

# MikroRNA-Bioinformatik

Peter N. Robinson

Institut für medizinische Genetik  
Charité Universitätsmedizin Berlin

4. Januar 2016

# MiRNA-Wirkmechanismen und offene Fragen

## MikroRNA

Peter N. Robinson

Nomenklatur

miR29

Mikroarray

GSEA

MREs

- Die jeweilige MikroRNA (miRNA) wird in einen aus mehreren Proteinen bestehenden Komplex, den "RNA-induced silencing complex" (RISC) integriert
- Der RISC-Komplex bindet an die Ziel-mRNA an eine zur miRNA komplementäre Basensequenz in deren 3'-UTR-Region
- Diese Bindung führt zur Hemmung der Proteinsynthese *und/oder* Degradation der Ziel-mRNA



Grafik: van Rooij (2014) Development of microRNA therapeutics is coming of age. *EMBO Mol Med* 6:851-64

s. auch Guo H et al (2010) Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature*

466:835-40.

# MiRNAs: Bindung an Ziel-mRNA

## MikroRNA

Peter N. Robinson

Nomenklatur

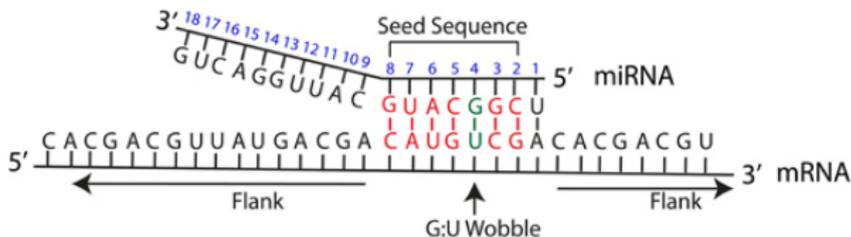
miR29

Mikroarray

GSEA

MREs

- Die miRNA:mRNA-Bindung wird insbesondere durch Komplementarität im Bereich der Nukleotide 2-8 (gezählt vom 5' Ende der miRNA) zur 3'-UTR-Sequenz der mRNA begünstigt



Grafik: Peterson SM et al. (2014) Common features of microRNA target prediction tools. *Front Genet* 5:23.

# MiRNAs: Bindung an Ziel-mRNA

MikroRNA

Peter N. Robinson

Nomenklatur

miR29

Mikroarray

GSEA

MREs

## Heute

wir werden heute folgende Themen besprechen

- mikroRNA-Nomenklatur
- Beispiel einer Analyse der Wirkung einer miRNA-Familie
- Wie kann man nach Veränderungen im Expressionsprofil der mRNA durch die Wirkung einer miRNA suchen?
- Wie findet man miRNA-Zielsequenzen in mRNAs?

# Rückblick

## MikroRNA

Peter N. Robinson

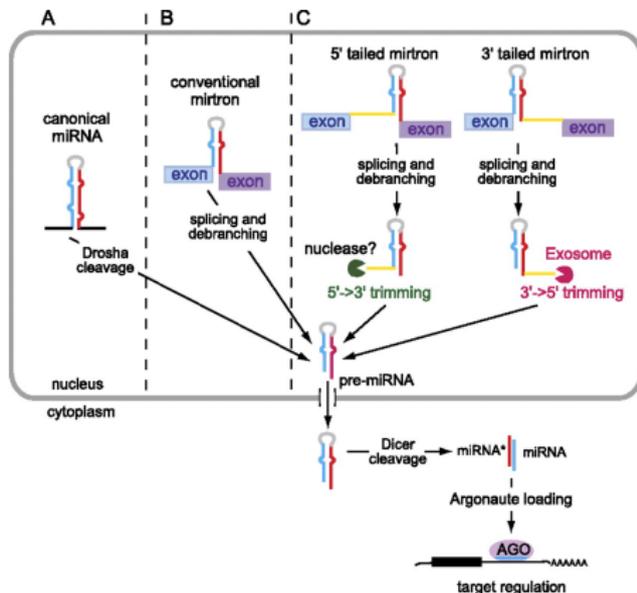
Nomenklatur

miR29

Mikroarray

GSEA

MREs



Ladewig E et al (2012) Discovery of hundreds of mirtrons in mouse and human small RNA data. *Genome Res*

# Outline

## MikroRNA

Peter N. Robinson

## Nomenklatur

miR29

Mikroarray

GSEA

MREs

- 1 mikroRNA-Nomenklatur
- 2 MicroRNAs Differentially Expressed in Postnatal Aortic Development Downregulate Elastin via 3' UTR and Coding-Sequence Binding Sites
- 3 Mikroarrays und globale Transkriptionsprofile
- 4 Gene-set enrichment analysis (GSEA)
- 5 MikroRNA-Ziel-Vorhersage: miRNA-regulatorischen Elementen (MREs)

# mikroRNA-Nomenklatur

MikroRNA

Peter N. Robinson

Nomenklatur

miR29

Mikroarray

GSEA

MREs

## miRNAs

Die MikroRNAs werden fortlaufend nummeriert

- Falls also die letzte publizierte MikroRNA in einer Spezies mir-374 war, würde die nächste die Zahl mir-375 erhalten.
- Falls jedoch eine neu entdeckte miRNA mit einer bereits publizierten miRNA einer anderen Spezies identisch ist, sollte dieselbe Zahl verwendet werden
- Die MikroRNAs einer bestimmten Spezies erhalten einen Namen der Form hsa-mir-121 für Spezies (hsa=Homo sapiens) – miRNA – Nr. 121

# mikroRNA-Nomenklatur

## MikroRNA

Peter N. Robinson

## Nomenklatur

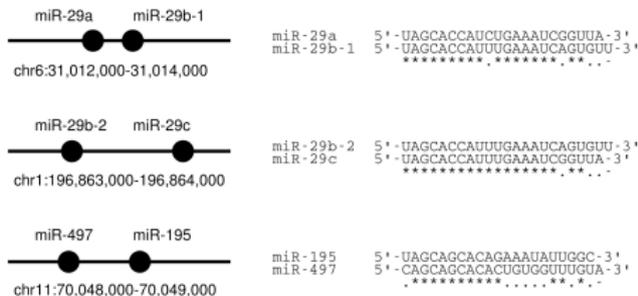
miR29

Mikroarray

GSEA

MREs

- Unterschiedliche Vorläufersequenzen und genomische Loci, welche die identische reife miRNA-Sequenz erzeugen erhalten Namen der Form **hsa-mir-121** und **hsa-mir-121-2**
- Buchstaben-Suffixe (a,b,c,...) bezeichnen eng verwandte reife miRNA-Sequenzen



Ott et al 2011

- Erklären Sie die Benennung der gezeigten miRNAs

# mikroRNA-Nomenklatur

## MikroRNA

Peter N. Robinson

## Nomenklatur

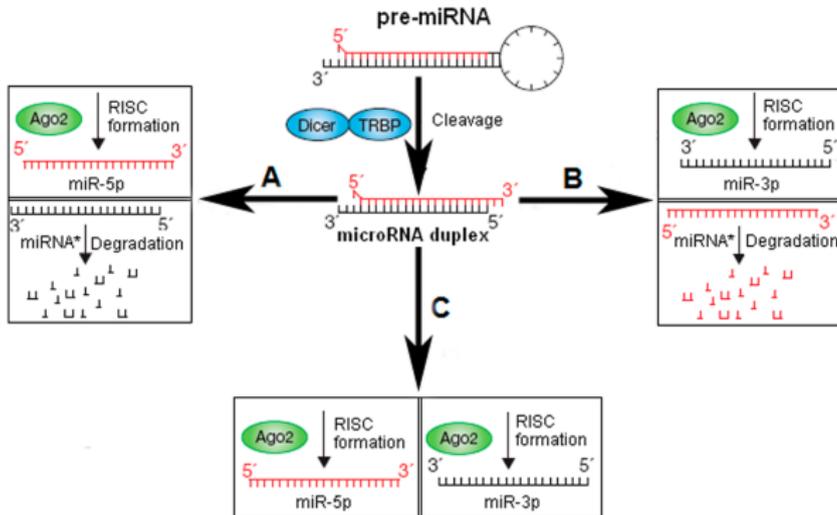
miR29

Mikroarray

GSEA

MREs

- Prinzipiell können zwei ~ 22nt Sequenzen von derselben prä-miRNA entstehen
- In der Regel herrscht eine der möglichen Sequenzen vor (z.B. **mir-123**), die andere wird abgebaut und wird z.B. **mir-123\*** bezeichnet



# mikroRNA-Nomenklatur

## MikroRNA

Peter N. Robinson

## Nomenklatur

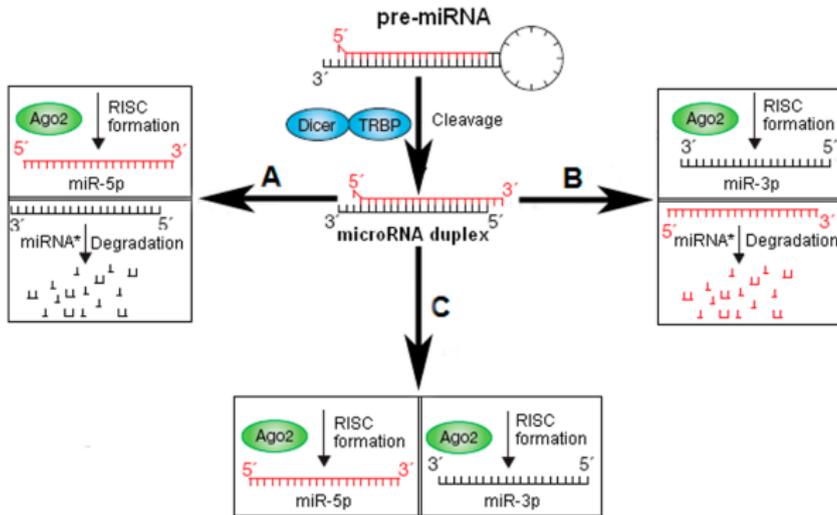
miR29

Mikroarray

GSEA

MREs

- Ist es unbekannt, welche der beiden Stränge vorherrscht, werden stattdessen Namen wie **mir-123-5p** (vom 5' Arm) und **mir-123-3p** (vom 3' Arm) verwendet.



# Outline

## MikroRNA

Peter N. Robinson

Nomenklatur

miR29

Mikroarray

GSEA

MREs

- 1 mikroRNA-Nomenklatur
- 2 MicroRNAs Differentially Expressed in Postnatal Aortic Development Downregulate Elastin via 3' UTR and Coding-Sequence Binding Sites**
- 3 Mikroarrays und globale Transkriptionsprofile
- 4 Gene-set enrichment analysis (GSEA)
- 5 MikroRNA-Ziel-Vorhersage: miRNA-regulatorischen Elementen (MREs)

# MiRNAs in der Aortenentwicklung

## MikroRNA

Peter N. Robinson

Nomenklatur

miR29

Mikroarray

GSEA

MREs

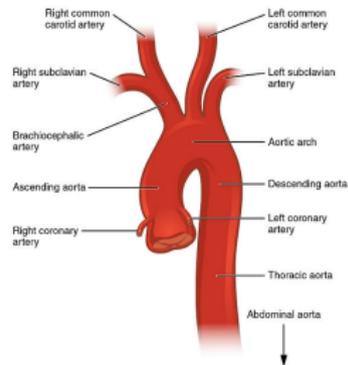
- Wir untersuchten mRNA und miRNA-Expressionsmuster in der Mauseorta in zwei Altersstufen
- Unsere Fragestellung betraf die seit langem bekannte Beobachtung, dass die Produktion von Elastin in der Aorta nach abgeschlossener Entwicklung quasi ausgeschaltet wird.
- Unsere Hypothese: Eine negative Regulation durch eine oder mehrere miRNAs könnte zu dieser Herunterregulation von Elastin beitragen

Ott CE, Grünhagen J, Jäger M, Horbelt D, Schwill S, Kallenbach K,

Guo G, Manke T, Knaus P, Mundlos S, Robinson PN (2011)

MicroRNAs differentially expressed in postnatal aortic development  
downregulate elastin via 3' UTR and coding-sequence binding sites.

*PLoS One* 6:e16250.



Grafik: Wikipedia commons

# MiRNAs in der Aortenentwicklung

MikroRNA

Peter N. Robinson

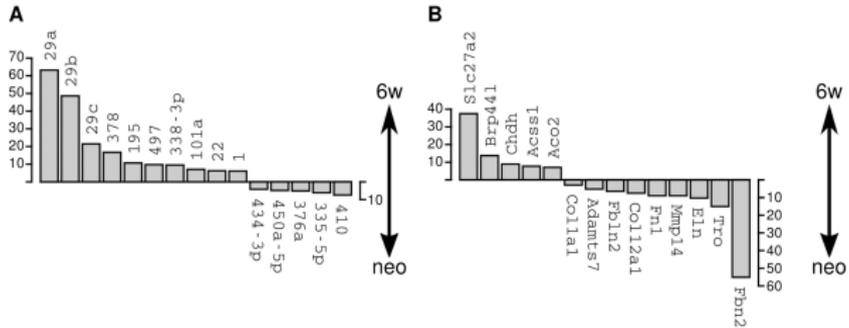
Nomenklatur

miR29

Mikroarray

GSEA

MREs



- Weitreichende Verschiebungen der Expression von mRNAs und von miRNAs während der frühen postnatalen Entwicklung (*neo*=neugeboren; *6w*=6 Wochen).
- Wie hängt das alles zusammen?

# MiRNAs in der Aortenentwicklung

## MikroRNA

Peter N. Robinson

Nomenklatur

miR29

Mikroarray

GSEA

MREs

| Significantly Higher Expression in Aortic Specimens of Neonatal Mice |   |          |             |                  |
|--|---|----------|-------------|------------------|
| ID   | Name                                      | Marginal | Study Count | Population Count |
| GO:0031012   | extracellular matrix                      | 0.969    | 71          | 269              |
| GO:004029  | regulation of gene expression, epigenetic | 0.916    | 14          | 39               |
| GO:016055  | Wnt receptor signaling pathway            | 0.821    | 30          | 129              |
| GO:007053  | transcriptional repressor complex         | 0.506    | 7           | 16               |

| Significantly Higher Expression in Aortic Specimens of 6-week old Mice |   |          |             |                  |
|--|---|----------|-------------|------------------|
| ID   | Name                                    | Marginal | Study Count | Population Count |
| GO:000739  | mitochondrion                           | 1.00     | 127         | 896              |
| GO:005811  | lipid particle                          | 0.964    | 6           | 11               |
| GO:016411  | acylglycerol O-acyltransferase activity | 0.810    | 6           | 11               |
| GO:019882  | antigen processing and presentation     | 0.734    | 17          | 49               |
| GO:043353  | erucate erythrocyte differentiation     | 0.700    | 4           | 7                |
| GO:006084  | acetyl-CoA metabolic process            | 0.541    | 23          | 32               |

For the analysis, 2,142 genes with significantly higher expression in the neo aortic samples and 1,327 genes with significantly higher expression in the w6 aortic samples were compared to all 27,827 genes represented on the microarray. Study count and population count indicate the number of genes in the study sets of differentially regulated genes and the population set of all genes on the microarray that are annotated to the GO term in question. Terms with marginal probabilities higher than 0.5 are more likely than not to be related to the observed profile of differential expression [25].  
doi:10.1371/journal.pone.0016250.t001

- Gene Ontology (GO) Analyse zeigt die charakteristischen Eigenschaften der differentiell exprimierten Gene
- Neugeborenen: Aufbau der Matrix (*extracellular matrix*)
- 6 Wochen: Energieproduktion für das schlagende Herz des nun aktiven Tiers (*mitochondrion*)

# MiRNA-Signaturen

## MikroRNA

Peter N. Robinson

Nomenklatur

miR29

Mikroarray

GSEA

MREs

- Jede miRNA kann bis zu mehrere Hundert mRNA-Moleküle regulieren
- Mindestens 20–30% aller menschlichen Gene werden von einer oder mehreren miRNAs mitgesteuert
- Dadurch, dass eine mikroRNA zahlreiche Ziel-mRNAs steuert, kann die Zelle durch eine Veränderung in der Expression von einer einzelnen mikroRNA die Expression von zahlreichen Gene beeinflussen
- Daher spielen die miRNAs eine extrem wichtige Rolle bei zellulären Genexpressionsnetzwerken

# MiRNA-Signaturen

MikroRNA

Peter N. Robinson

Nomenklatur

miR29

Mikroarray

GSEA

MREs

## Ziel der bioinformatischen Analyse

Die Vorhersage der wichtigsten miRNAs bei einem biologischen Prozess

- Unsere Daten: Globale mRNA-Expressionsdaten (mikroRNA oder RNA-seq)
- Gewünschtes Ergebnis der Analyse: Eine Liste von miRNAs, welche mit dem Expressionsprofil der mRNAs “korrelieren”

# Outline

## MikroRNA

Peter N. Robinson

Nomenklatur

miR29

Mikroarray

GSEA

MREs

- 1 mikroRNA-Nomenklatur
- 2 MicroRNAs Differentially Expressed in Postnatal Aortic Development Downregulate Elastin via 3' UTR and Coding-Sequence Binding Sites
- 3 Mikroarrays und globale Transkriptionsprofile**
- 4 Gene-set enrichment analysis (GSEA)
- 5 MikroRNA-Ziel-Vorhersage: miRNA-regulatorischen Elementen (MREs)

# Affymetrix-Technologie

MikroRNA

Peter N. Robinson

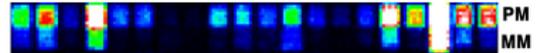
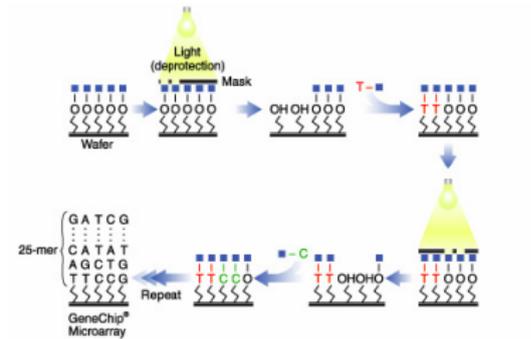
Nomenklatur

miR29

Mikroarray

GSEA

MREs



- Photolithographische Synthese von Oligonucleotiden auf Objektträgern (Mikroarrays)
- Ein Objektträger ("Chip") kann bis zu 1,6 Millionen Sonden enthalten
- Zwei 25-mer Oligonucleotide bilden ein Sondenpaar (probe pair). Hierbei weist ein Oligo eine perfekte Übereinstimmung mit der Zielsequenz auf (PM: perfect match) und das andere weist eine Fehlpaarung an der 13. Position auf (MM: mismatch)
- Die Sondenpaare erlauben die Quantifizierung des spezifischen Signals nach Abzug des durch das Mismatch-Oligo gemessene Hintergrundsignals.
- PM - MM  $\Rightarrow$  Konzentration der Zielsequenz

# Affymetrix-Technologie

## MikroRNA

Peter N. Robinson

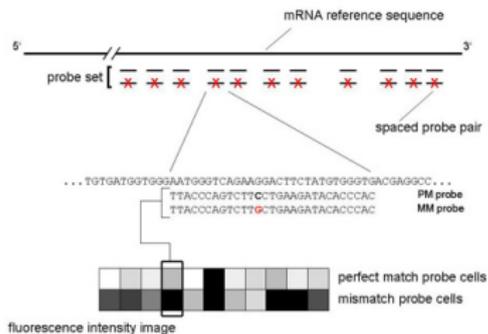
Nomenklatur

miR29

Mikroarray

GSEA

MREs



- Eine mRNA wird durch eine über die Länge des Transkriptes verstreute Reihe von Sondenpaaren detektiert ("probe set")
- Hybridisierung der fluoreszent markierte mRNA an diese Sonden wird durch einen Laser-Scanner gemessen.
- Ein probe set enthält 11 Sondenpaare – die Expression wird anhand der Intensitäten aller 11 Paare berechnet

# Affymetrix-Technologie

MikroRNA

Peter N. Robinson

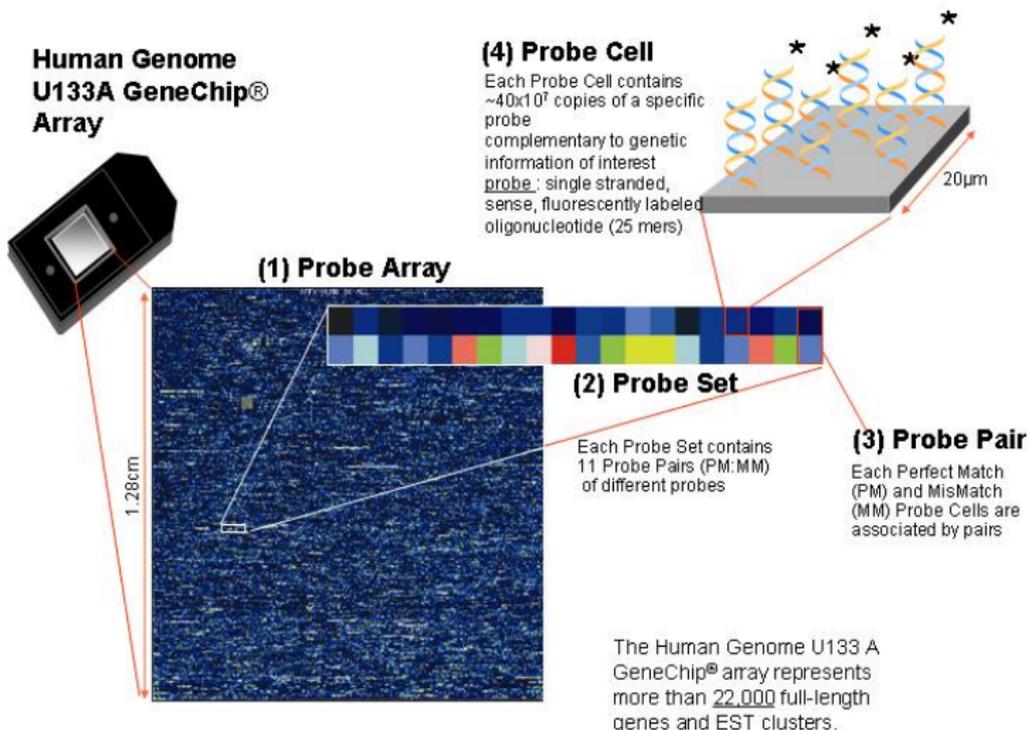
Nomenklatur

miR29

Mikroarray

GSEA

MREs



# Outline

## MikroRNA

Peter N. Robinson

Nomenklatur

miR29

Mikroarray

GSEA

MREs

- 1 mikroRNA-Nomenklatur
- 2 MicroRNAs Differentially Expressed in Postnatal Aortic Development Downregulate Elastin via 3' UTR and Coding-Sequence Binding Sites
- 3 Mikroarrays und globale Transkriptionsprofile
- 4 Gene-set enrichment analysis (GSEA)
- 5 MikroRNA-Ziel-Vorhersage: miRNA-regulatorischen Elementen (MREs)

# MiRNA-Signaturen

## MikroRNA

Peter N. Robinson

Nomenklatur

miR29

Mikroarray

GSEA

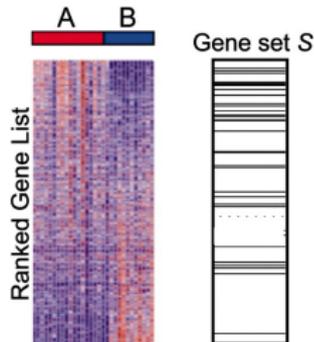
MREs

- GSEA und ähnliche Methoden: Suche nach einer "Signatur" von spezifischen miRNAs in Mikroarray oder RNA-seq Daten

Gewebsspezifische miRNAs induzieren gewebsspezifische Signaturen im Expressionsprofil der mRNAs

Sood P et al (2006) Cell-type-specific signatures of microRNAs on target mRNA expression. *Proc Natl Acad*

*Sci U S A* **103**:2746-51



Subramanian A et al (2005) Gene set enrichment analysis:

a knowledge-based approach for interpreting

genome-wide expression profiles. *Proc Natl Acad Sci U S*

*A* **102**:15545-50.



# GSEA: MiRNA-Signaturen

MikroRNA

Peter N. Robinson

Nomenklatur

miR29

Mikroarray

GSEA

MREs

## Ziel der bioinformatischen Analyse

Es gibt mehrere Variationen desselben Themas der GSEA. Ich werde eine relative einfache Version der GSEA vorstellen, s. weiterführende Literatur (letzte Folie) für Hinweise auf andere verwandte Verfahren.

- Die GSEA wird i.d.R. mit Mikroarray oder RNA-seq Daten durchgeführt.

# GSEA: Methode

MikroRNA

Peter N. Robinson

Nomenklatur

miR29

Mikroarray

GSEA

MREs

Eingabe (*input*):

- Datensatz  $D$  mit den Expressionswerten von  $N$  Genen in  $k$  Proben
- Ein Verfahren, um eine Rangliste der  $N$  Gene zu erzeugen
- Eine Menge von Genen (*gene set*), welche eine bestimmte Eigenschaft teilen (z.B. Gene Ontology Klasse, zytogenetische Bande, oder in unserem Falle den Status als Ziel-mRNA einer bestimmten miRNA)

In dieser Vorlesung werden wir der Einfachkeit halber annehmen, dass  $k = 1$  und dass wir die Gene nach ihren gemessenen Intensitäten anordnen

# Wilcoxon-Rangsummen-Test

MikroRNA

Peter N. Robinson

Nomenklatur

miR29

Mikroarray

GSEA

MREs

## GSEA: Wilcoxon

GSEA kann mit verschiedenen statistischen Tests durchgeführt werden. Einer der einfachsten ist der Wilcoxon-Rangsummen-Test

- Der Wilcoxon-Rangsummen-Test prüft, ob sich die Verteilungen der Grundgesamtheiten zweier Stichproben bezüglich ihrer Lage unterscheiden

# Rang

MikroRNA

Peter N. Robinson

Nomenklatur

miR29

Mikroarray

GSEA

MREs

## Rang

Der Rang einer Beobachtung ist die entsprechende Position in einer von der kleinsten nach der größten Beobachtung geordneten Liste.

Beispiel:

- $N$  Datenpunkte  $\{x_1, x_2, \dots, x_i, x_j, \dots, x_N\}$  mit  $x_i \in \mathbb{R}$
- Der Rang von  $x_i$ , den wir als  $R(x_i)$  angeben werden, ist dann einfach die Anzahl aller Datenpunkte, die kleiner gleich  $x_i$  sind:

$$R(x_i) = \sum_{j=1}^N \mathbf{1}_{x_j \leq x_i} \quad (1)$$

# Wilcoxon

## MikroRNA

Peter N. Robinson

Nomenklatur

miR29

Mikroarray

GSEA

MREs

- Beispiel: Blutdruckwerte von Patienten, welche mit einer neuen Medikationen behandelt werden (Gruppe 1,  $n$  Probanden) oder die ein Placebo erhalten (Gruppe 2,  $m$  Probanden)
- Seien nun  $S_1, \dots, S_n$  die Ränge der  $n$  Werte der Gruppe 1 innerhalb der  $N = n + m$  Werte aller Probanden (aus beiden Gruppen)
- Der Wertebereich für  $S_i$  ist also  $\{1, 2, \dots, N\}$
- Unter der Annahme, dass keine zwei Werte gleich sind, wird die Wilcoxon-Statistik definiert als

$$W_S = S_1 + S_2 + \dots + S_n \quad (2)$$

# Wilcoxon

## MikroRNA

Peter N. Robinson

Nomenklatur

miR29

Mikroarray

GSEA

MREs

- Der kleinste Wert von  $W_S$  gilt dann, wenn die Proben von Gruppe 1 alle kleiner sind als die Werte von Gruppe 2. Dann

$$W_S = 1 + 2 + \dots + n = n(n+1)/2 \quad (3)$$

- Falls die Werte von Gruppe 1 alle größer sind dann

$$\begin{aligned} W_S &= (N-n+1) + (N-n+2) + \dots + N \\ &= (N-n) + 1 + (N-n) + 2 + \dots + (N-n) + n \\ &= n(N-n) + (1+2+\dots+n) \\ &= n(N-n) + n(n+1)/2. \end{aligned}$$

- Daher können wir den Wertebereich von  $W_S$  angeben als

$$W_S \in \{n(n+1)/2, n(n+1)/2 + 1, \dots, n(N-n) + n(n+1)/2\} \quad (4)$$

# Wilcoxon: Beispiel

## MikroRNA

Peter N. Robinson

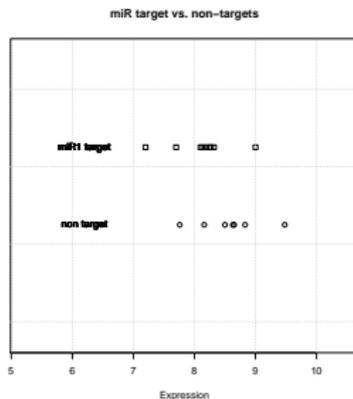
Nomenklatur

miR29

Mikroarray

GSEA

MREs



---

|             |      |      |      |      |      |      |      |      |      |
|-------------|------|------|------|------|------|------|------|------|------|
| miR1 target | 8.50 | 9.48 | 8.65 | 8.16 | 8.83 | 7.76 | 8.63 | -    | -    |
| non target  | 8.27 | 8.20 | 8.25 | 8.14 | 9.00 | 8.10 | 7.20 | 8.32 | 7.70 |

---

- Sind die Daten für miR-Zielgene und Nicht-Zielgene identisch?
- Ist der Mittelwert signifikant unterschiedlich?
- Sind die Daten normal verteilt?

# Wilcoxon: Beispiel

MikroRNA

Peter N. Robinson

Nomenklatur

miR29

Mikroarray

GSEA

MREs

- Wir haben also Stichproben von zwei Populationen (miR1-Ziele und Nicht-Ziele, im folgenden **A** und **B**)
- Wir wollen die Hypothese prüfen, dass A und B eine identische Verteilung haben
- $H_0: A=B$
- $H_A: A<B$
- Der Wilcoxon-Test basiert auf der Rangfolge der  $n_A$  bzw.  $n_B$  Beobachtungen von Gruppe A und B
- $\mathbf{w}_A$ =Summe der Ränge der Beobachtungen von Gruppe A

# Wilcoxon: Beispiel

MikroRNA

Peter N. Robinson

Nomenklatur

miR29

Mikroarray

GSEA

MREs

---

|        |      |      |      |      |      |      |      |      |      |
|--------|------|------|------|------|------|------|------|------|------|
| Wert   | 7.20 | 7.70 | 7.76 | 8.10 | 8.14 | 8.16 | 8.20 | 8.25 | 8.27 |
| Gruppe | A    | A    | B    | A    | A    | B    | A    | A    | A    |
| Rang   | 1    | 2    | 3    | 4    | 5    | 6    | 7    | 8    | 9    |

---

(fortgesetzt)

---

|        |      |      |      |      |      |     |      |
|--------|------|------|------|------|------|-----|------|
| Wert   | 8.32 | 8.50 | 8.63 | 8.65 | 8.83 | 9.0 | 9.48 |
| Gruppe | A    | B    | B    | B    | B    | A   | B    |
| Rang   | 10   | 11   | 12   | 13   | 14   | 15  | 16   |

---

- $w_A = 1 + 2 + 4 + 5 + 7 + 8 + 9 + 10 + 15 = 61$
- $w_B = 3 + 6 + 11 + 12 + 13 + 14 + 16 = 75$

# Wilcoxon: Beispiel

MikroRNA

Peter N. Robinson

Nomenklatur

miR29

Mikroarray

GSEA

MREs

- Wie können wir nun einen P-Wert für  $w_A = 61$  berechnen?
- Um eine Antwort auf diese Frage zu erhalten, müssen wir das erwartete Verhalten von Rangsummen unter  $H_0$  kennen
- Das ist im Prinzip die Verteilung, die man dann erhält, wenn man die Markierung aller Datenpunkte als “A” oder “B” randomisiert.
- Unsere Hypothese lautet, dass miR-Zielgene eine geringere Expression aufweisen als die Nicht-Zielgene (durch die repressorische Wirkung von mikroRNAs).
- Daher lautet  $H_A: A < B$

# Wilcoxon: Beispiel

MikroRNA

Peter N. Robinson

Nomenklatur

miR29

Mikroarray

GSEA

MREs

- Wir bezeichnen mit  $W_A$  die Verteilung von  $w_A$  unter der Nullhypothese  $H_0$ . Dann beträgt der  $P$ -Wert für den Wilcoxon-Test

$$P\text{-Wert} = \Pr(W_A \leq w_A) \quad (5)$$

# Wilcoxon: in R

MikroRNA

Peter N. Robinson

Nomenklatur

miR29

Mikroarray

GSEA

MREs

```
> A <- c(8.50,9.48,8.65,8.16,8.83,7.76,8.63)
> B <- c(8.27,8.20,8.25,8.14,9.00,8.10,7.20,8.32,7.70)
> wilcox.test(B,A,paired=FALSE,alternative="less")
```

Wilcoxon rank sum test

data: B and A

W = 16, p-value = 0.05708

alternative hypothesis: true location shift is less than 0

# Wilcoxon: Woher kommen die P-Werte?

MikroRNA

Peter N. Robinson

Nomenklatur

miR29

Mikroarray

GSEA

MREs

- Für kleine Stichproben kann der P-wert  $P\text{-Wert} = \Pr(W_A \leq w_A)$  im Prinzip durch Aufzählung der möglichen Rangsummen berechnet werden, wobei jede Permutation der Ränge als gleich wahrscheinlich betrachtet wird
- Tabellen stehen zur Verfügung (wie beim Chiquadratstest usw)
- Für größere Stichproben gilt eine Approximation durch eine Normalverteilung.

# Wilcoxon: Woher kommen die P-Werte?

MikroRNA

Peter N. Robinson

Nomenklatur

miR29

Mikroarray

GSEA

MREs

## Lemma 1

Hat die Stichprobe 1  $n_1$  Beobachtungen und die Rangsumme  $R_1$  und Stichprobe 2  $n_2$  und die Rangsumme  $R_2$ , dann gilt  $R_1 + R_2 = n(n+1)/2$ , wobei  $n = n_1 + n_2$ .

Die Summe der ersten  $n$  positiven Ganzzahlen beträgt  $\frac{n(n+1)}{2}$  (Gauß).

# Wilcoxon: Woher kommen die P-Werte?

MikroRNA

Peter N. Robinson

Nomenklatur

miR29

Mikroarray

GSEA

MREs

## Lemma 2

Haben die beiden Stichproben einen ausreichenden Umfang (z.B.  $n > 20$ ), dann ist  $w_A$  annähernd normal verteilt  $\mathcal{N}(\mu, \sigma)$  mit  $\mu = n_1 \cdot (n_1 + n_2 + 1)/2$

Sei  $x_i$  der Rang von Beobachtung  $i$  in Stichprobe 1. Dann können wir den Erwartungswert für den Rang einer einzelnen Beobachtung berechnen als:

$$\begin{aligned}\mathbb{E}[x_i] &= \frac{1}{n} \cdot \sum_{i=1}^n i \\ &= \frac{1}{n} \cdot \frac{n(n+1)}{2} \\ &= \frac{n+1}{2}\end{aligned}$$

# Wilcoxon: Woher kommen die P-Werte?

MikroRNA

Peter N. Robinson

Nomenklatur

miR29

Mikroarray

GSEA

MREs

Nun können wir den Erwartungswert für  $R_1$  berechnen

$$\begin{aligned}\mu &= \mathbb{E}[R_1] \\ &= \mathbb{E}\left[\sum_{i=1}^{n_1} x_i\right] \\ &= \sum_{i=1}^{n_1} \mathbb{E}[x_i] \\ &= n_1 \cdot \frac{n+1}{2} \\ &= n_1 \cdot \frac{n_1 + n_2 + 1}{2}\end{aligned}$$

# Wilcoxon: Woher kommen die P-Werte?

MikroRNA

Peter N. Robinson

Nomenklatur

miR29

Mikroarray

GSEA

MREs

## Lemma 3

Haben die beiden Stichproben einen ausreichenden Umfang (z.B.  $n > 20$ ), dann ist  $w_A$  annähernd normal verteilt  $\mathcal{N}(\mu, \sigma)$  mit  $\sigma = \frac{n_1 n_2}{12} \cdot (n_1 + n_2 + 1)$

Beweis: ähnlich

# Wilcoxon: Woher kommen die P-Werte?

MikroRNA

Peter N. Robinson

Nomenklatur

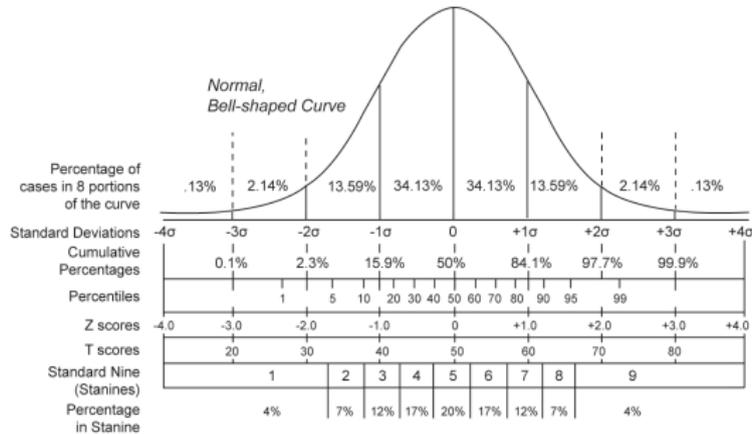
miR29

Mikroarray

GSEA

MREs

- Mit den Werten für  $\mu$  und  $\sigma$  Können wir einen P-Wert berechnen



## Z-Score

$$z = \frac{x - \mu}{\sigma} \quad (6)$$

# Wilcoxon für die Analyse experimenteller Daten?

MikroRNA

Peter N. Robinson

Nomenklatur

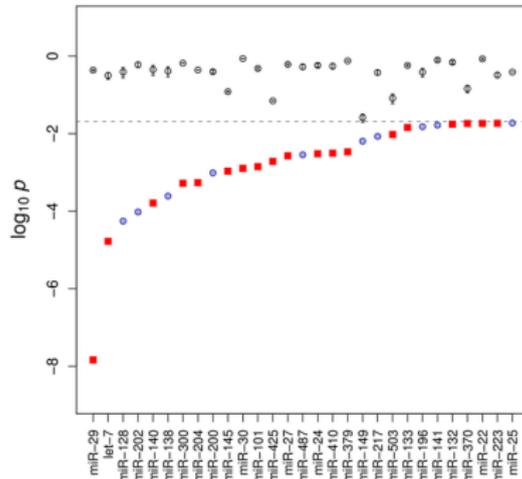
miR29

Mikroarray

GSEA

MREs

Für jede der 567 mikroRNAs für die Expressionswerte mittels Mikroarray gewonnen wurden, ist ein Wilcoxon-Test wie oben beschrieben durchgeführt worden.



Rot: stimmt mit experimenteller Daten überein

# Wilcoxon für die Analyse experimenteller Daten?

MikroRNA

Peter N. Robinson

Nomenklatur

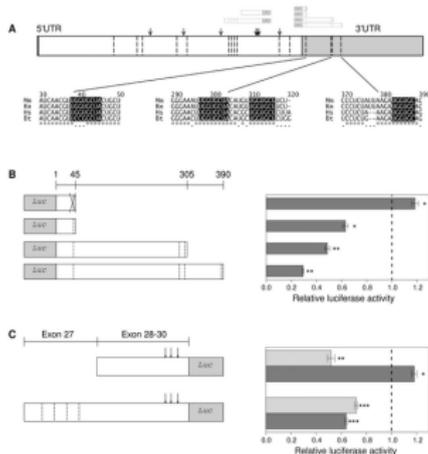
miR29

Mikroarray

GSEA

MREs

Schlussfolgerung: miR-29 ist überexprimiert und beeinflusst das Genexpressionsprofil der Aorta



Unsere Ergebnisse unterstützten die Hypothese, dass diese miRNA in der adulten Aorta hoch exprimiert ist und dadurch einen Faktor für die physiologische Unterdrückung der Elastinproduktino im erwachsenen Organismus darstellen könnte.

# Outline

## MikroRNA

Peter N. Robinson

Nomenklatur

miR29

Mikroarray

GSEA

MREs

- 1 mikroRNA-Nomenklatur
- 2 MicroRNAs Differentially Expressed in Postnatal Aortic Development Downregulate Elastin via 3' UTR and Coding-Sequence Binding Sites
- 3 Mikroarrays und globale Transkriptionsprofile
- 4 Gene-set enrichment analysis (GSEA)
- 5 MikroRNA-Ziel-Vorhersage: miRNA-regulatorischen Elementen (MREs)



# Grundlagen der Vorhersage von miRNA-Zielgenen: Das 5'-Ende

## MikroRNA

Peter N. Robinson

Nomenklatur

miR29

Mikroarray

GSEA

MREs

- 5'-Ende der miRNA: 'Saat' (seed)
- 6–8 Nukleotide, perfekte Basenpaarung mit Zielgen
- Das 5'-Ende ist typischerweise ungepaart, oder beginnt mit Uracil und enthält i.d.R. keine G:U-Wobble-Paare
- Die Bindungsstelle im Zielgen ist oft von Adenosinresten flankiert.

# Grundlagen der Vorhersage von miRNA-Zielgenen: Das 3'-Ende

## MikroRNA

Peter N. Robinson

Nomenklatur

miR29

Mikroarray

GSEA

MREs

- Perfekte Basenpaarung am 3'-Ende kann eine imperfekte Basenpaarung am 5'-Ende kompensieren
- Viele miRNA-Bindungsstellen haben einen 'Buckel' (Bulge) im mittleren Bereich

|        |  |    |                |                            |
|--------|--|----|----------------|----------------------------|
| 3'     | GAUGGUAUCCCAUU--UUGGUGAc   | 5' | hsa-miR-140-5p | Alignment score: 161.0     |
|        |  |    |                | PhastCons score: 0.6432655 |
| 164:5' | CUA <sup>164</sup> CCCUAGUUC <sup>168</sup> AAAA <sup>172</sup> CA <sup>176</sup> CCACU <sup>180</sup> u | 3' | FBN1           | Energy: -16.02             |

Bindungsstelle für miRNA-140 in *FBN1*





# Klassifikation der miRNA-Bindungsstellen

MikroRNA

Peter N. Robinson

Nomenklatur

miR29

Mikroarray

GSEA

MREs

**7mer m8:** Genaue Übereinstimmung der Nukleotide 2-8 der reifen miRNA (Seed + Position 8) gefolgt von einem anderen Nukleotid außer A





# Klassifikation der miRNA-Bindungsstellen

MikroRNA

Peter N. Robinson

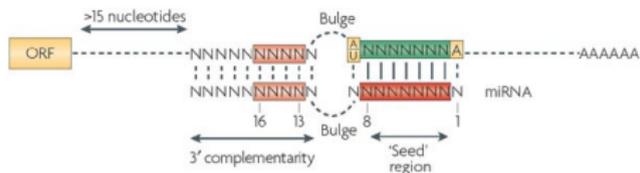
Nomenklatur

miR29

Mikroarray

GSEA

MREs



Filipowicz (2008) Nature Reviews Genetics 9,102–114

- Fehlpaarungen (“bulge”) im mittleren Abschnitt des miRNA–mRNA Duplexes
- teilweise Komplementarität des 3' Bereiches der miRNA, insbesondere Nukleotide 16–16

# Bewertung der miRNA-Bindungsstellen

## MikroRNA

Peter N. Robinson

Nomenklatur

miR29

Mikroarray

GSEA

MREs

Eine Suche nach MREs allein auf Grund einer Sequenzidentität findet zahlreiche falsch positive MREs. Die biologischen Regeln der mikroRNA:mRNA Interaktionen sind noch nicht vollständig geklärt.

- Stärke: m8 > 7mer-m8 > 7mer-A1 >> 6mer
- Sequenzkontext (AU reich)
- Sekundärstruktur (“Zugänglichkeit”)
- Cluster von MREs
- Bindungen anderer Sequenzabschnitte

Bartel DP (2009) MicroRNAs: target recognition and regulatory functions. Cell **136**:215-33.

# Vorhersage von miRNA-Zielgenen

## MikroRNA

Peter N. Robinson

Nomenklatur

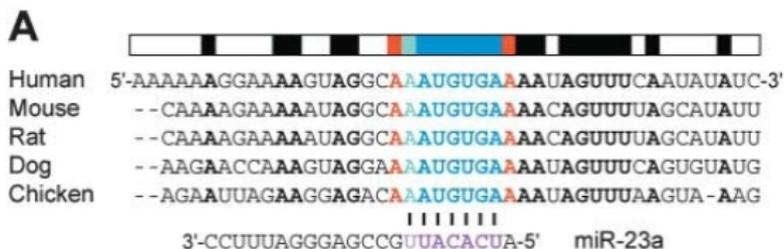
miR29

Mikroarray

GSEA

MREs

- Konservierte miRNA-Saatsequenzen, oft mit flankierenden A's, deuten auf miRNA-Zielgene. Alignment der orthologen Sequenzen der 3'-UTR des HIC-Gens mit konservierter Saatsequenz für miR-23a<sup>1</sup>





# MikroRNA-Ziel-Vorhersage

## MikroRNA

Peter N. Robinson

Nomenklatur

miR29

Mikroarray

GSEA

MREs

- Reicht es zur Identifikation von miRNA-regulatorischen Elementen (MREs) nach Sequenzkomplementarität zu suchen?
- Überschlagsrechnung
  - typische 3' UTR hat eine Länge von 700 nt
  - ca. 20.000 proteinkodierende Gene (nehmen wir hier 1 Transkript pro Gen an), daher  $20.000 \times 700 \text{ nt}$  3' UTR Sequenz im Genom
  - kürzeste MRE-Sequenz umfasst 6 nt, d.h., eine bestimmte 6-nt Sequenz begegnet per Zufall ca. alle  $4^6 = 4096$  nt
  - Für eine bestimmte miRNA erwarten wir also  $\frac{20.000 \times 700 \text{ nt}}{4096} \approx 3148$  "Hits" per Zufall
  - ca. 1000 miRNA sind im menschlichen Genom bekannt
- Man findet also reichlich MRE-"Kandidaten" per Zufall
- Die Bioinformatik versucht daher auch andere Eigenschaften zu berücksichtigen um die Spezifität der Vorhersagen zu erhöhen.



# miRanda: DP

MikroRNA

Peter N. Robinson

Nomenklatur

miR29

Mikroarray

GSEA

MREs

## miRanda

Der miRanda-Algorithmus besitzt eine Ähnlichkeit mit dem Smith-Waterman-Algorithmus. Statt jedoch das Alignment anhand von übereinstimmenden Nukleotiden zu berechnen (z.B. A–A, oder G–G), bewertet miRanda die Komplementarität der Nukleotide (A=U or G≡C) bzw. auch Wobble-Paare (G-U)

Score-Matrize  $M(i,j)$ :

- G≡C:  $M(i,j)=+5$
- A=U:  $M(i,j)=+5$
- G-U:  $M(i,j)=+2$
- $M(i,j)=-3$  für alle anderen Nukleotidpaare

# miRanda: DP

## MikroRNA

Peter N. Robinson

Nomenklatur

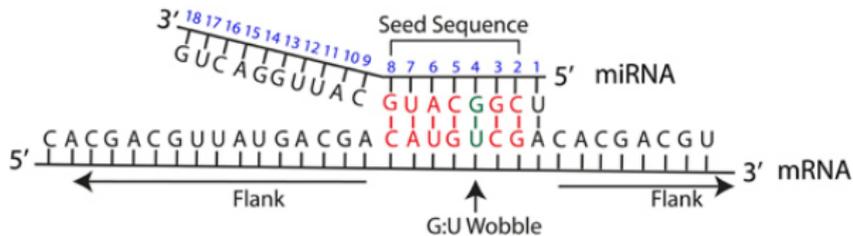
miR29

Mikroarray

GSEA

MREs

- Affine Gap-penalties (linear in der Anzahl der Nukleotide in der “Lücke”): Gap-Eröffnen: -8, Gap-Verlängern: -2
- Der Score in den ersten 11 Positionen der miRNA wird durch zwei multipliziert, um die besondere Wichtigkeit des 5'-Bereichs der miRNA zu berücksichtigen



# miRanda: DP

MikroRNA

Peter N. Robinson

Nomenklatur

miR29

Mikroarray

GSEA

MREs

Der DP-Algorithmus von miRanda funktioniert ähnlich wie bei Smith Waterman indem die Matrix  $M$  Schritt für Schritt berechnet wird:

$$M(i, j) = \max \begin{cases} 0 & \\ M(i-1, j-1) + S(i, j) & \text{match/mismatch} \\ \max_{k \geq 1} \{M(i-k, j) + W_k\} & \text{Deletion} \\ \max_{l \geq 1} \{M(i, j-l) + W_l\} & \text{Insertion} \end{cases} \quad 1 \leq i \leq m, 1 \leq j \leq n \quad (7)$$



# miRanda: freie Energie

MikroRNA

Peter N. Robinson

Nomenklatur

miR29

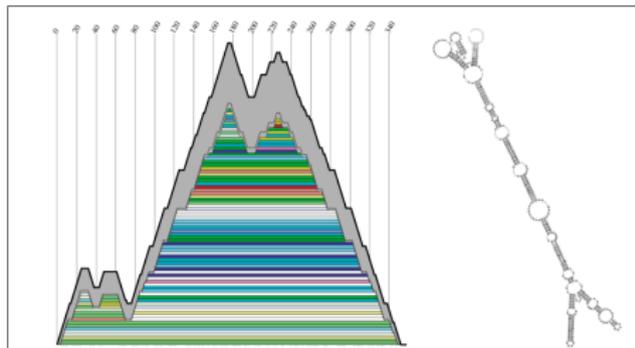
Mikroarray

GSEA

MREs

## Min. freie Energie

Die Berechnung der freien Energie des miRAN.mRNA-Alignments erfolgt im Prinzip wie beim letztes Mal vorgestellten DP Algorithmus



"Mountant plot": Eine Basenpaarung wird durch eine "Scheibe" dargestellt

# miRanda: freie Energie

## MikroRNA

Peter N. Robinson

Nomenklatur

miR29

Mikroarray

GSEA

MREs

- Vienna 1.3 RNA secondary structure programming library (RNAlib)
- Die miRNA-Sequenz und die 3' UTR Sequenz werden zu einer einzelnen Sequenz durch einen 8nt-“Linker” (XXXXXXXX) kombiniert
- Die minimale freie Energie aller Strukturen wird berechnet

# miRanda: Sequenzkonservierung

## MikroRNA

Peter N. Robinson

Nomenklatur

miR29

Mikroarray

GSEA

MREs

- All miRNA-Sequenzen werden dann gegen alle 3' UTR-Sequenzen im Genom verglichen
- Schwellen für Hit-Detektion
  - 1 Smith-Waterman Alignment  $S \geq 80$
  - 2 Freie Energie des miRNA:mRNA-Duplex  $\Delta G \leq -14$  kcal/mol.
  - 3 Hits werden unter evolutionär verwandten Organismen verglichen (z.B. D. melanogaster und D. pseudoobscura oder Mensch und Maus).
  - 4 Alignments der Zielsequenzen werden über eine geteilte (oder homologe) miRNA definiert (UTR→miRNA→UTR)
  - 5 Positionen der Zielsequenzen sollen dann innerhalb von  $\pm 10$  nt in den alignierten 3UTR Sequenzen fallen und eine Sequenzidentität von  $\geq 90\%$  (Maus-Mensch) aufweisen



# miRanda

## MikroRNA

Peter N. Robinson

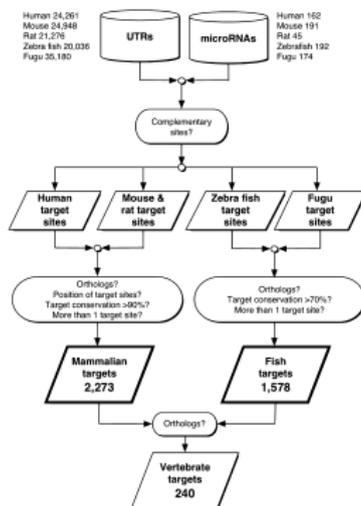
Nomenklatur

miR29

Mikroarray

GSEA

MREs



- Konservierte miRNA-Ziele in Säugern bzw. Fischen
- Kleinere Menge von “super-conserved” Zielen (Vertebraten)

# miRanda

## MikroRNA

Peter N. Robinson

Nomenklatur

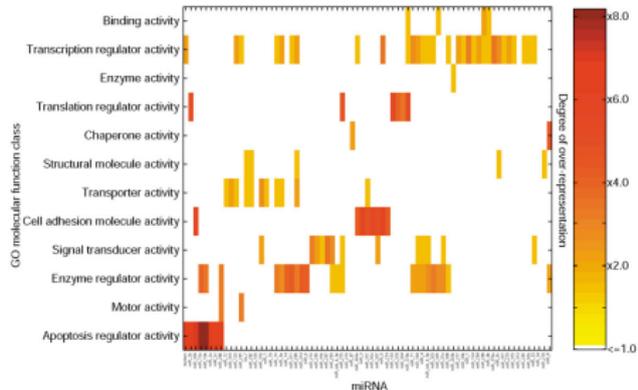
miR29

Mikroarray

GSEA

MREs

- Funktionen von miRNAs und deren Zielgenen
- Links: GO-Klassen
- X-Achse: 73 miRNAs
- Farbkodierung: Grad der Überrepräsentierung (z.B. hell rot: 6-8mal mehr Ziele in der GO-Klasse als per Zufall erwartet)



# zum Schluss

## MikroRNA

Peter N. Robinson

Nomenklatur

miR29

Mikroarray

GSEA

MREs

- Email: [peter.robinson@charite.de](mailto:peter.robinson@charite.de)

## weiterführende Literatur

- John B et al (2004) Human MicroRNA targets. *PLoS Biol* **2**:e363. (miRanda-Algorithmus)
- Bartel DP (2009) MicroRNAs: target recognition and regulatory functions. *Cell* **136**:215-33.