Andrej Aderhold

Problemseminar Thema

Inference of miRNA targets using evolutionary conservation and pathway analysis

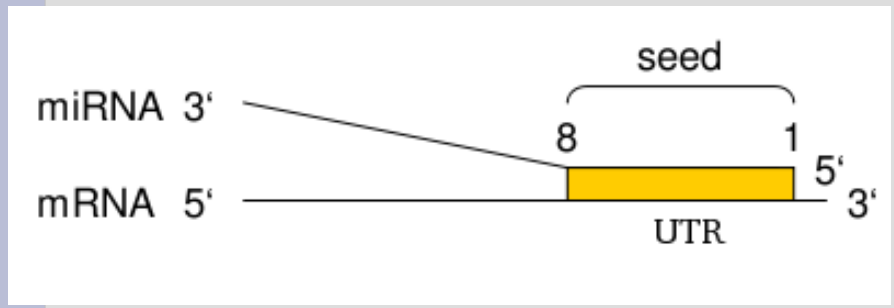
Gaidatzis, Nimwegen, Hausser, Zavolan, March2007, BMC Bioinformatics

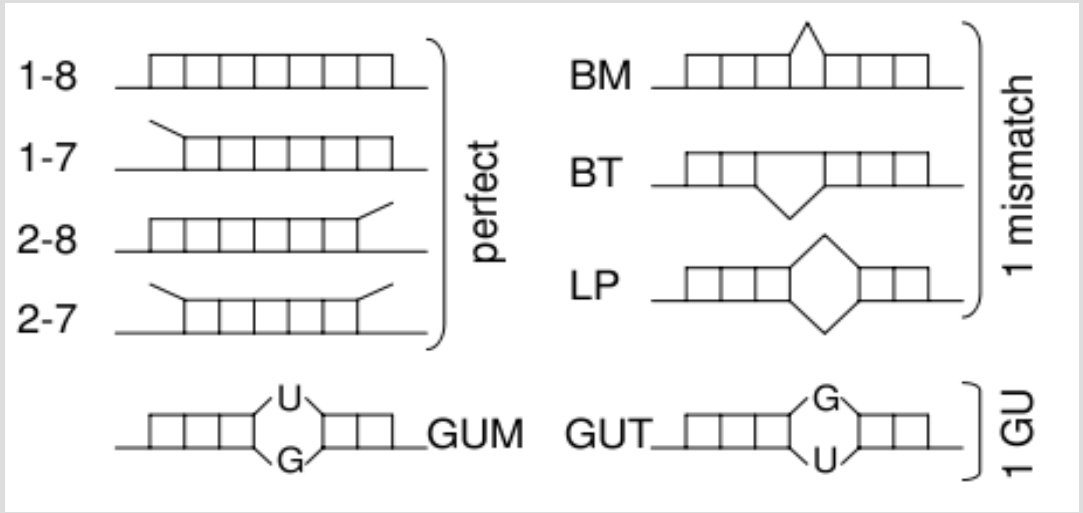
Übersicht

- miRNA – Ziel Interaktion: 'seed' Typen
- MiRNA Ziel Vorhersage: Anwenden eines Bayesianischen phylogenetischen Modells
- miRNA Funktion: pathway analysis (KEGG)

Seed Typen 1

- Ersten 8 Nucleotide am 5' Ende der miRNA spielen eine Rolle bei Bindung zum Ziel
- Welche Bindungskonstellation (seed type) ist günstig?
- Diejenigen mit stärkster Konservierung in orthologen Genen
- Analyse von 9 verschiedenen 'seed' Typen





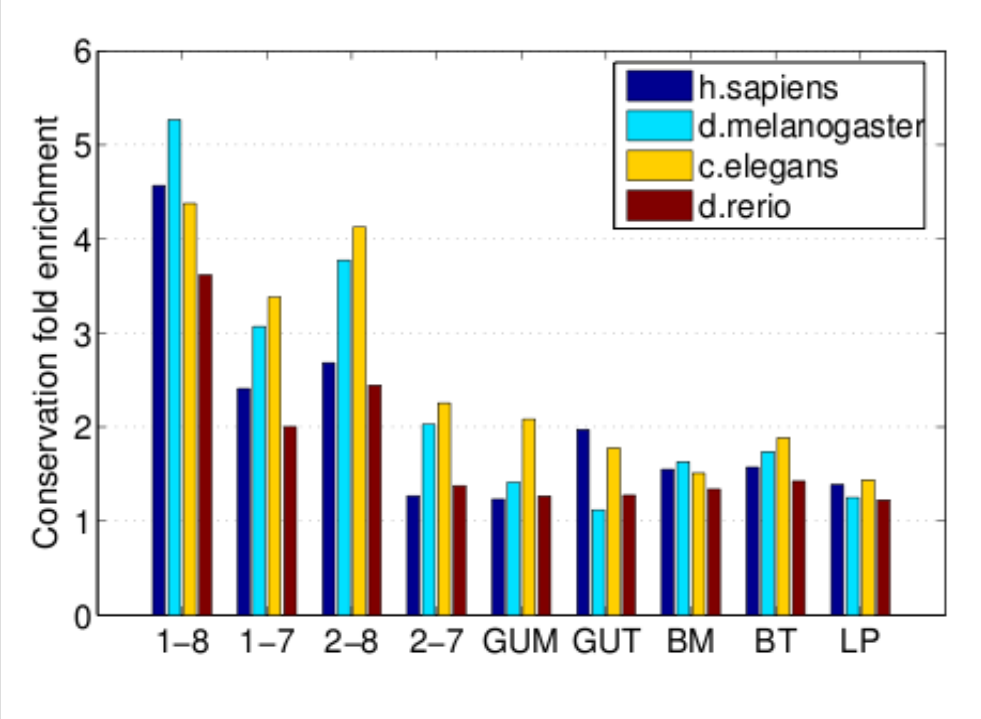
Seed types 2 – Konservierungs Statistik

- Idee: Umso stärker die Konservierung des miRNA Ziels in orthologen Genen umso signifikanter das Ziel (Indiz für Funktionalität)

T : Anteil der Ziele die perfekt konserviert in allen Spezien einer Menge von verwandten Arten sind

B : Anteil der perfekt konservierten Zufallssequenzen aus den 3' UTRs

T/B : conservation fold enrichment. Umso höher der Wert, umso stärker die Konservierung

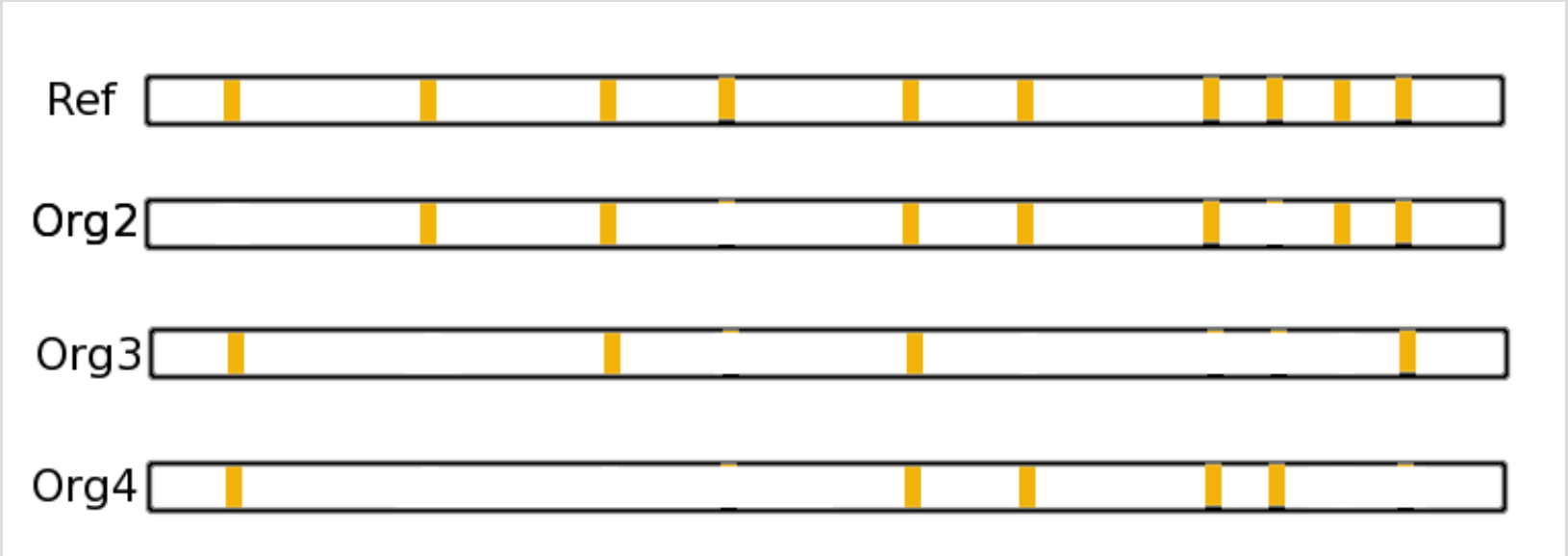


Ermitteln eines miRNA Ziels: populäre Methode 1

- Suche miRNA Ziele in Referenz Organismus
- Identifiziere orthologe Sequenzen in verwandten Organismen

(pairwise alignment)

- Nutze evolutionäre Konservierung als statistisches Mass für die Signifikanz der Funktionalität eines Zieles



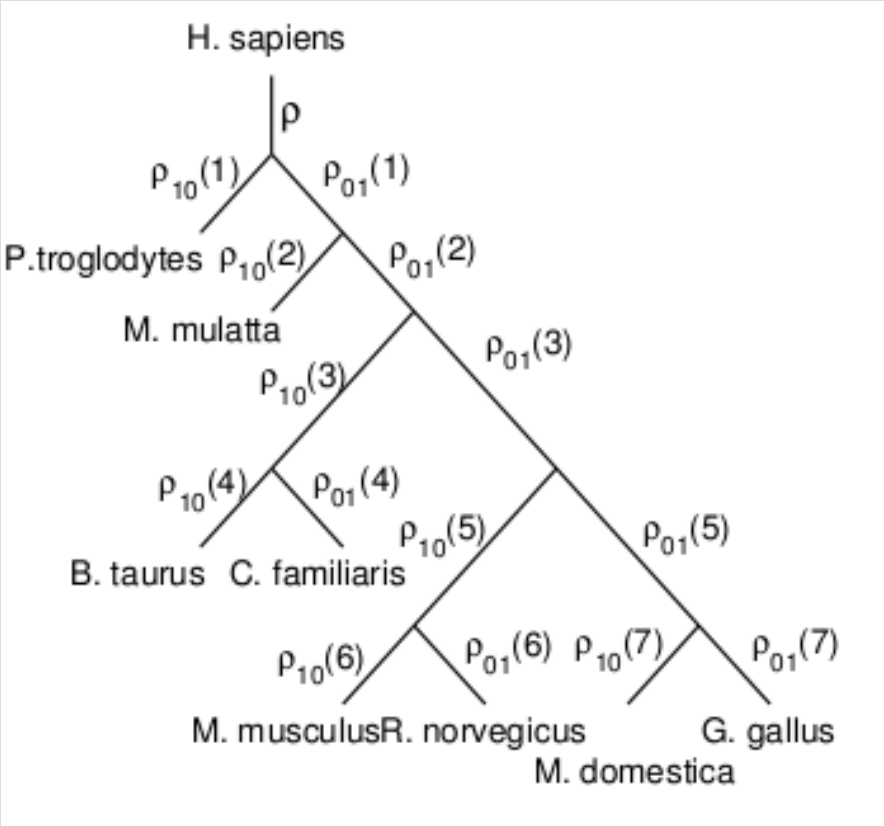
Ermitteln eines Zieles: populäre Methode 2

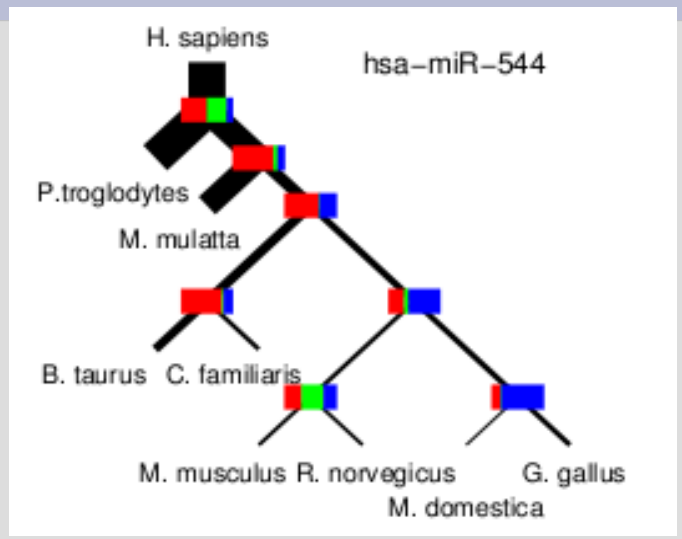
- Nachteile:
  - Phylogenetische Beziehung bei Gewichtung der Konservierung wird nicht eingerechnet
  - alle miRNAs werden gleich behandelt.
- Doch: Selektionsdruck auf funktionale Ziele kann zwischen verwandten Arten für verschiedene miRNAs verschieden sein

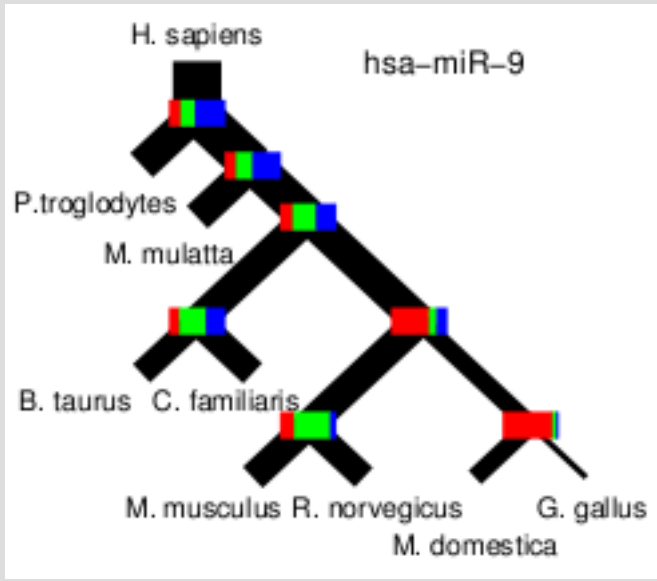
-> starker Einfluss auf Konservierungsstatistik

MiRNA Ziel Vorhersage 1

Für jede miRNA wird die Evolution von orthologen miRNA Ziel Seqünzen modelliert.







MiRNA Ziel Vorhersage 2

Die posterior Wahrscheinlichkeit der Bayes Methode gibt den Grad der Funktionalität des miRNA Ziels im Referenzorganismus wieder (Also für alle Ziele in Homo-sapiens bei Saugetieren z.B.).

MiRNA Ziel Vorhersage 3

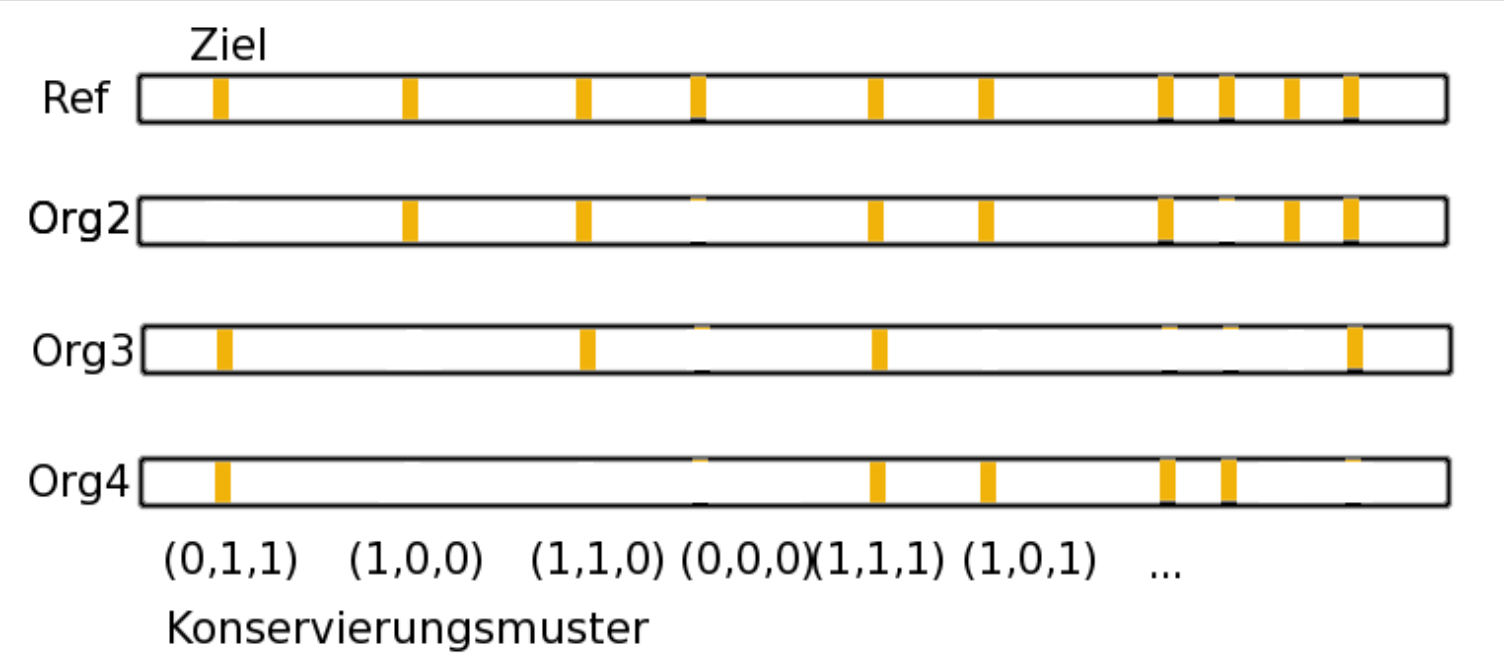
Wir brauchen:

- Konservierungsmuster des miRNA Ziels
- Konservierung des miRNA Gens
- Hintergrund Konservierung (noise)
- Selektionsmuster der einzelnen miRNA Ziele (simplifiziert)

Für jede miRNA kann dann die Wahrscheinlichkeitsverteilung der Selektionsmuster (prior) bestimmt werden.
-> Ziel: Angleich an die beobachteten Frequenzen der Konservierungsmuster durch Maximierung (likelihood, EM)

Konservierungsmuster 1

- Jedem miRNA Ziel wird ein binärer Vektor c zugeordnet mit c_i=1, falls Ziel konserviert in Organismus i, c_i=0 andernfalls.



Konservierungsmuster 2

- Ermittle die Frequenzen der Konservierungsmustern über alle miRNA Ziele des Referenzorganismus:

  n(c,t), wobei t dem seed Typ entspricht

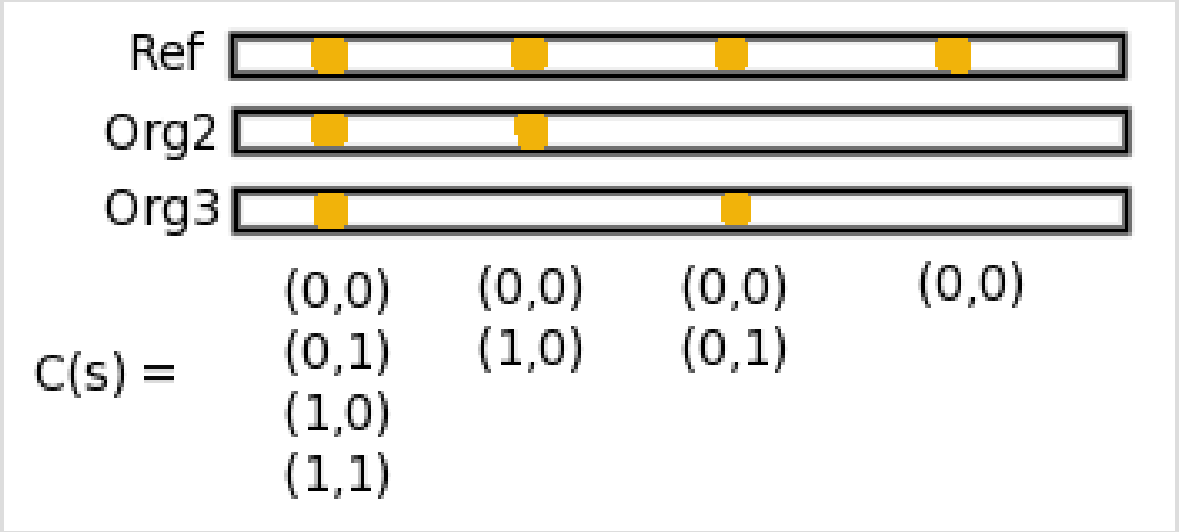
Hintergrund Modell (noise)

- Ermittle relative Frequenzen aller möglichen Konservierungsmuster unter zufällig gewählten Seqünzabschnitten

  - wähle zufällige Abschnitte aus 3' UTR der Referenz (konform zu seed Typ!)
  - suche orthologe Abschnitte in verwandten Arten und erstelle jeweils Konservierungsmuster
  - Ermittle relative Freqünz für jedes Konservierungsmuster über alle ermittelten Randomabschitte

Selektionsmuster

- Jede konservierte Stelle kann 2 Zustände haben:
  Funktional: Selektionsdruck hat gewirkt und Abschnitt konserviert
  Nicht-Funktional: Abschnitt hat sich gemäss Hintergrund Modell entwickelt
- Jedem Konservierungsmuster kann eine Menge von Selektionsmustern (C(s)) zugeordnet werden mit denen es konsistent ist.



Wahrscheinlichkeit

- Daraus lässt sich die Wahrscheinlichkeit

$$p(\vec{c}|t,\vec{s}) = \frac{p(\vec{c}|t,bg)}{\sum_{\vec{c}\,'\in C(\vec{s})} p(\vec{c}\,'|t,bg)}$$

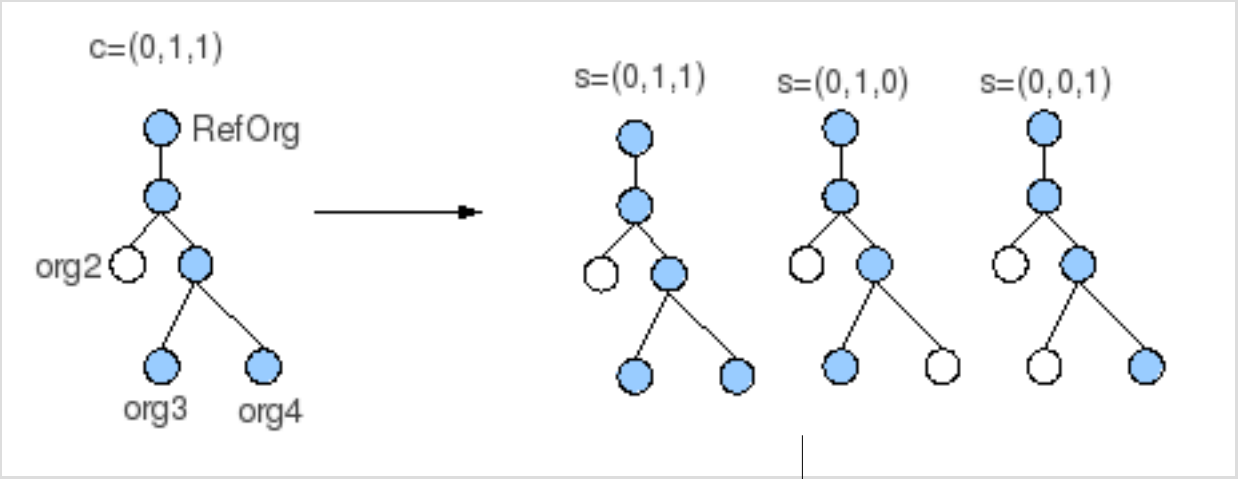
Pr die c zufällig beobachtet unter dem Hintergrund Modell mit seed Typ t

Alle c konsistent mit s

ableiten: die Möglichkeit ein Konservierungsmuster c mit gegebenen Selektionsmuster s zu beobachten (und seed Typ t).

p(s) 1:
A Priori Wahrscheinlichkeitsverteilung der Selektionsmuster

- Wie Wahrscheinlich ist es, dass ein miRNA Ziel in einer bestimmten Untermenge der Verwandten Arten unter Selektion steht. (blau : konserviert(links), unter Selektion(rechts)).



p((0,1,1)),p((0,1,0)), p((0,0,1)) gehen a priori in das Bayes Modell

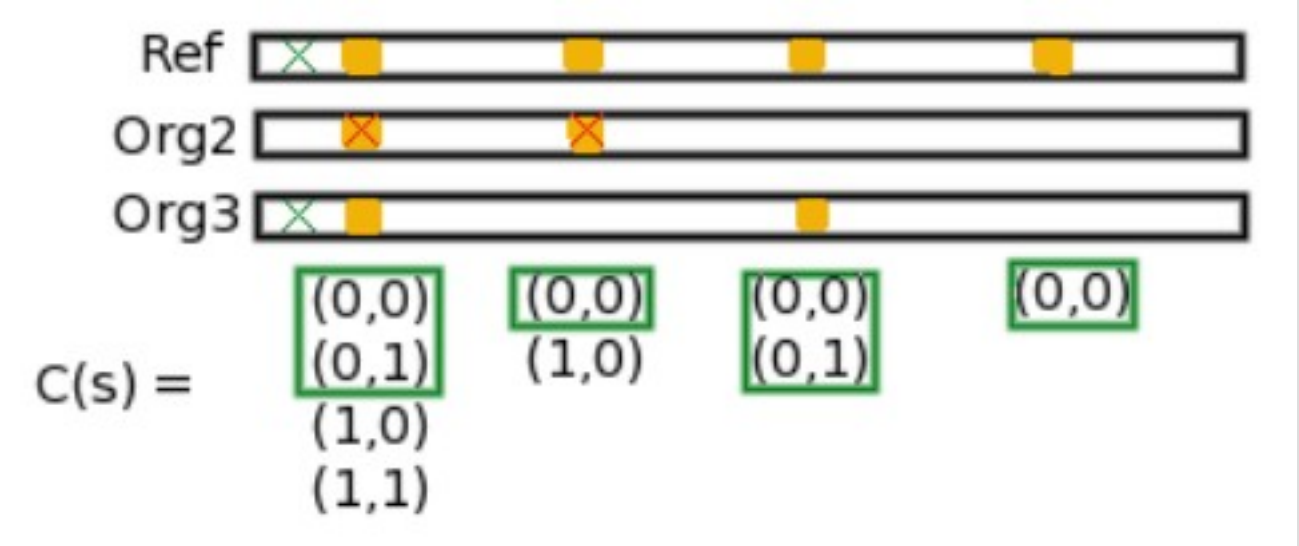
p(s) 2

Jedem möglichen Selektionsmuster s muss eine Wahrscheinlichkeit zum Auftreten zugeordnet werden, die die Konservierung funktionaler Ziele innerhalb einer Verwandten Gruppe (z.B. Saugetiere) berücksichtigt (n(c,t)).

+ Info über Konservierung des jeweiligen miRNA Gens geht auch ein

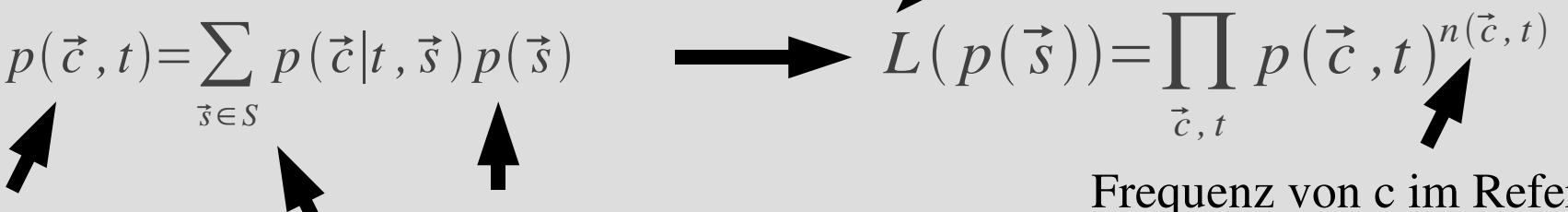
MiRNA Gen Konservierung

- Angenommen die miRNA wurde nicht in Org2 gefunden, dann setze die Wahrscheinlichkeit aller Selektionsmuster auf 0, die in Org2 eine Selektion haben, also
  p((1,0)) = 0, p((1,1)) = 0



p(s) 3

Suche Verteilung für p(s) die am besten die ermittelten Daten p(c|t,s) und n(c,t) erklärt!



Bei gegebenen Konservierungsmuster und seed Typ

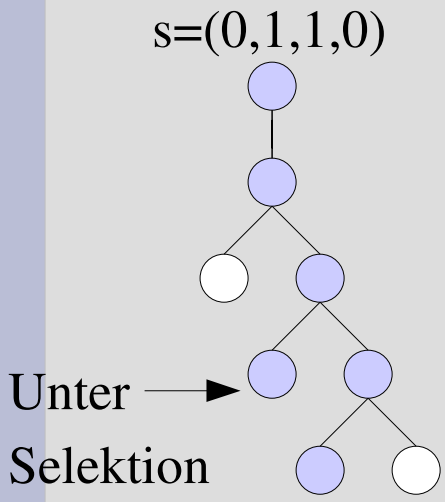
zu ermitteln

Frequenz von c im Referenz Organismus

Menge der Selektionsmuster konsistent mit miRNA Gen Konservierung, d.h. ignoriere s die Selektion zeigen wo keine miRNA existiert.

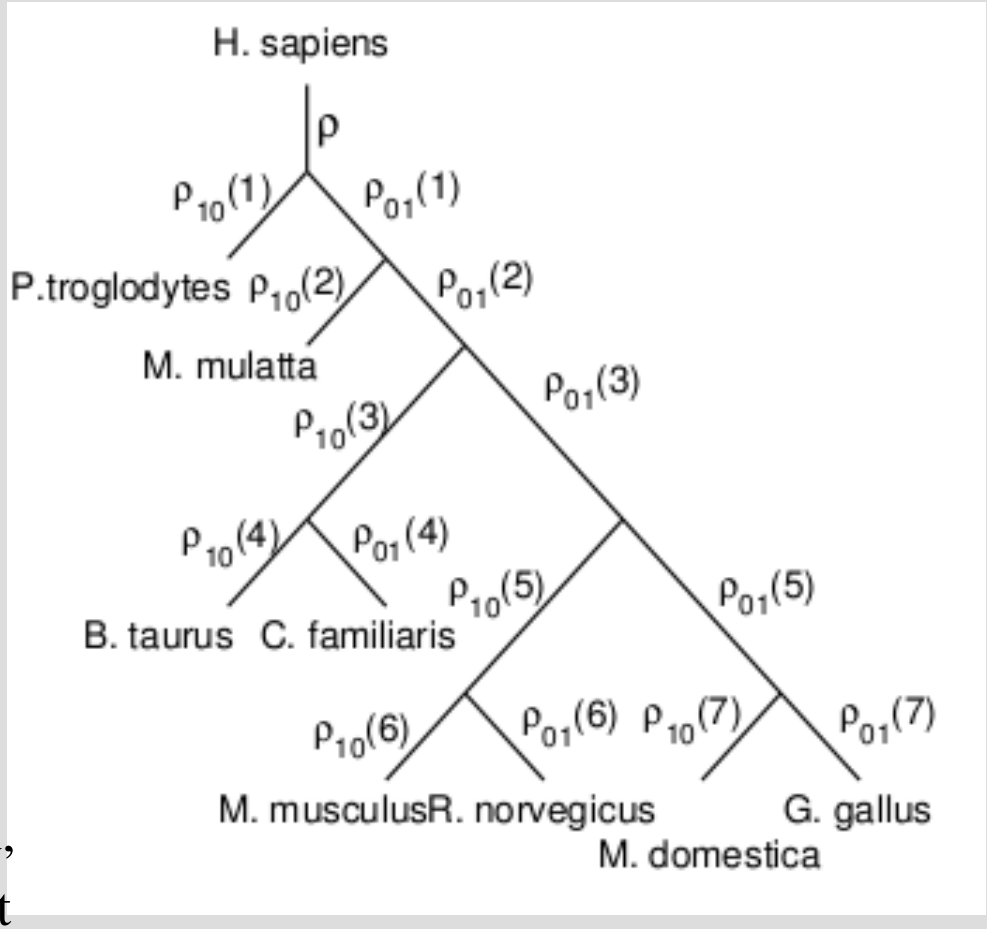
p(s) – Parametrisierung 1

Statt für jedes mögliche Selektionsmuster eine Wahrscheinlichkeit zu errechnen werden andere Parameter ausgewählt.



$$p(\vec{s}=(0,1,1,0)) = p\, p_{01}(1)\, p_{11}(2)\, p_{10}(3)$$

p: Anteil der miRNA Ziele die Funktional sind,
d.h. Unter Selektion in mind. 1 verwandten Art



p(s) – Parametrisierung 2

Grund:

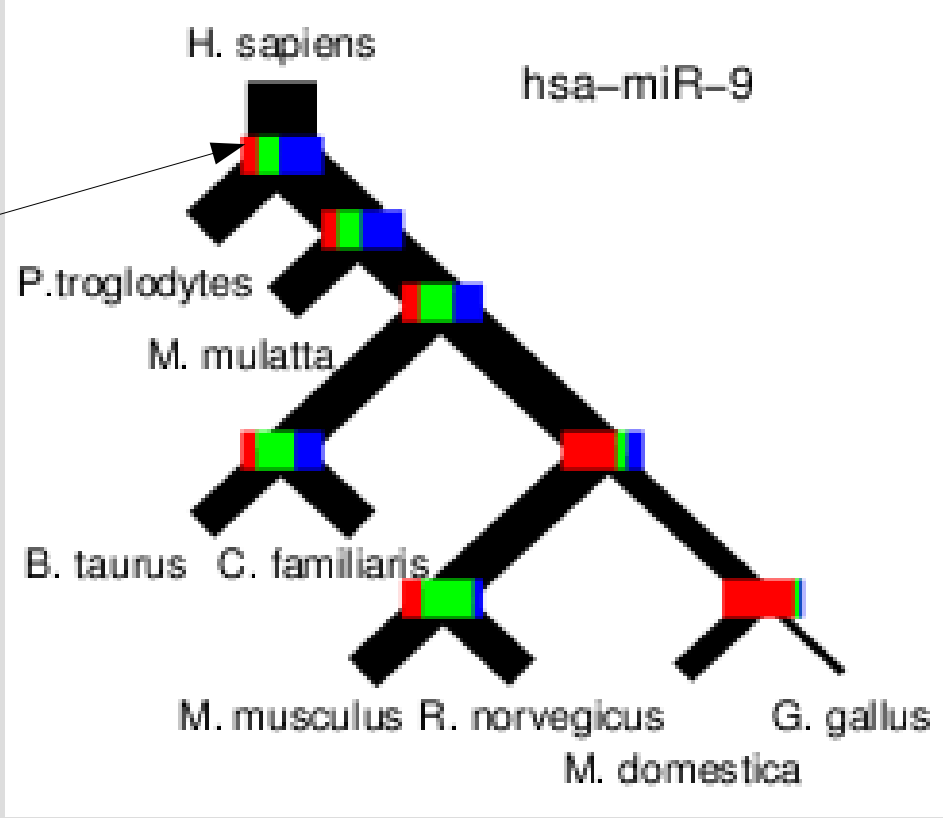
- Overfitting durch zu viele Parameter soll vermieden werden (2^g Parameter bei g Unterarten, Reduktion auf ~ g Parameter)

<u>+ Berücksichtigung von:</u>

• Verschiedenen Raten von Verlust und Gewinn von Selektion entlang des Baumes. (Selektionsdruck nicht konstant)

•Topologie des Baumes

•Asymmetrie des Baumes (da nur in der Referenz(root) neu nach Zielen gesucht und von diesem pairwise aligned wird).



p(s) – Expectation Maximum

- Maximierung durch Expectation Maximum Prozedur
- Iterationen über Parameter p_w(k) maximiert L(p(s)).

Bayes Formel 1

-> Wahrscheinlichkeit für die Funktionalität des miRNA Ziels

Symbol: leeres Selektionsmuster s=(0,...,0)

Hintergrund Modell für c (Zufall von c)

A priori für keine Selektion

$$p(\vec{s} \neq \vec{0} | t, \vec{c}) = 1 - \frac{p(\vec{c}|t, bg)(1-p)}{\sum_{\vec{s} \in S} p(\vec{c}|t, \vec{s})\, p(\vec{s})}$$

Alle Selektionsmuster konsistent mit miRNA Gen Konservierung

Schliesse Nicht-Selektionsmuster aus!

Bayes Formel 2

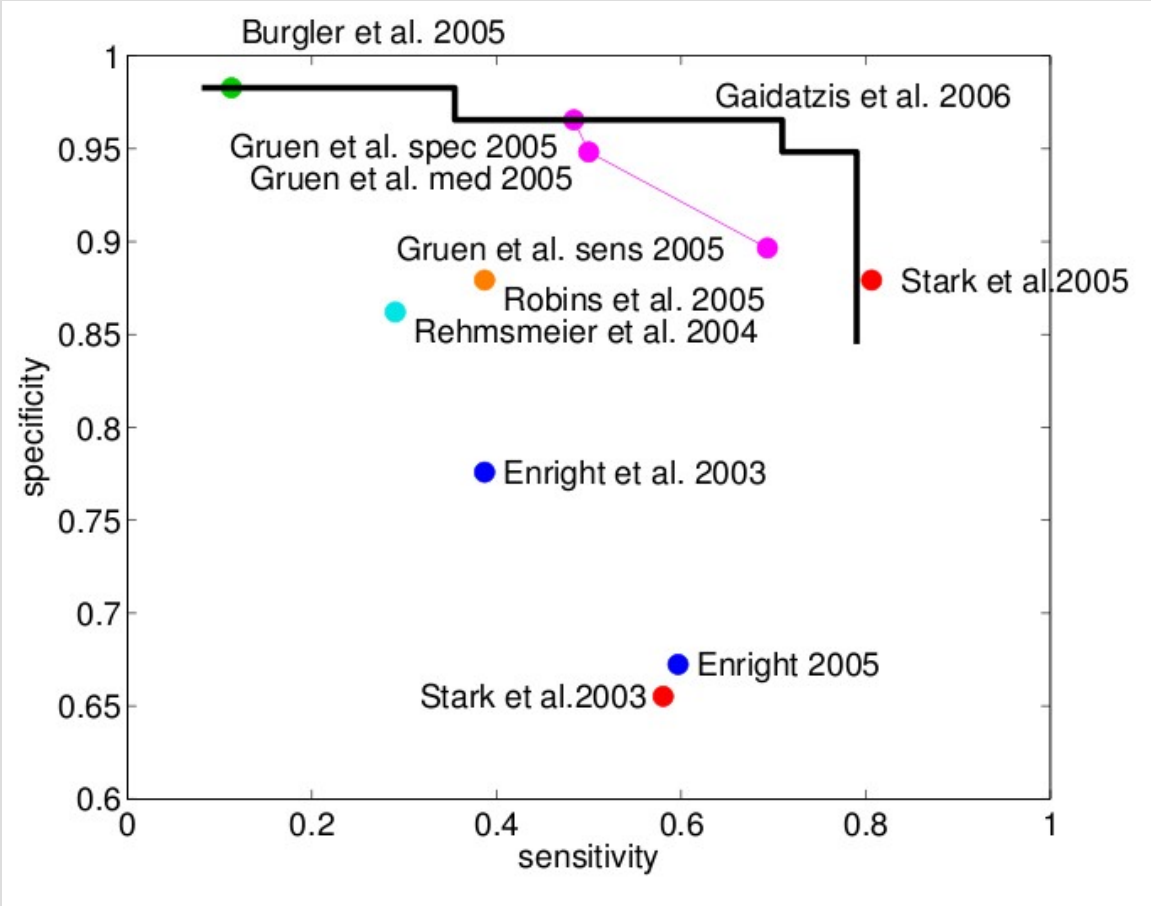
- Ein miRNA Ziel im Referenz Organismus kann als relevant für eine Interaktion angesehen werden, wenn die posterior Wahrscheinlichkeit eine Grenze (level of confidence) überschreitet.
- Hier meist >= 0.5

Resultat: Performance

•Erstelle mehrere Mengen mit unterschiedlichen cut-offs für die posterior Wahrscheinlichkeitenen : Eine mRNA ist Ziel einer miRNA wenn eine Zielsequenz enthalten ist, die den posterior grösser als den cut-off hat.

Sensitivity:
TP/(TP+FN)

Spezifität:
TN/(FP+TN)



Test an 120 experimentell getesteten miRNA-RNA Interaktionen bei den Fliegen:
Restlicher Datenvergleich von Stark et al., 2005

Vergleich zu Stark et al. Und Grün et. al 1

- Nehme miRNA Ziele mit einer posterior Wahrscheinlichkeit >= 0.5
- Umgang mit Überlappungen bei Spleissvarianten: Immer wenn eine andere Methode ein miRNA Ziel in einer Splice-Variante vorhersagt, teilen sich alle anderen Varianten dieses Ziel.

Vergleich zu Stark et al. und Grün et. al 2

Ergebnis:

Überlappung zwischen dieser Methode und der von Stark, Grün variiert stark zwischen miRNAs, z.B.:

Bantam (Zuckermais) miRNAs:	Stark(76%)	Grün(86%)
mir-1:	Stark(70%)	Grün(75%)
mir-281:	Stark(38%)	Grün(50%)

-> starke Abweichungen, mit weniger als 50 % Identität

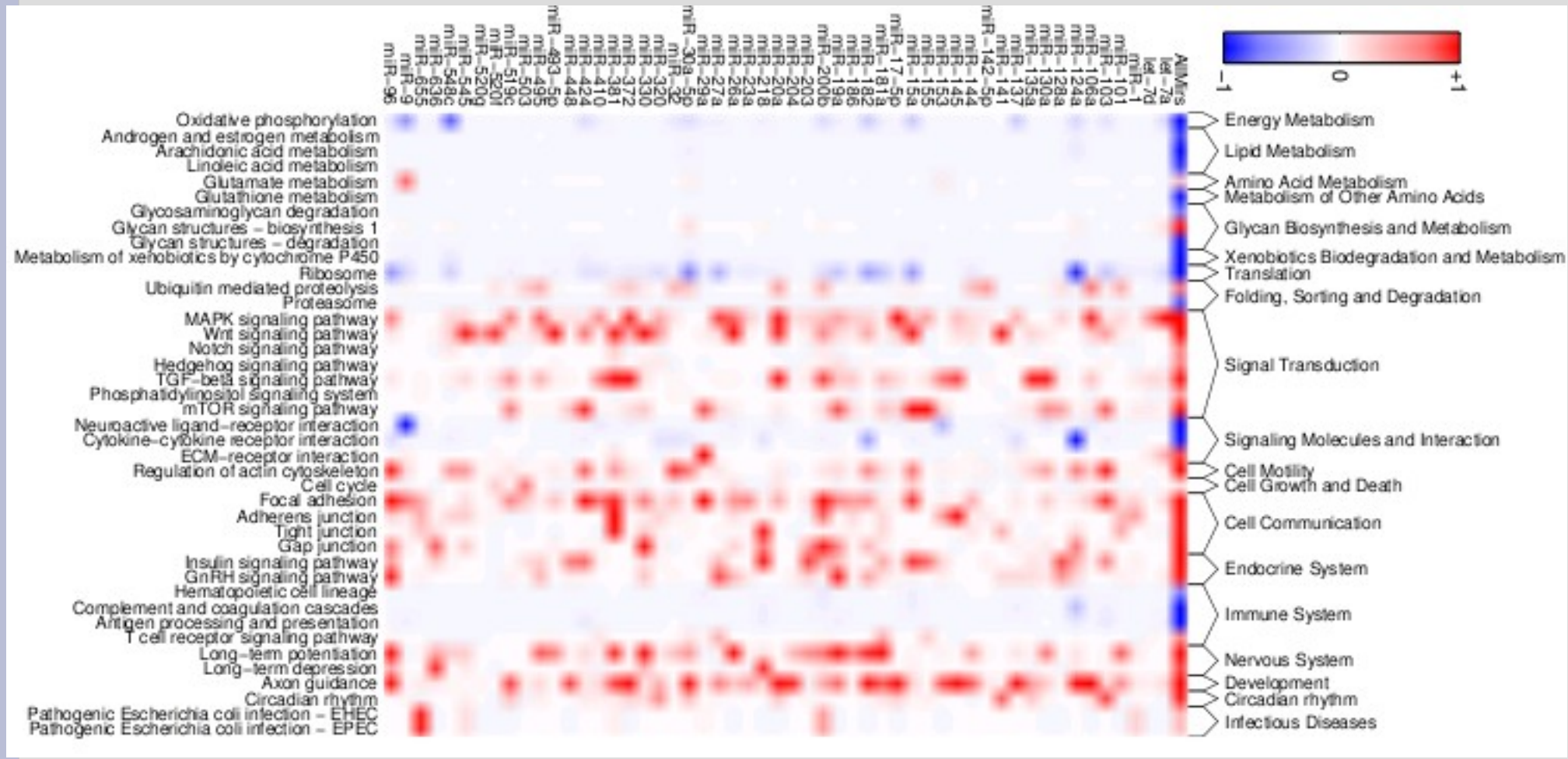
Fazit: zu wenig experimentelle miRNA-mRNA Interaktionsdaten um aussagekräftigen Vergleich zu führen.

Ableiten der miRNA Funktion: Pathway Analysis (via KEGG) 1

- Vorbereitung:
  - mRNAs mit posterior Wahrscheinlichkeit >= 0.5
  - 4011 humane Refseqs (NCBI)
  - 2 Modelle (log-likelihood ratio darüber):
    - Unabhängiges Modell: Wahrscheinlichkeit der Interaktion einer mRNA mit einer miRNA ist unabhängig von der Zuordnung der mRNA in einen Pathway
    - Abhängig: Die Zuordnung der mRNA spielt eine Rolle, bei der Interaktion
  - positiver log-likelihood Ratio => Überrepräsentiert
  - negativer => Unterrepräsentiert

Ableiten der miRNA Funktion: Pathway Analysis (via KEGG) 2

Ausschnitt mit aktivsten Pathways



Ableiten der miRNA Funktion: Pathway Analysis (via KEGG) 3

- Wie schon von Stark et.al,2005 beobachtet sind weit verbreitete Gene, die grundlegene Metabolische Funktionen innehalten selten Ziele von miRNAs
- Ziele sind v.a. Gene der Transcriptions Regulation, Interzellulären Kommunikation, Zell Wachstum, Tod und Entwicklung
- Bsp. 19 von 119 miRNA sind signifikant bei der Axon Führung (axon guidance pathway) involviert.

Pathway Analysis (via KEGG) 4 - Fazit

- Methode vermittelt erwartete Assoziationen zwischen miRNA – Pathways.

- Bsp.
  - Pathway Analyse zeigt für mir-124a Interaktionspartner die bei Entwicklung des Nervensystems eine Rolle spielen. Für mir-124a ist bekannt, dass dessen Expression spezifisch für Nervensystem ist.

Fazit 1

- Methode erreicht gute Resultate (Spezifität,Sensivität) im Vergleich zu Anderen
- Aber: Qualitativer Vergleich schwer für grosse Daten, da experimentelle Tests fehlen.

Fazit 2

- Vorteile der Methode:
  - Kann auf jede Familie von Organismen angewandt werden
  - phylogenetische Beziehung wird automatisch im Modell integriert (via Konservierungsmuster)
  - einfache Erweiterung, je mehr Genome zur Verfügung stehen
  - Bayes Methode unabhängig von der Definition der Ziel Region (kann erweitert, abgeändert werden)

Fazit 3

- miRNA spezifische Modellierung der Evolution des Selektionsdruck (p(s)) ermöglicht den differenzierteren Umgang mit miRNAs, die in unterschiedlichen Evolutionsstufen ausgeprägt sind.
- Dadurch bessere Nachbildung der Einwirkung von Selektionsdruck in verschiedenen Stadien für entsprechende Ziele.

Referenz

[1] Gaidatzis D, van Nimwegen E, Hausser J, Zavolan M: Inference of miRNA targets using evolutionary conservation and pathway analysis. BMC Bioinformatics 2007, 8:69.