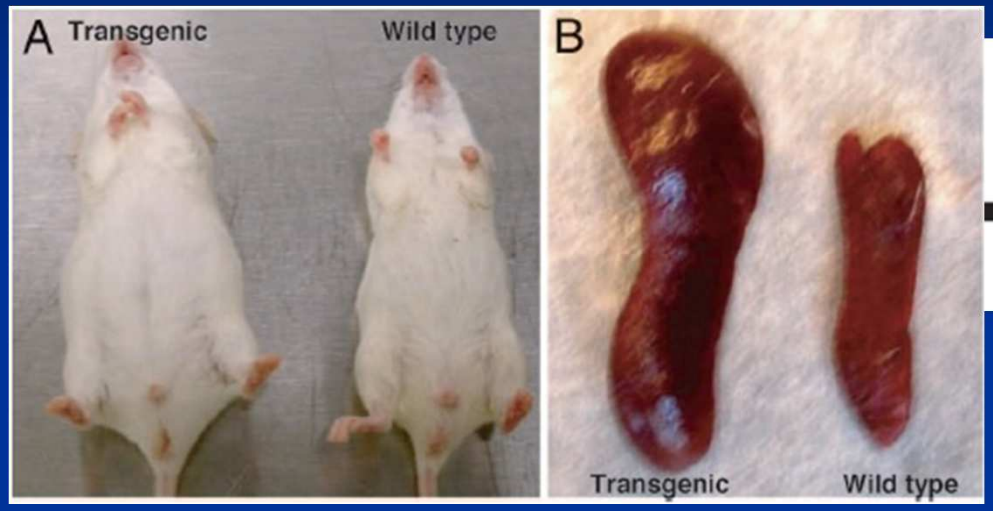
Tissue types have been ordered from TNM stage I to stage IV tumors. Bars indicate median value.

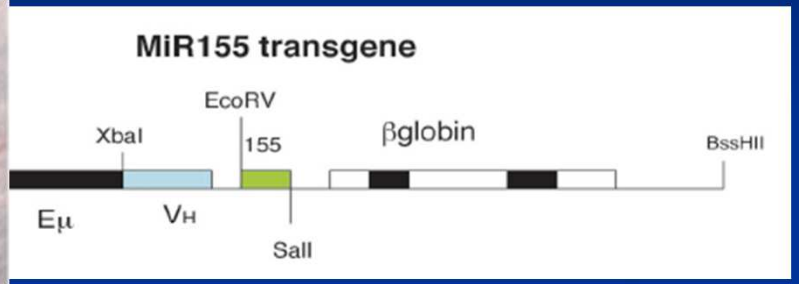
# High *miR-21* and Poor Survival in Patients With Typical Adenocarcinoma Histology




The association of high *miR-21* expression in tumors with poor prognosis is validated by qRT-PCR in an independent cohort (Honk Kong).
High expression is based on the highest tertile (3.3-fold to 8.7-fold higher than nontumor). Log-rank *P* values are from Kaplan-Meier analysis.

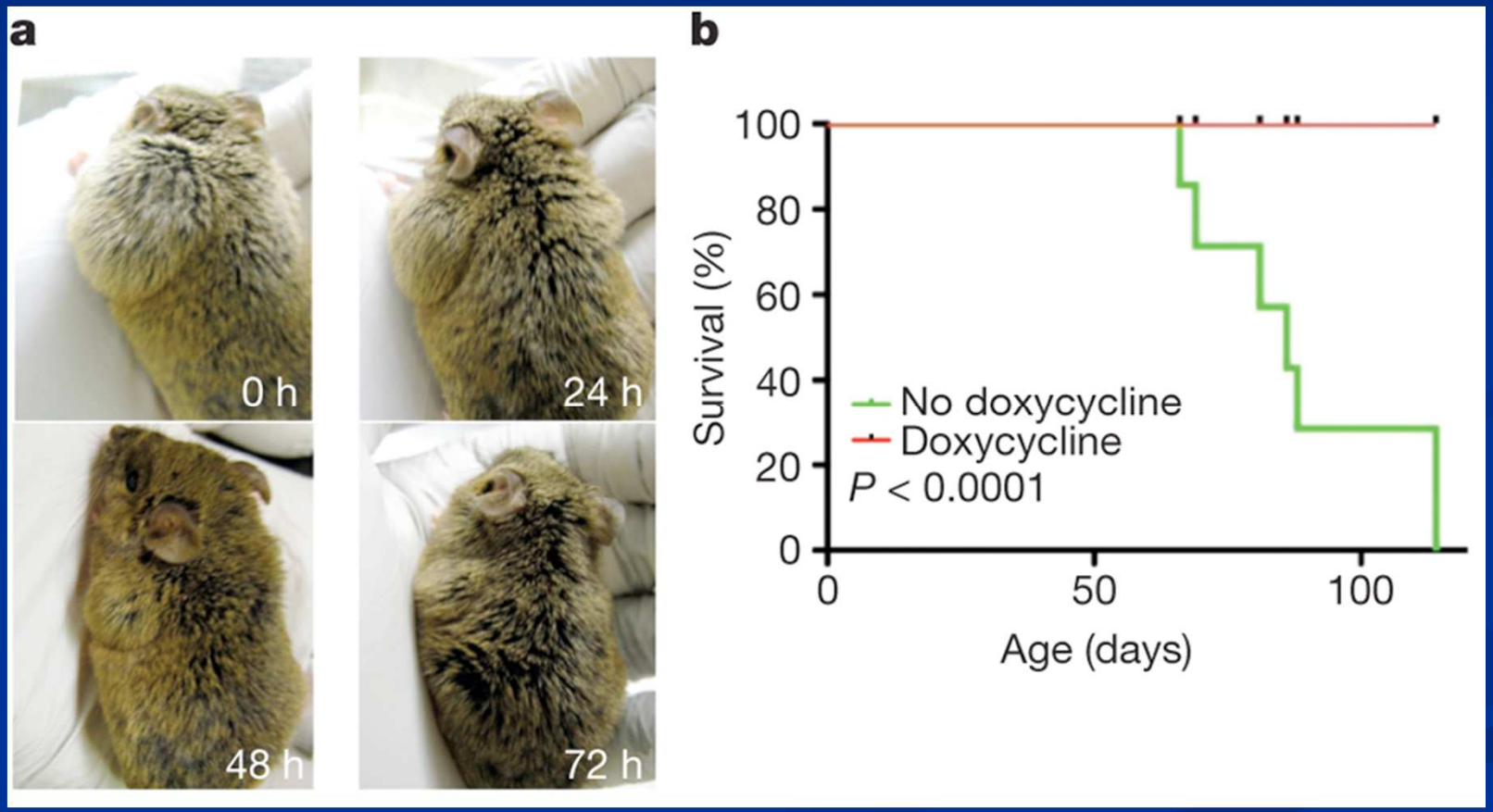
# miR-155 transgenic mice develop acute lymphoblastic leukemia/high grade lymphoma (Costinean et al, 2006, 2009)




The leukemias start at approximately 9 months and are preceded by a polyclonal pre-B cell proliferation, with variable clinical presentation, are transplantable and develop oligo/monoclonal expansion. The B cell precursors have the highest miR-155 expression and are at the origin of the leukemias. SHIP and C/EBPbeta, two regulators of the IL-6 signaling pathway, are direct targets of miR-155.

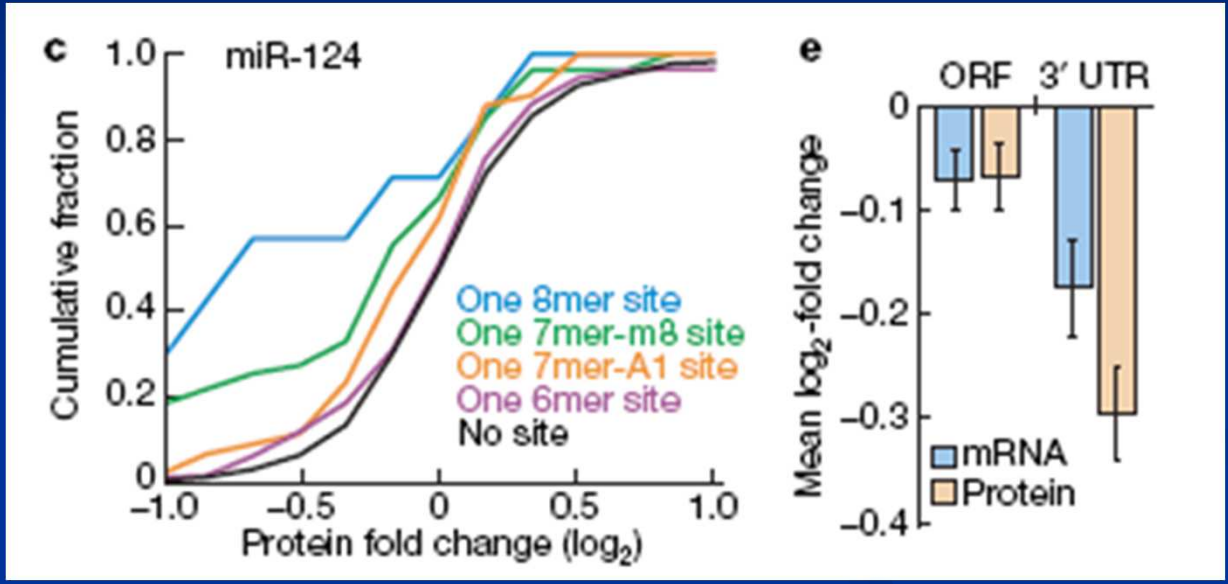
# Mir21 onco-mouse



# Question:

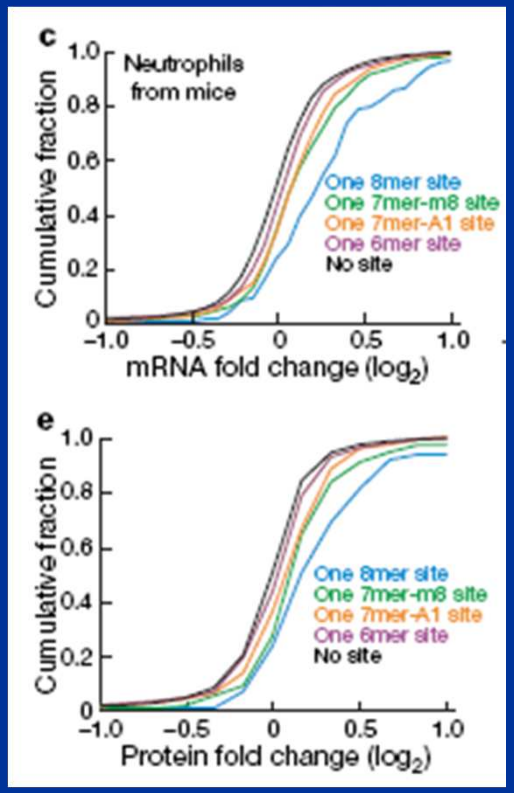
- What is the genomic impact of miRNA on proteins and mRNAs?

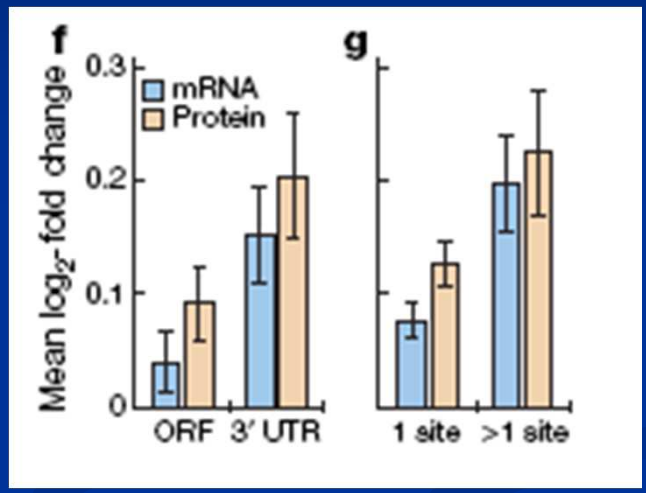
# The case of miR-124




Response of proteins from messages with single miR-124 3'-UTR sites.
Plotted is the fraction of proteins that change at least to the degree indicated on the x axis. Proteins from messages with multiple 3'-UTR sites were not considered.
6mer sites that were part of larger sites were not included in the 6mer distribution, and 7mers that were part of 8mers were not included in the 7mer distributions.

# The impact of deleting mir-223 in mouse neutrophils

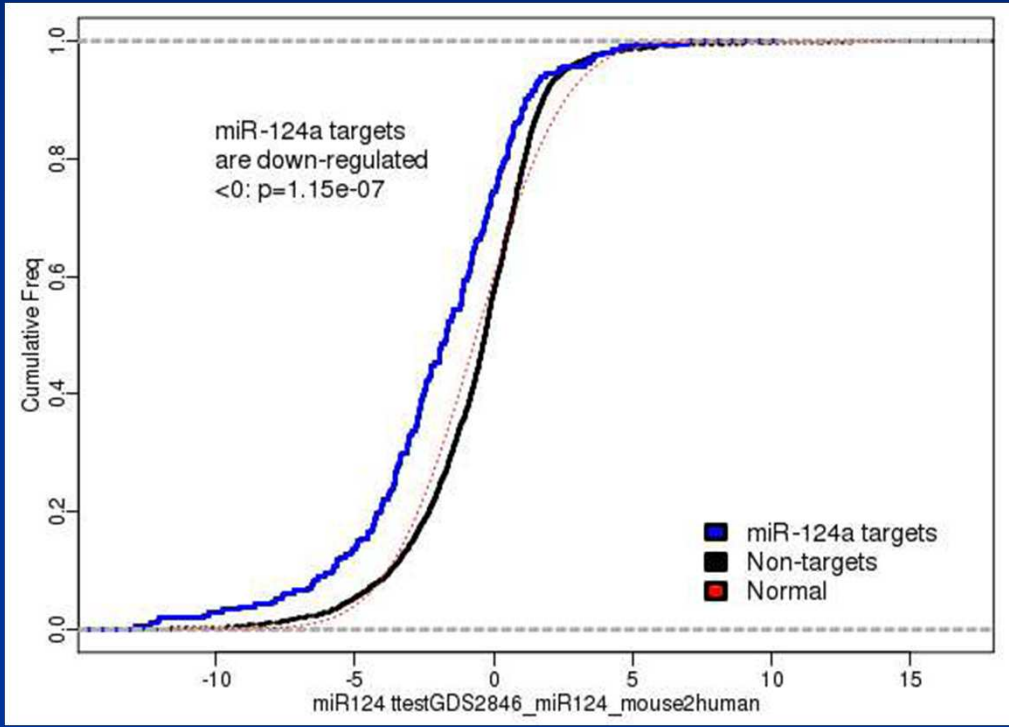

Analysis of neutrophils isolated from mice monitoring the effects of miR-223 loss on messages with single miR-223 sites in their 39 UTRs.
Plotted is the fraction of messages that changed at least to the degree indicated on the x axis

**Reverse assay:**
**from mRNAs to miRNAs**

Messenger RNA profiles
(Affymetrix)

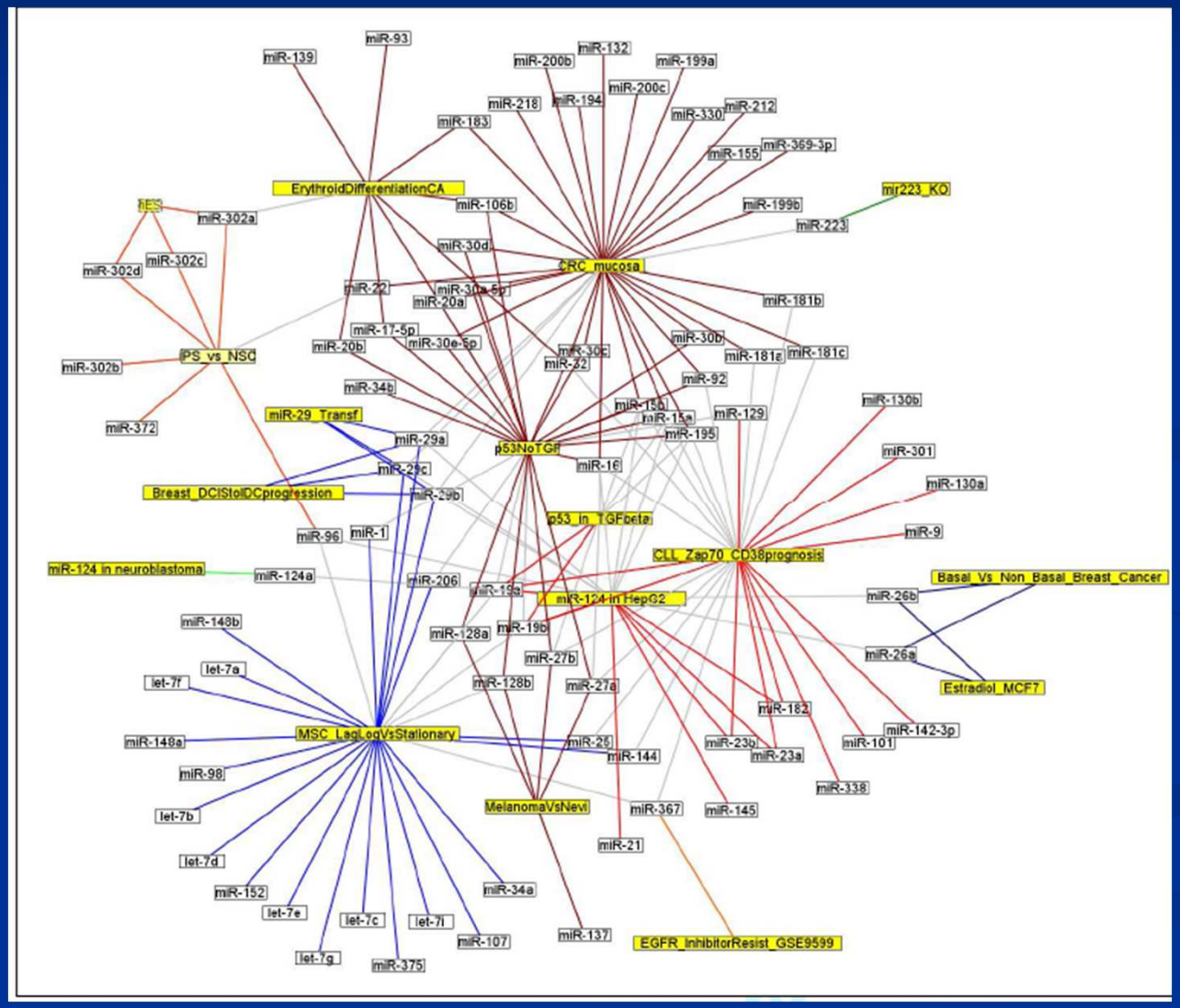
miRNA agents

# The cumulative distribution function (ECDF) plot of the Kolmogorov-Smirnov test correctly identify target coding genes specifically controlled by microRNAs.




T-test scores from the miRNA assay Vs. the control cells were analyzed. A) miR-124 over expression in mouse neuroblastoma (GDS2846). miR-124 correctly detected as the miRNA with the most down-regulated target genes. The blue curve (miR-124 target genes) on the left and above the black non–targets curve indicates down-regulation of messenger RNAs. The red dotted curve is that expected by random association.

# Extracting miRNA info from 66 Affy experiments

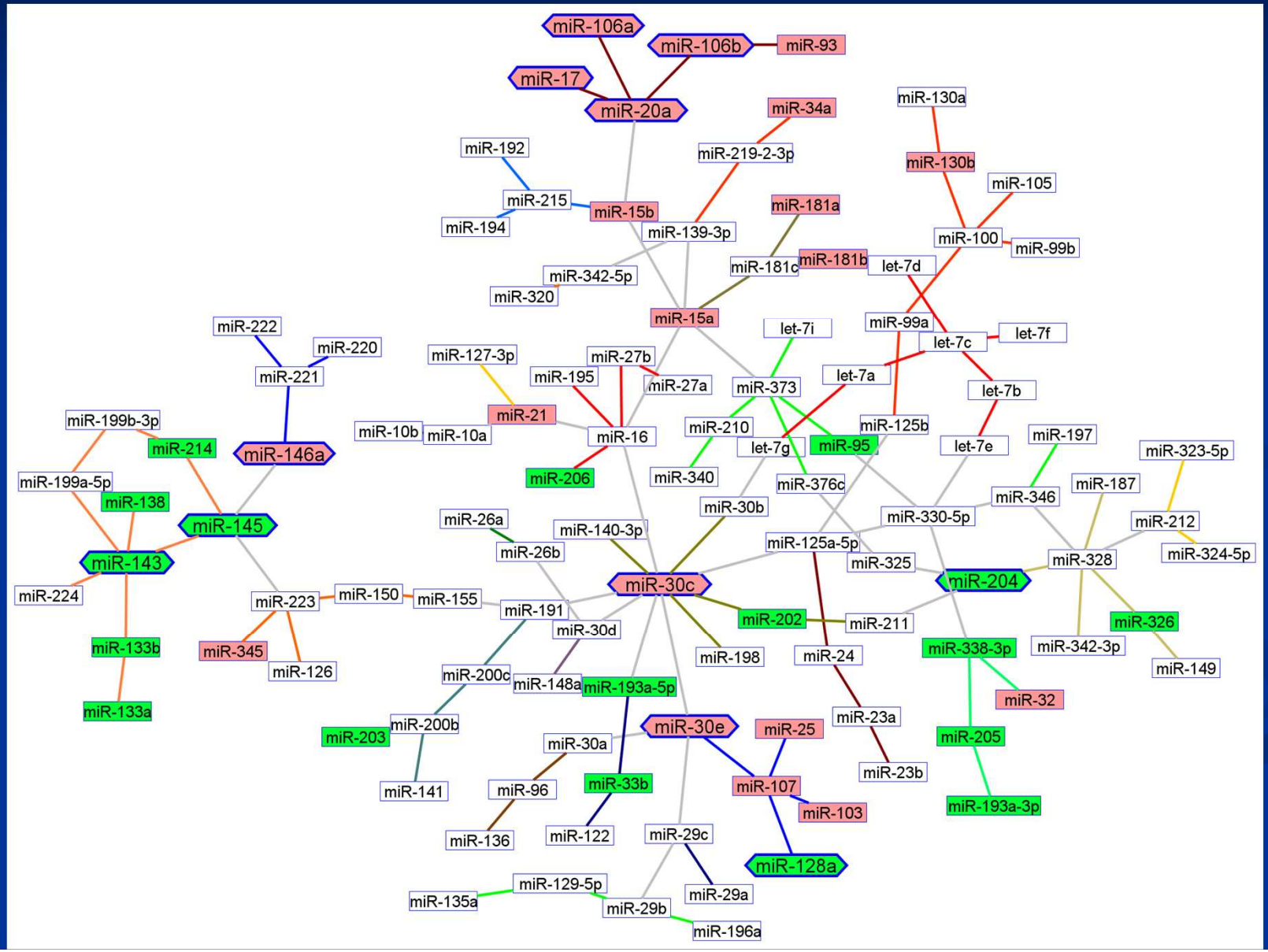



The different conditions are colored in yellow. miRNA grouping was obtained by MCL clustering and is indicated by the colored edges. Seventeen out of 66 tested cellular conditions display highly significant miRNAs.

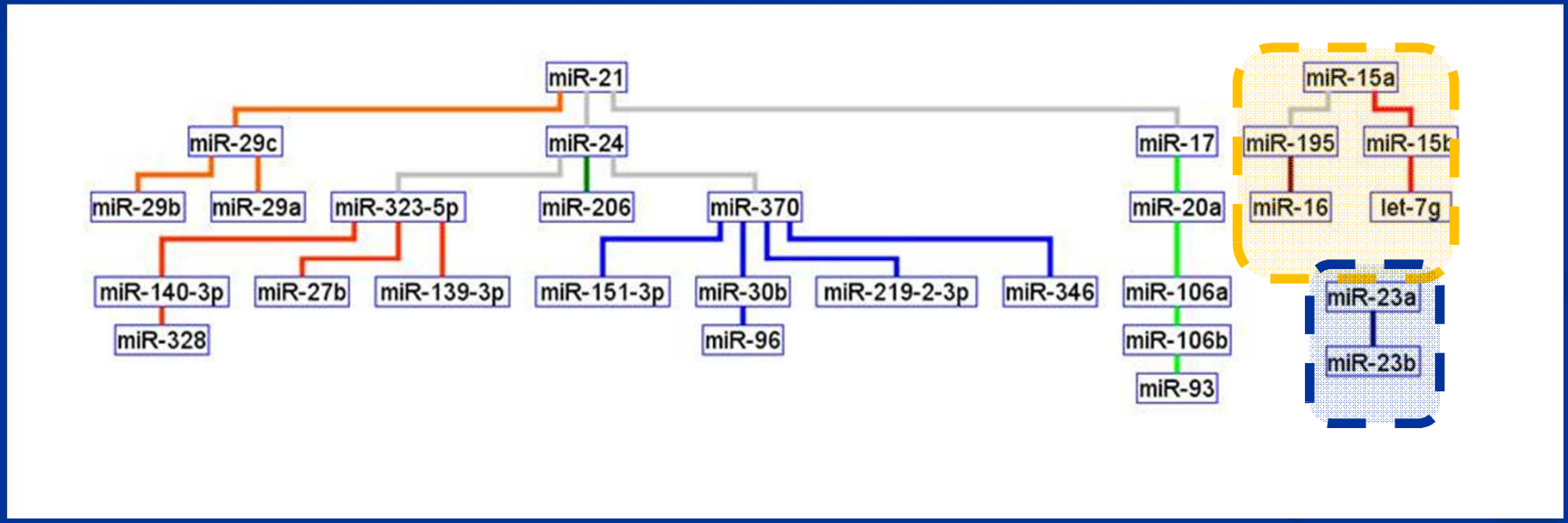
# miRNA gene networks

- Bayesian networks were built for different tissues and diseases.
- For each tissue or disease all the mature miRNAs were considered.
- Then the expression values were preprocessed to filter out non varying miRNAs.
- The MCL graph-based clustering algorithm to extract co-regulated genes from miRNA networks.

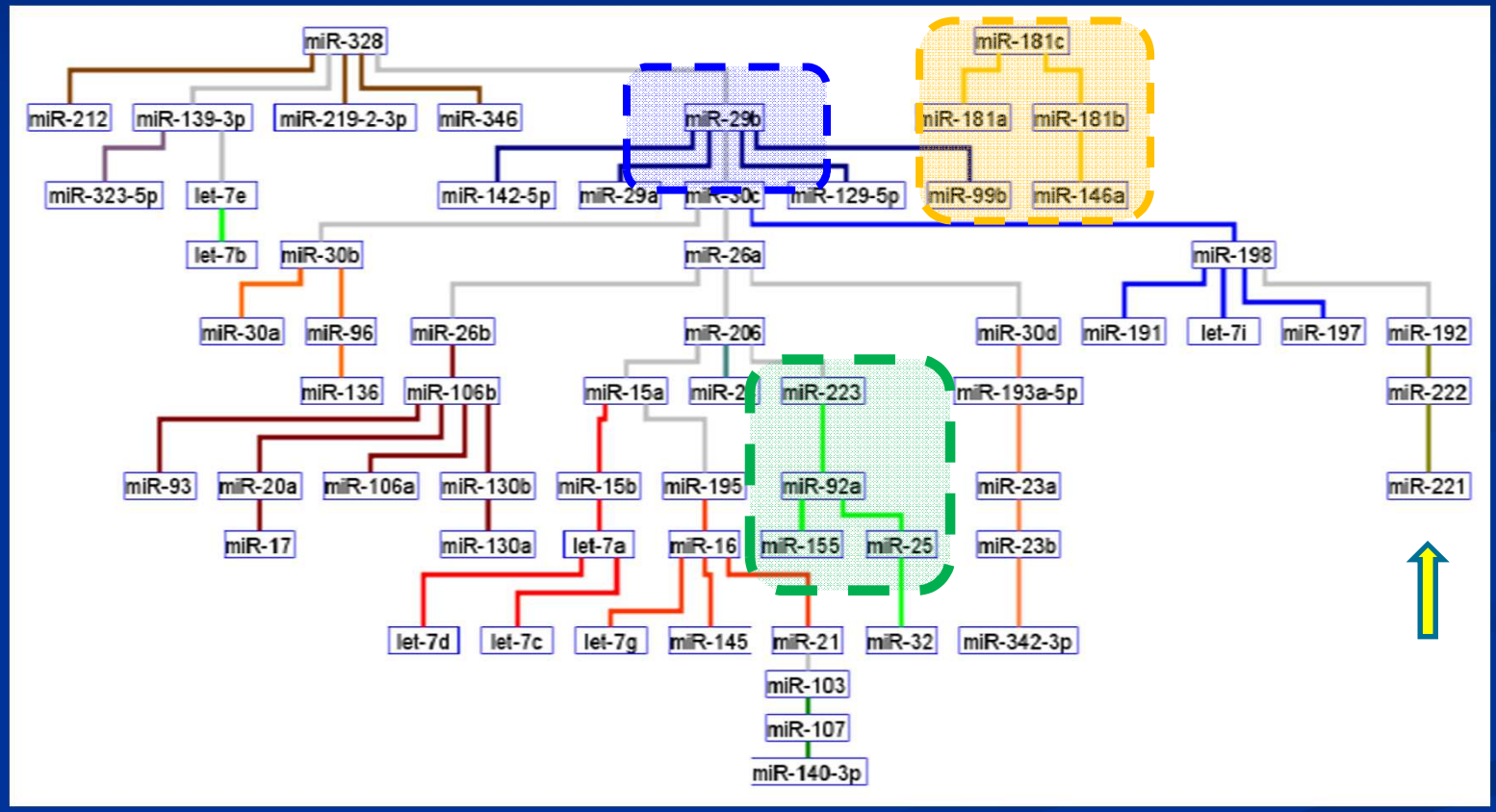
# The miRNA network in solid cancers (2532 samples, 31 cancer types, 120 miRNAs)



# miRNA network of CLL (254 samples, 3 MCL clusters)

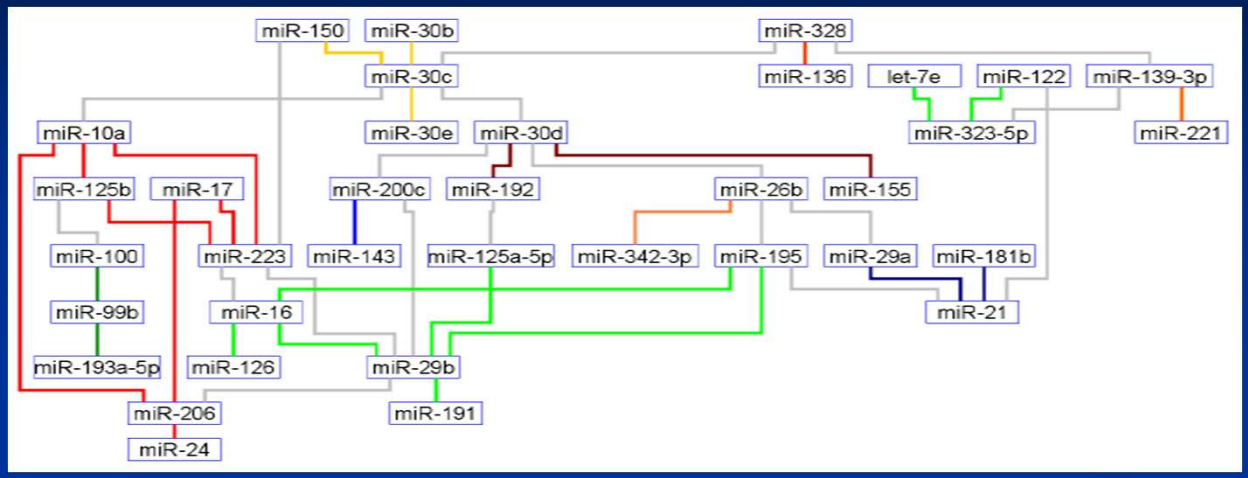


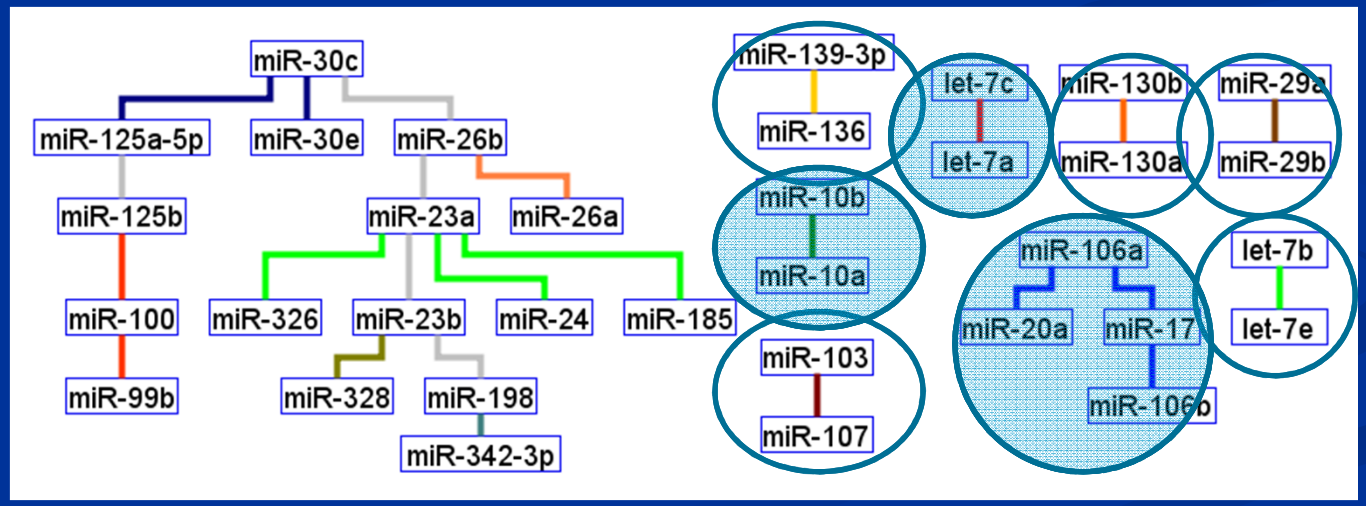
# miRNA network of AML (589 samples, 2 MCL clusters)



# Comparison of miRNA networks in normal lung and adenocarcinoma

A) Normal lung (71 samples,1 MCL cluster).




B) Lung adenocarcinoma (125 samples, 9 MCL clusters)

# Conclusions from chips

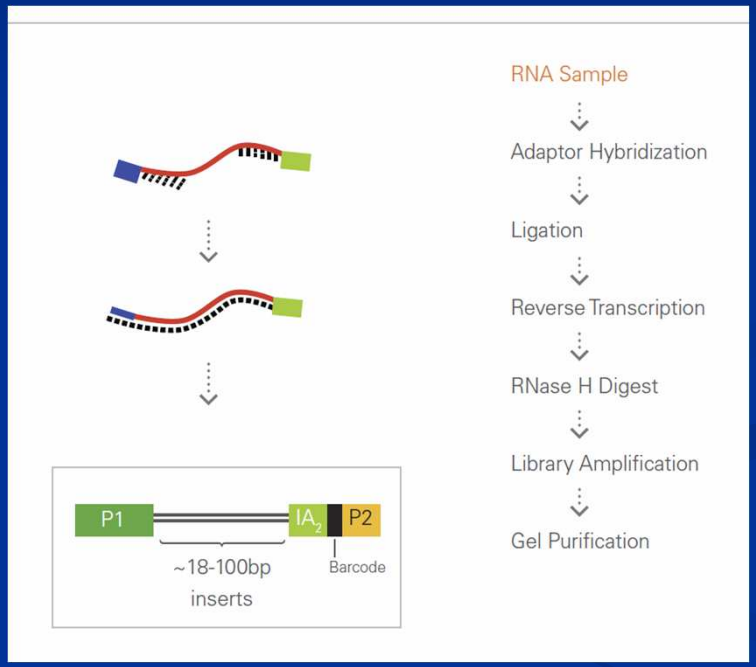
- miRNA signatures are excellent classifiers and related to fundamental pathways

- miRNA profiles can be inferred by mRNA profiles.

- miRNA networks are reprogrammed in cancer and leukemias and reveal “rebel” miRNAs.

# miRNA profiling and RNA discovery by Next Generation Sequencing

- Problems:
  - 1 Data Analysis
  - 2 Data Management



# Data Management

- miRNA sequencing (Profiling)
  - Short RNA sequencing
    - Multiplexing allows 48 samples to be run at the same time
    - Activation of pipelines
      - Small RNA pipeline (ABI)
      - In house novel pipeline
- Database
  - Annotation
  - Normalization

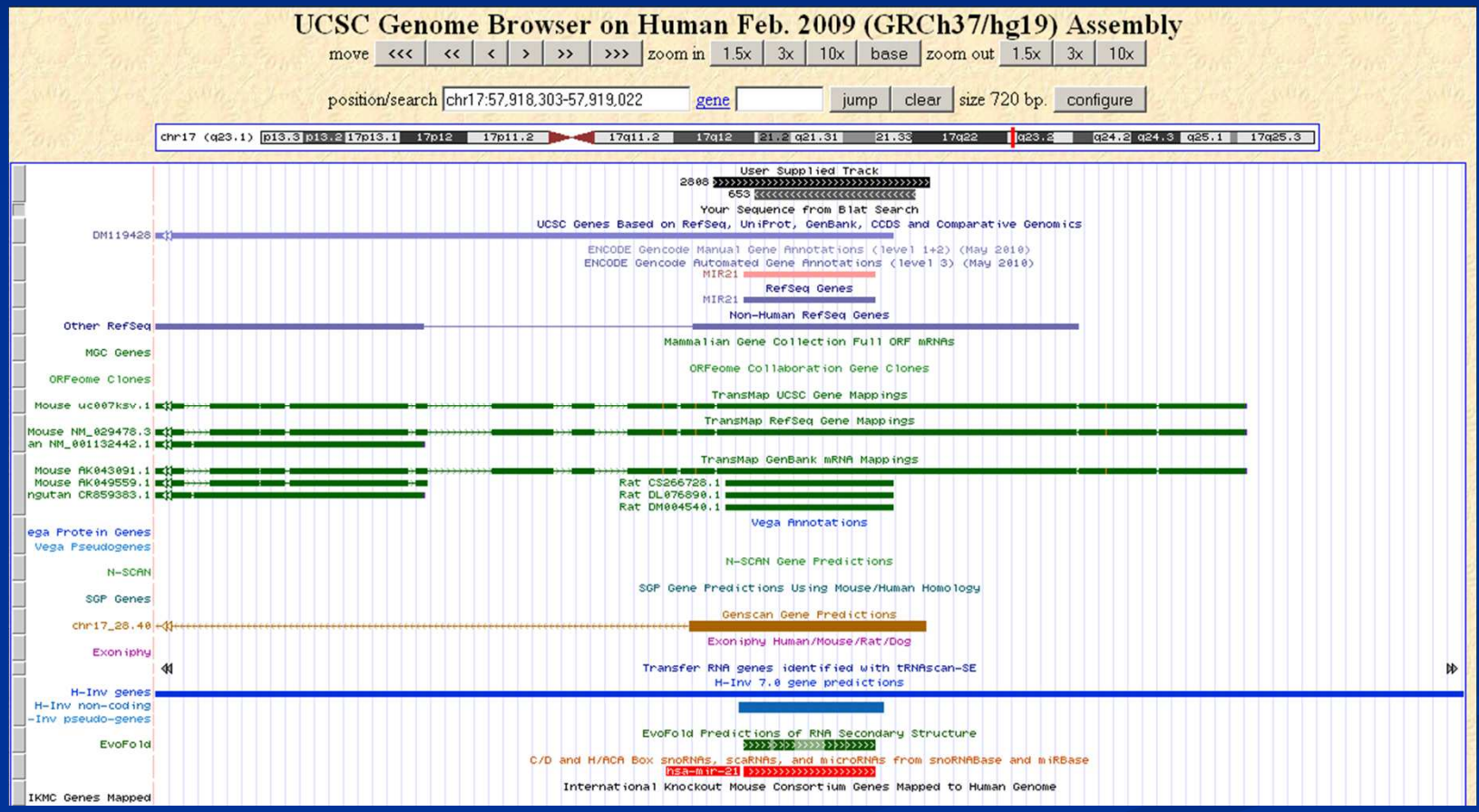
**Indexes:**

| Keyname | Type | Cardinality | Action | Field |
|---|---|---|---|---|
| **PRIMARY** | PRIMARY | 397 | | miRBase_Counts_FileID |
| **dir_and_csfasta_filename** | UNIQUE | 397 | | dir_and_csfasta_filename |
| **solid_machine** | INDEX | 3 | | solid_machine |
| **date_of_run** | INDEX | 39 | | date_of_run |
| **sample_name** | INDEX | 397 | | sample_name |
| **project_name** | INDEX | 49 | | project_name |
| **sample_ID** | INDEX | 397 | | sample_ID |

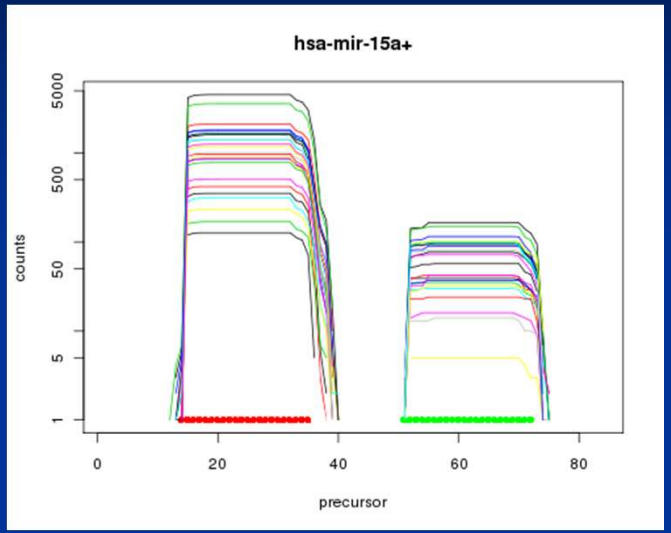
# hg19 contigs generated from short RNA reads

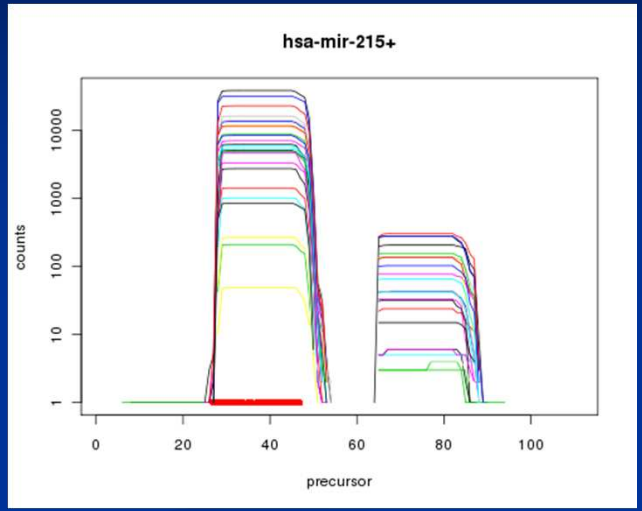
- Merging all short reads detected in at least 5 samples with at least 3 reads:
  - > 100K different short non repetitive sequences

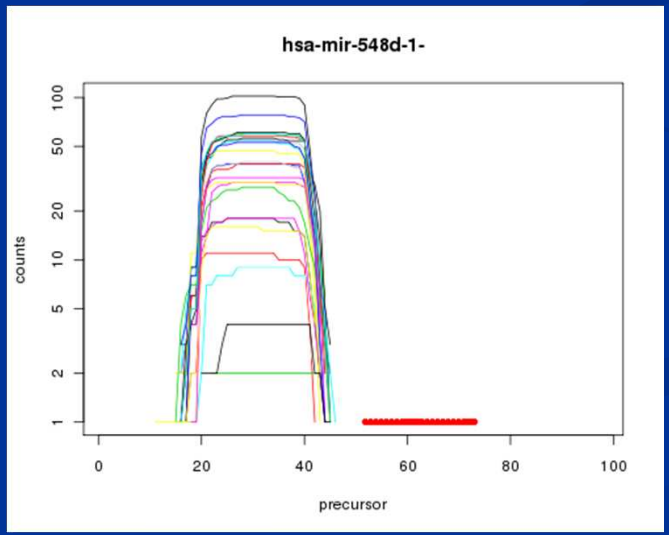
# miR-21 contigs



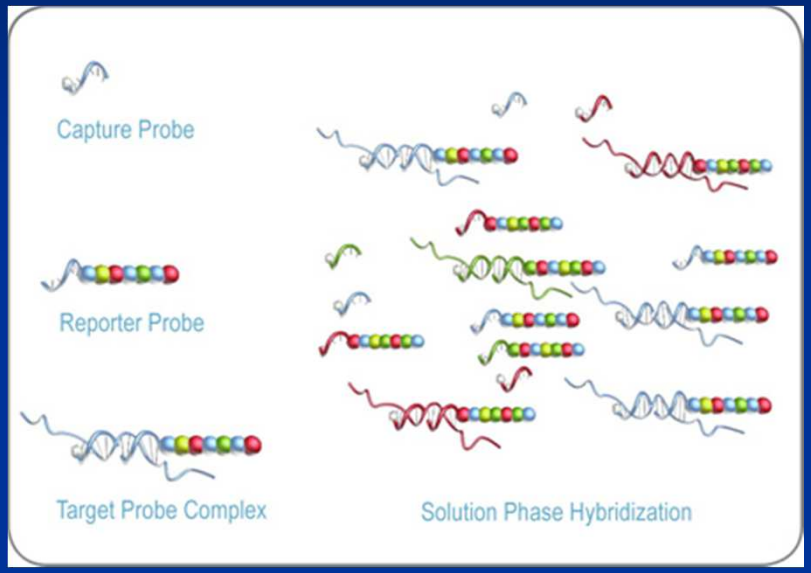
# Visualization of miRNA processing

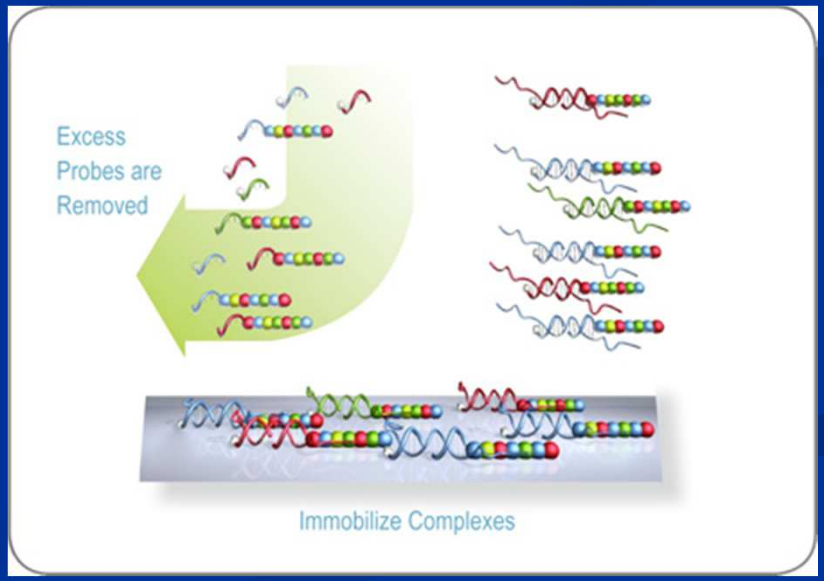


# NanoString


Capture Probe
Reporter Probe
Target Probe Complex
Solution Phase Hybridization


Excess
Probes are
Removed
Immobilize Complexes

# CRC and Normal Colon profiles using different detection techniques (average values)

| miR29a | Solid | RT-PCR (ABI) | Chips (CRC vs adenomas) | RT-PCR (Exiqon) | Northern | Nanostring |
|---|---|---|---|---|---|---|
| Control | 328.17 | 2.46 | 147.04 | 6.16 | 4866397.79 | 7491.28 |
| Cancer | 2574.11 | 2.01 | 526.25 | 13.45 | 9790703.92 | 13026.55 |
| miR31 | Solid | RT-PCR | Chips | | Northern | Nanostring |
| Control | 12.85 | 0.01 | Ns | | | 34.49 |
| Cancer | 97.62 | 0.12 | Ns | | | 140.73 |
| miR135b | Solid | RT-PCR | Chips | | Northern | Nanostring |
| Control | 10.00 | 0.02 | 224.61 | | | 42.74 |
| Cancer | 25.4 | 0.22 | 392.36 | | | 798.77 |
| miR223 | Solid | RT-PCR | Chips | | Northern | Nanostring |
| Control | 572.94 | 0.86 | 250.21 | | | 1200.20 |
| Cancer | 1432.18 | 1.13 | 2137.13 | | | 2724.59 |
| miR224 | Solid | RT-PCR | Chips | | Northern | Nanostring |
| Control | 12.06 | 0.04 | 65.26 | | | 95.77 |
| Cancer | 42.62 | 0.14 | 142.17 | | | 117.97 |

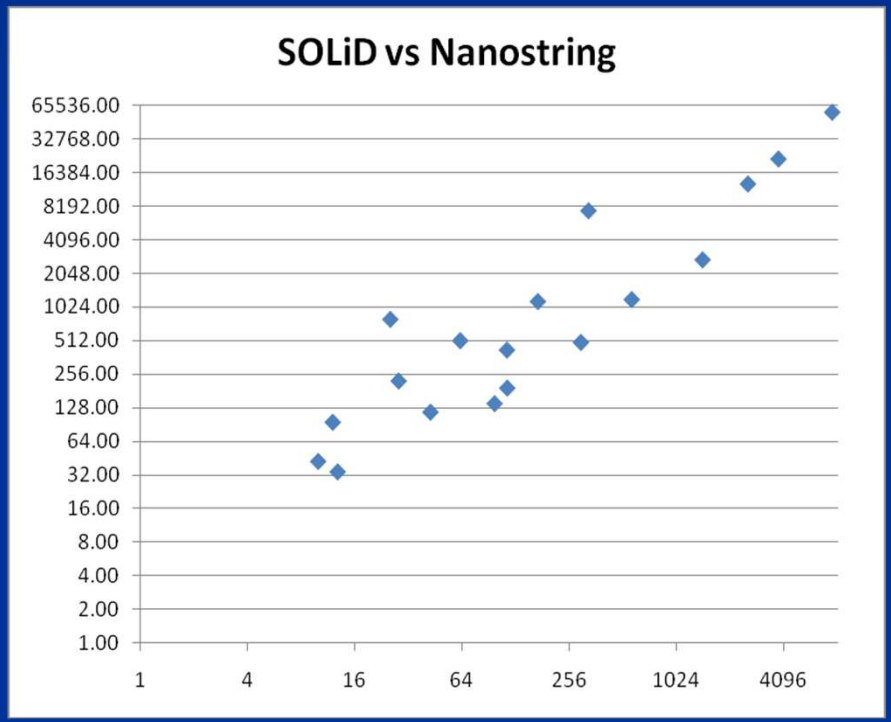
Up regulated miRNAs

# CRC and Normal Colon profiles using different detection techniques (average values)

| miR-497 | Solid | RT-PCR | chips | RT-PCR (Exiqon) | Northern | Nanostring |
|---|---|---|---|---|---|---|
| Control | 170.68 | 0.37 | | | | 1154.83 |
| Cancer | 62.61 | 0.14 | | | | 514.62 |
| miR-148a | Solid | RT-PCR | | | Northern | Nanostring |
| Control | 102.88 | 0.29 | Ns | | | 1368.73 |
| Cancer | 38.67 | 0.27 | Ns | | | 2843.71 |
| miR-215 | Solid | RT-PCR | | | Northern | Nanostring |
| Control | 114.44 | 0.52 | | | 851396.62 | 424.10 |
| Cancer | 28.28 | 0.10 | | | 651444.00 | 224.35 |
| miR-378 | Solid | RT-PCR | | | Northern | Nanostring |
| Control | 297.47 | 2.27 | | | | 497.43 |
| Cancer | 114.89 | 0.80 | | | | 193.85 |
| miR-145 | Solid | RT-PCR | Chips | | Northern | Nanostring |
| Control | 7640.96 | | 830.76 | | 10165254.98 | 57050.29 |
| Cancer | 3815.46 | | 1765.8 | | 3911435.14 | 21756.85 |

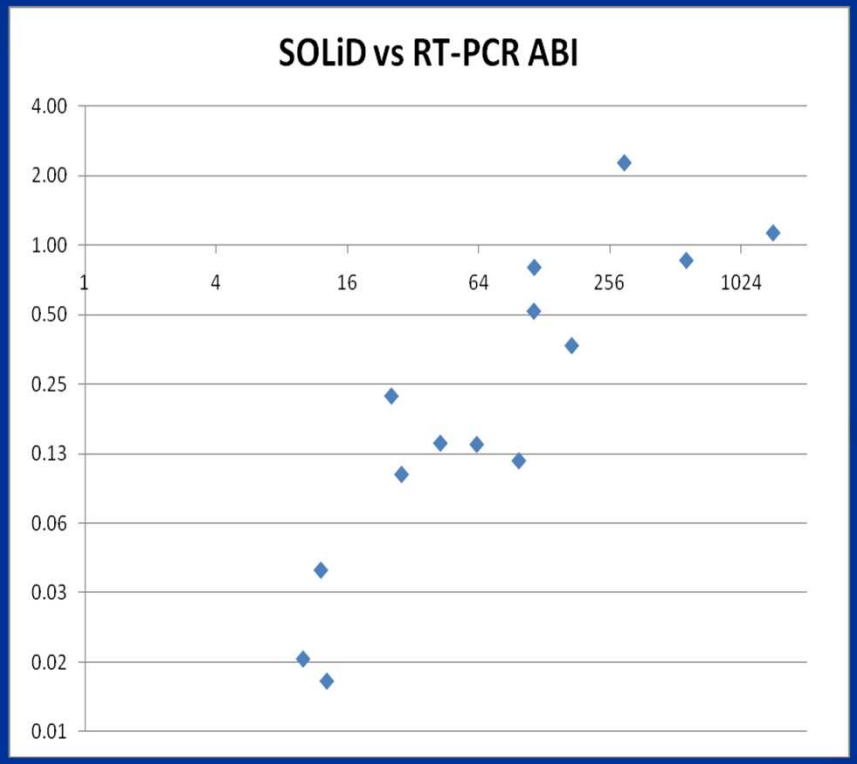
Down regulated miRNAs

# CRC: Validation across platforms



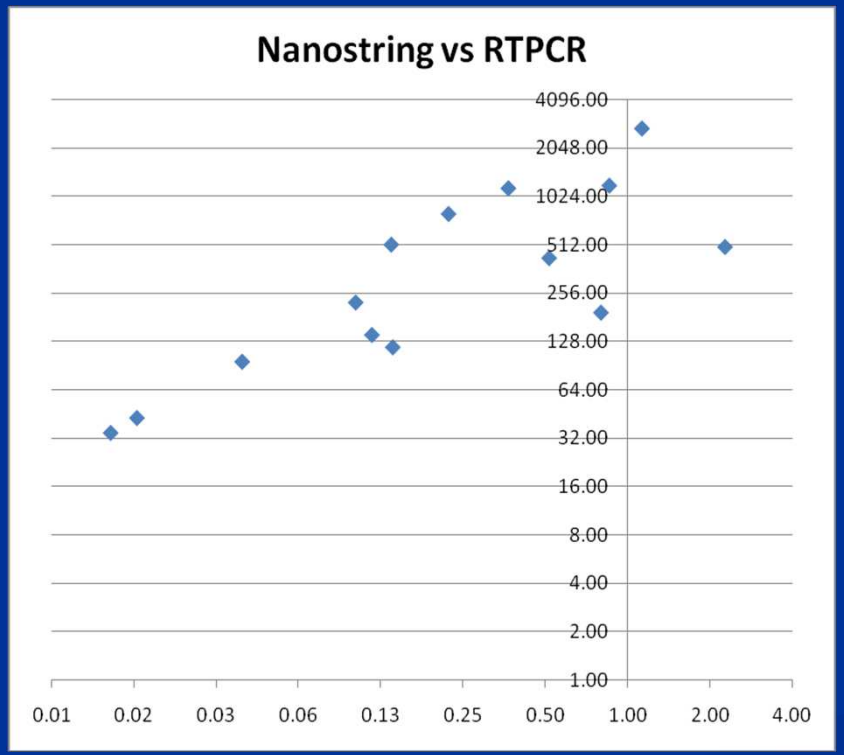

- SOLiD vs Nanostring (9/10 concordant trends within CRC/Normal – 1 yellow discordant not included)
- R = 0.98, pvalue =2.73E-12

# CRC: Validation across platforms




- SOLiD vs RTPCR (7/9 concordant trends within CRC/Normal – 2 yellow discordant not included)
- R = 0.51, pvalue =0.06

# CRC: Validation across platforms



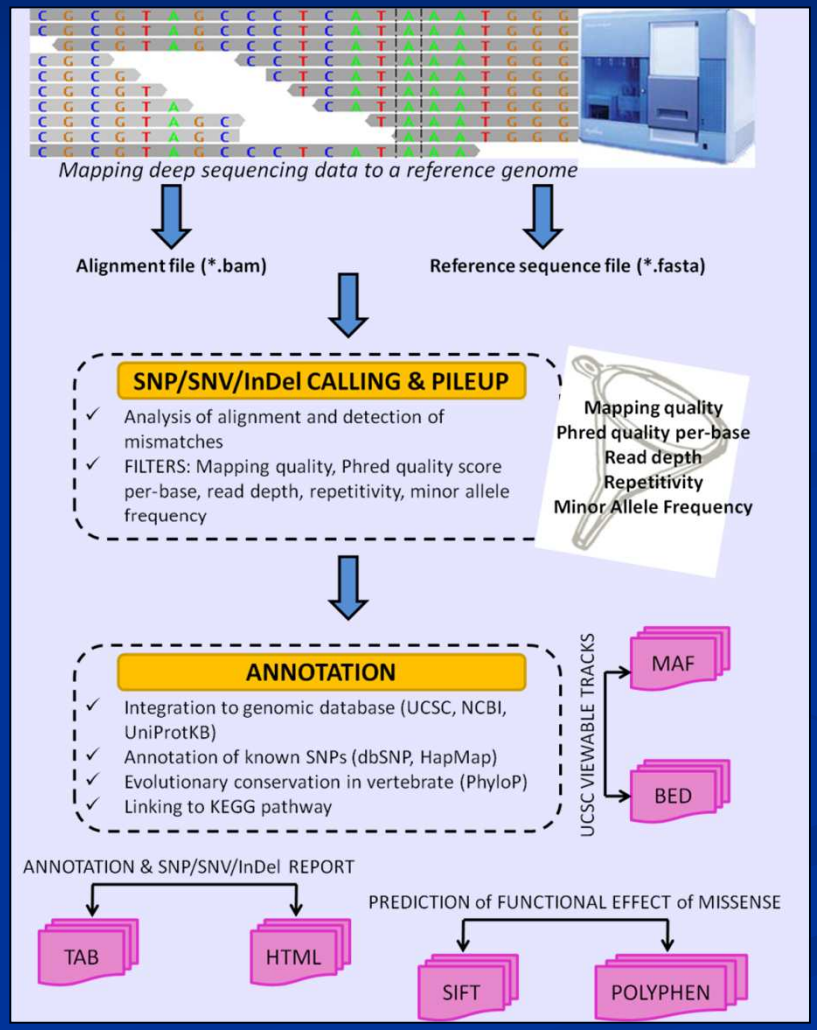

- RTPCR vs Nanostring (7/9 concordant trends within CRC/Normal – 2 yellow discordant not included)
- R = 0.41, pvalue =0.15

# Some more reasons to use Next Generation Sequencing for miRNAs?

- Detection of isoforms (isomiRNAs)
- Extreme specificity
- Detection of Viral miRNAs
- Detection of novel miRNAs
- Detection of other short RNAs
- Detection of mutations (reliable in RNA?)

# Genomic Analysis of Mutations Extracted by Sequencing

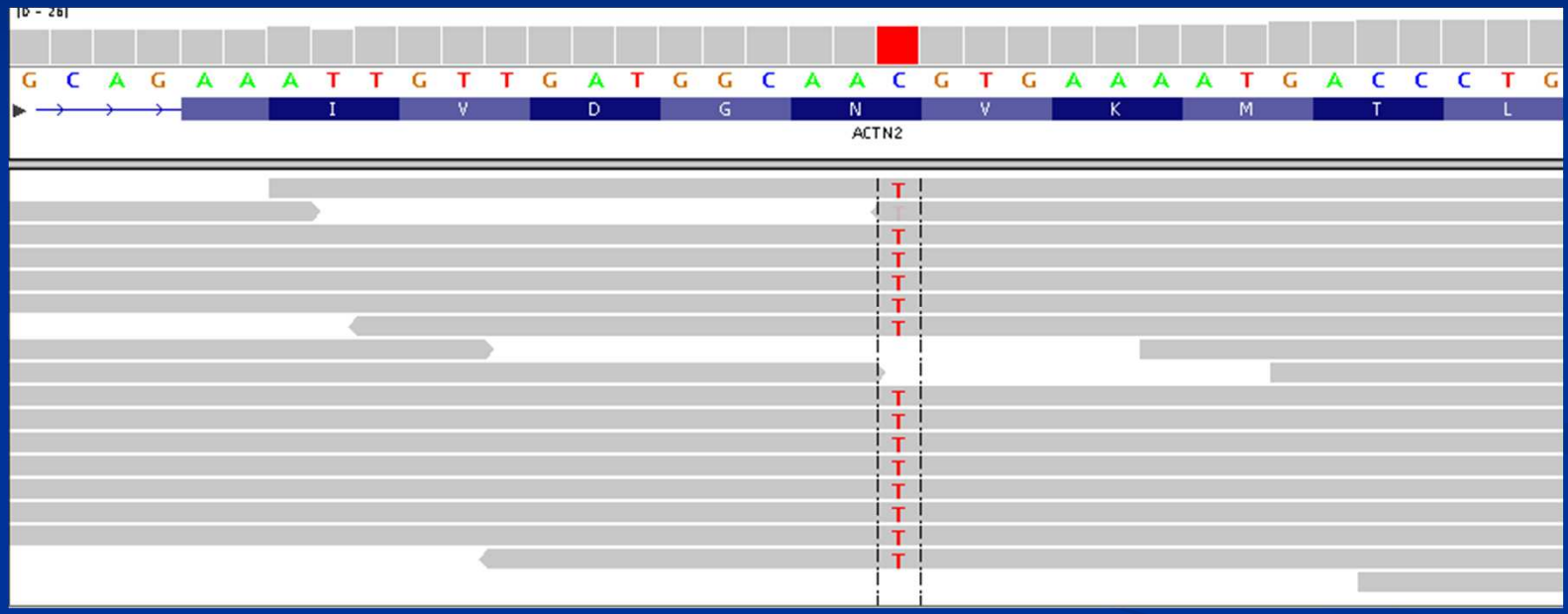
Sana et al,
*Bioinformatics*
2011



# INPUT

- **SAM (Sequence Alignment/Map)** format
  - ✓ Supports:
    - short/long reads
    - compatible with different sequencing platforms
    - single-read/paired-end
  - ✓ Generable by several aligning packages, such as BWA, MAQ, SHRIMP, BFAST, PerM, Mosaik, ELAND, PASS, SOLiD LifeScope™

# SNP/SNV/InDel CALLING




For each mismatch, extracts the genomic coordinates (chromosome and position), reference base, the first and second calls, with respective quality and counts, and repetitivity.

On the agenda:

•miRNA and other short RNA profiling in leukemia and cancer (about 500 sequenced samples)

•miRNA mutation detection from RNA (editing?)

•miRNA mutations detection from genomic sequencing (SNPs?)

Many Thanks for Your Attention!