

Nachlese zu den Referaten

Referate

Die Themen

Phylogenie

Aminosäurematrizen für die Eukaryontenphylogenie

Ansätze für die Prokaryontenphylogenie

Vergleichende Sequenzierung

Spezieswahl zur Informationsoptimierung

Genomevolution

Die Entstehung des *S. cerevisiae* Genoms

Induzierte Duplikationen in *S. cerevisiae*

Metagenomics

Vergleich ökologischer Nischen/Stoffwechselleistungen

Transkriptomics

Transkription

Die Bestandteile

Polymerasen

exprimieren das Gen

Promotoren

enthalten die Information für die
geregelt Expression eines Genes

Transkriptionsfaktoren

modulieren die Genaktivität

Chromatinstrukturen und
epigenetische Faktoren

können Gene an- oder abschalten

(DNA und Histon- Methylierung,
Histon- Acetylierung)

Transkription

Polymerasen

:

RNA Polymerase besteht aus 5 Untereinheiten in Bakterien

Untereinheit	Funktion
α_2	die Untereinheiten assemblieren das Enzym und erkennen regulatorische Faktoren
β	Polymeraseaktivität
β'	Bindung an DNA
ω	Chaperonfunktion?

Drei Typen von RNA-Polymerasen in Eukaryonten

Typ	Funktion
I	rRNA Synthese
II	mRNA und snRNA Synthese
III	tRNAs, 5S rRNA, small RNA Synthese

Transkription

Promotoren

Proteinbindestellen auf der DNA, die die Expression eines Genes regulieren

Polymerase-Bindestelle

Transkriptionsfaktoren-Bindestellen

Enhancer

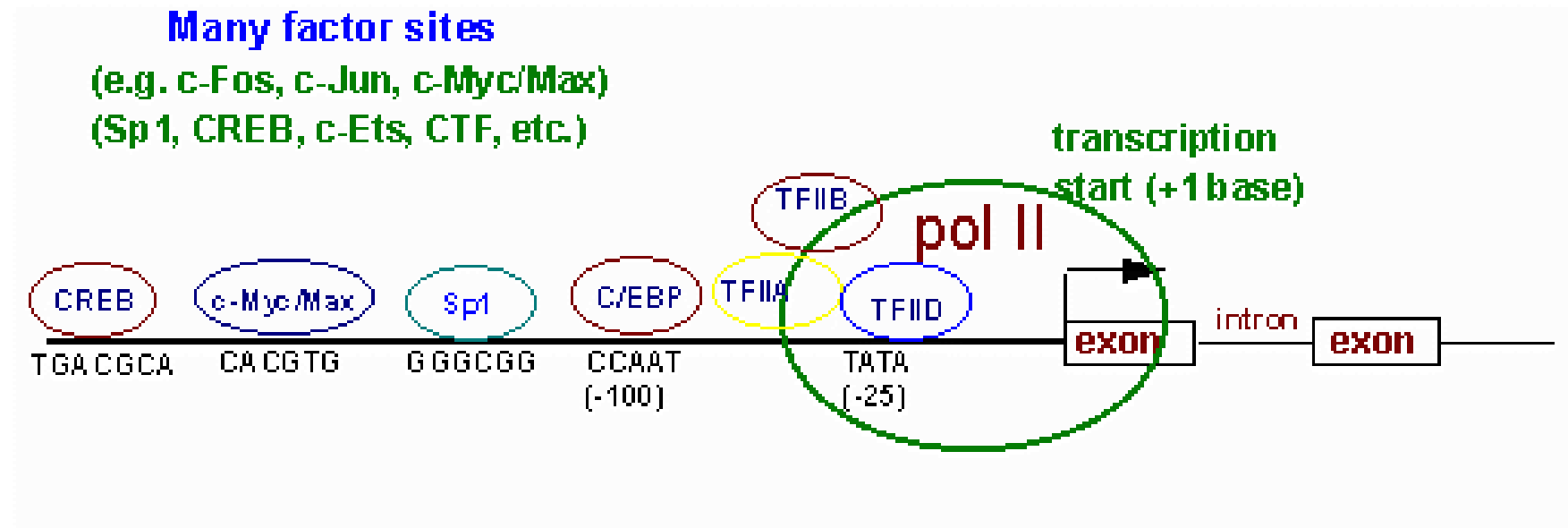
Silencer

Gewebespezifische Expressionsregulatoren

Durch epigenetische Modifikation kann das Bindungsverhalten
geändert werden

Promotoren

Definition einer Regionen

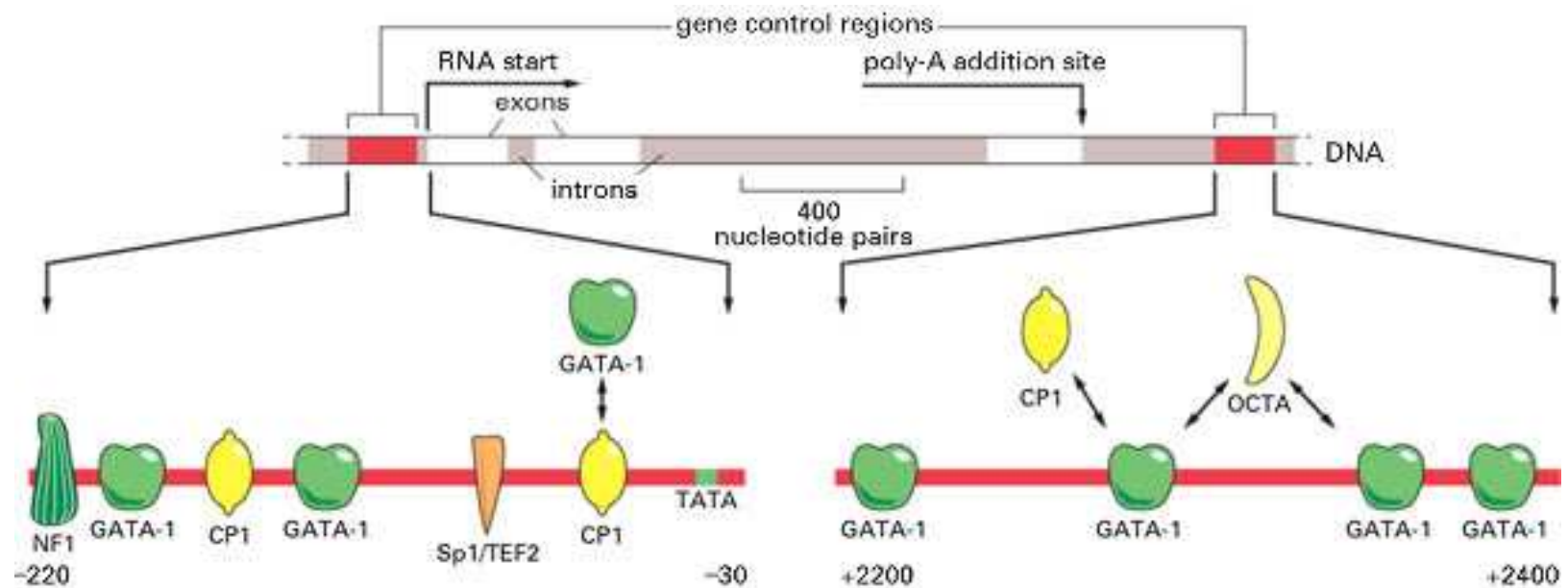


nigen Basen

Promotoren sind komplex aus mehreren Bindungsstellen zusammengesetzt!

Promotoren

Genkontrollregionen



Transkriptionsbeeinflussende Sequenzen müssen nicht nur im 5' Bereich gesucht werden!

Transkription

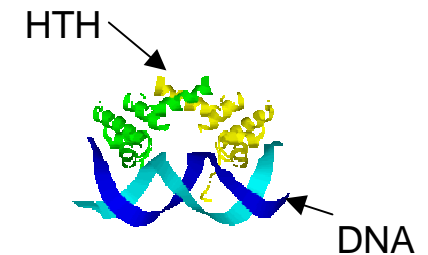
Transkriptionsfaktoren

Transkriptionsfaktoren (TF) binden an spezifische Sequenzmotive (TFBS)

Es gibt verschiedene Proteinmotive, durch die Transkriptionsfaktoren erkannt werden können, z.B.:

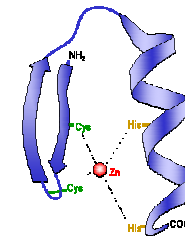
Helix-Turn-Helix

bindet die große Grube in DNA



Zinkfinger

2x β -Faltblatt + Helix



bZip

basische Region mit Leucin-Zipper

bHLH

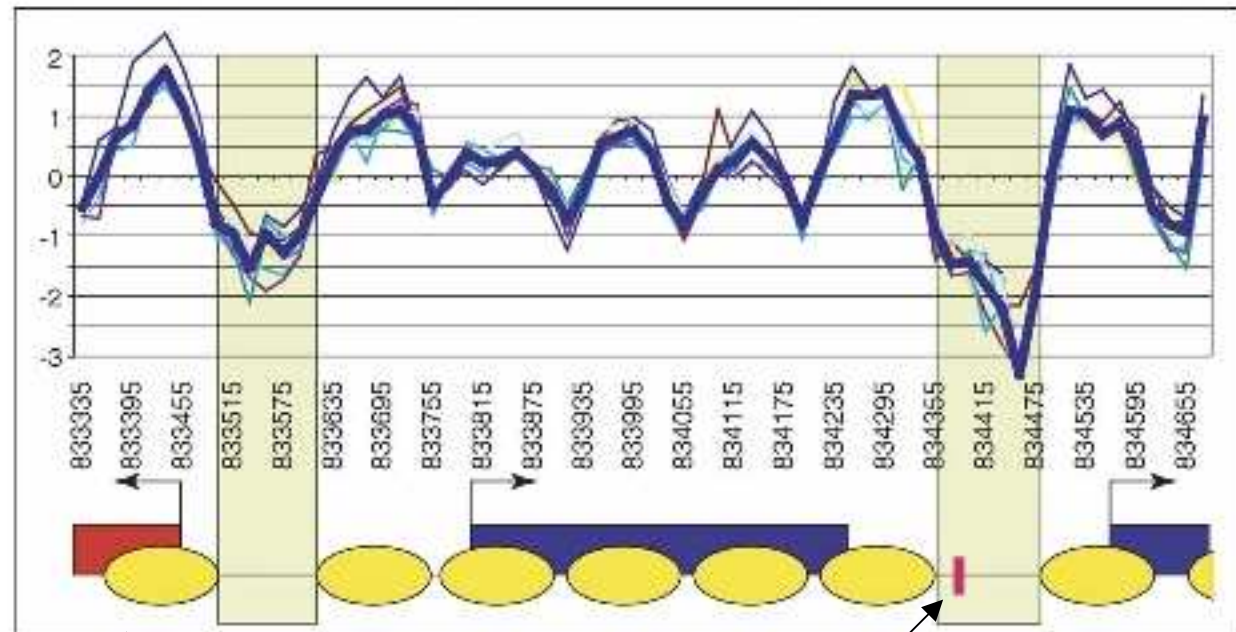
basische Region in einer Helix mit zweiter Helix über Loop verbunden

Transkription

Chromatinstruktur

Hefechromosom XVI

Microarray-
Ergebnisse zu
Nukleosomenpositionen

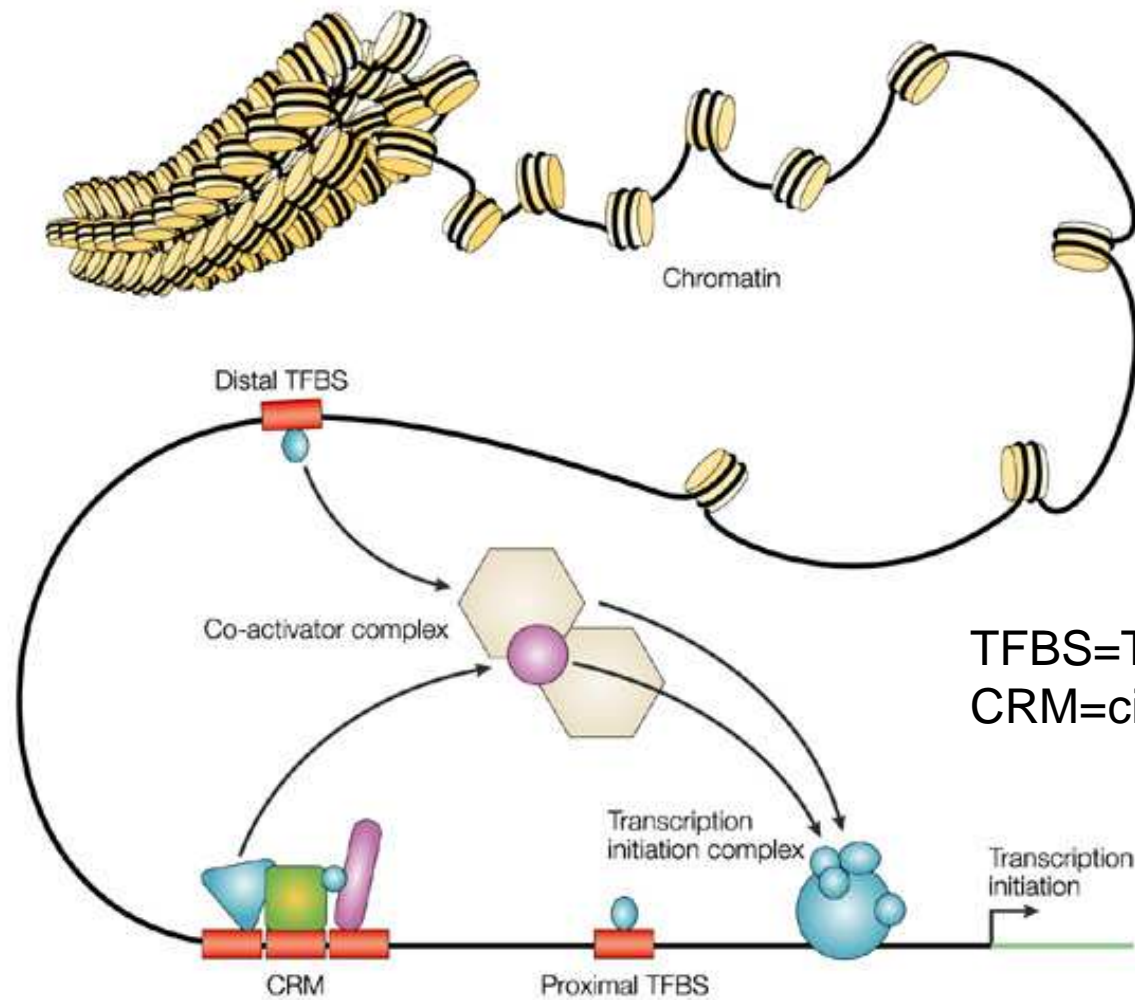


Nukleosomen

bekanntes Promotormotiv

Transkription

Die Maschinerie



TFBS=Transkriptionsfaktorbindungstelle
CRM=cis regulatorische Module

Detektion

Wie können wir Promotoren finden?

De novo aus einer Sequenz: Nur experimentell!

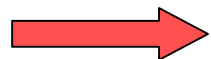
Sequenzvergleiche

**Transkriptions
-faktor**

seq A **GCCCATATCCGACCCGTTTCATCGTA**

seq B **AAGCTGCCAGAACCTGTGATATGCA**

Homolog von seq A



Hindernis: Die gesuchten Sequenzen sind kurz.
Keine allgemeinen Regeln existieren!

Detektion

Konservierung als funktioneller Marker

Theorie:

nicht translatierte Sequenzen, die Funktionen vermitteln unterliegen einem negativen Selektionsdruck

Konsequenz:

Suche nach kaum veränderten Sequenzen im 5' Bereich eines Gens in verschiedenen Organismen

Probleme:

Konservierung kann nur wenige Basen betreffen.
Welche Gene werden verglichen?

Detektion

Inter-Spezies Vergleich

Fragen:

Haben die von mir verglichenen regulatorischen Elemente irgendetwas miteinander zu tun?

Oder vergleiche ich Äpfel mit Birnen?

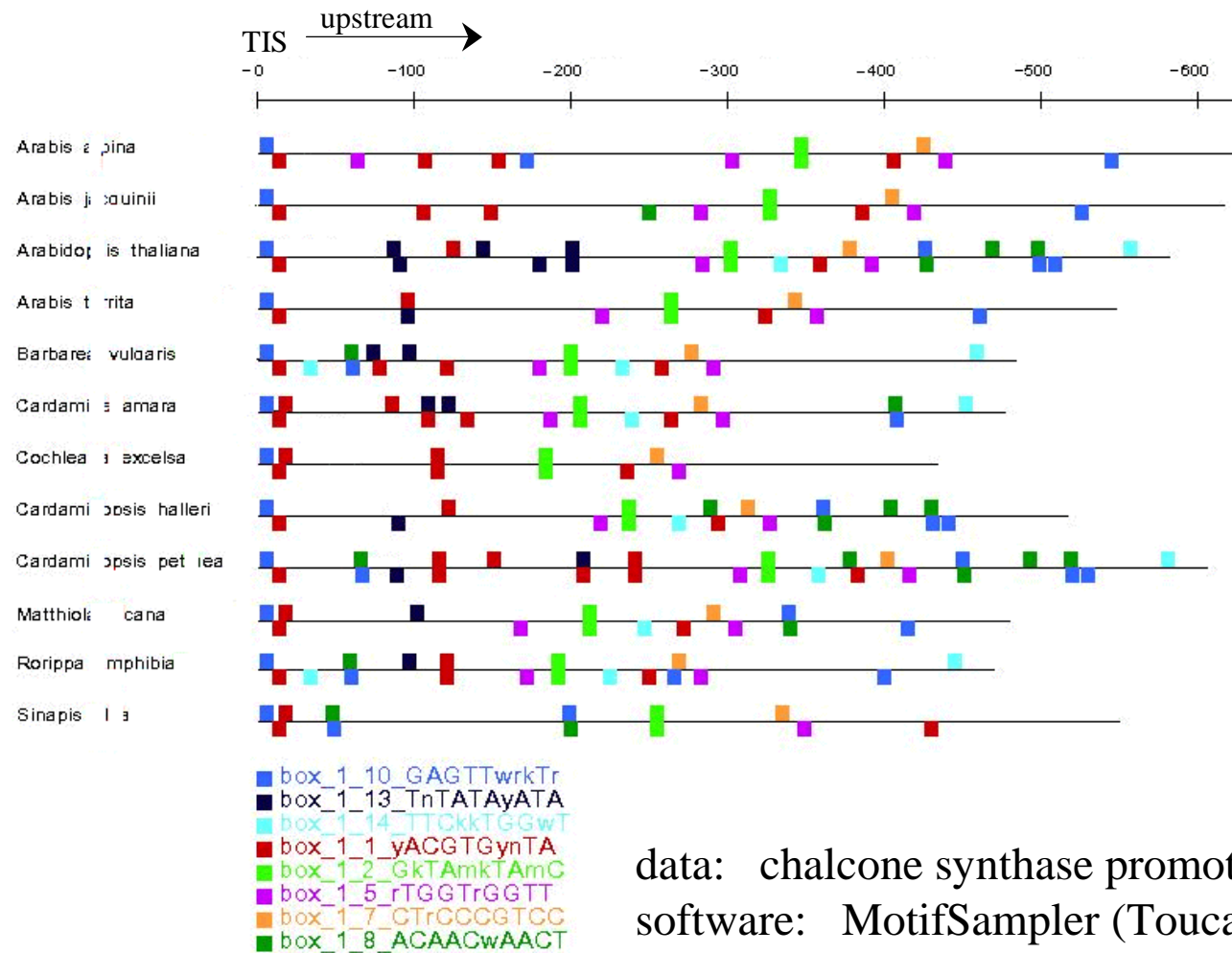
Kann ich aussagekräftige Vergleiche machen?

Wo setze ich Ähnlichkeitsschwellenwerte an?

Wie vergleiche ich, welche Methoden stehen mir zur Verfügung?

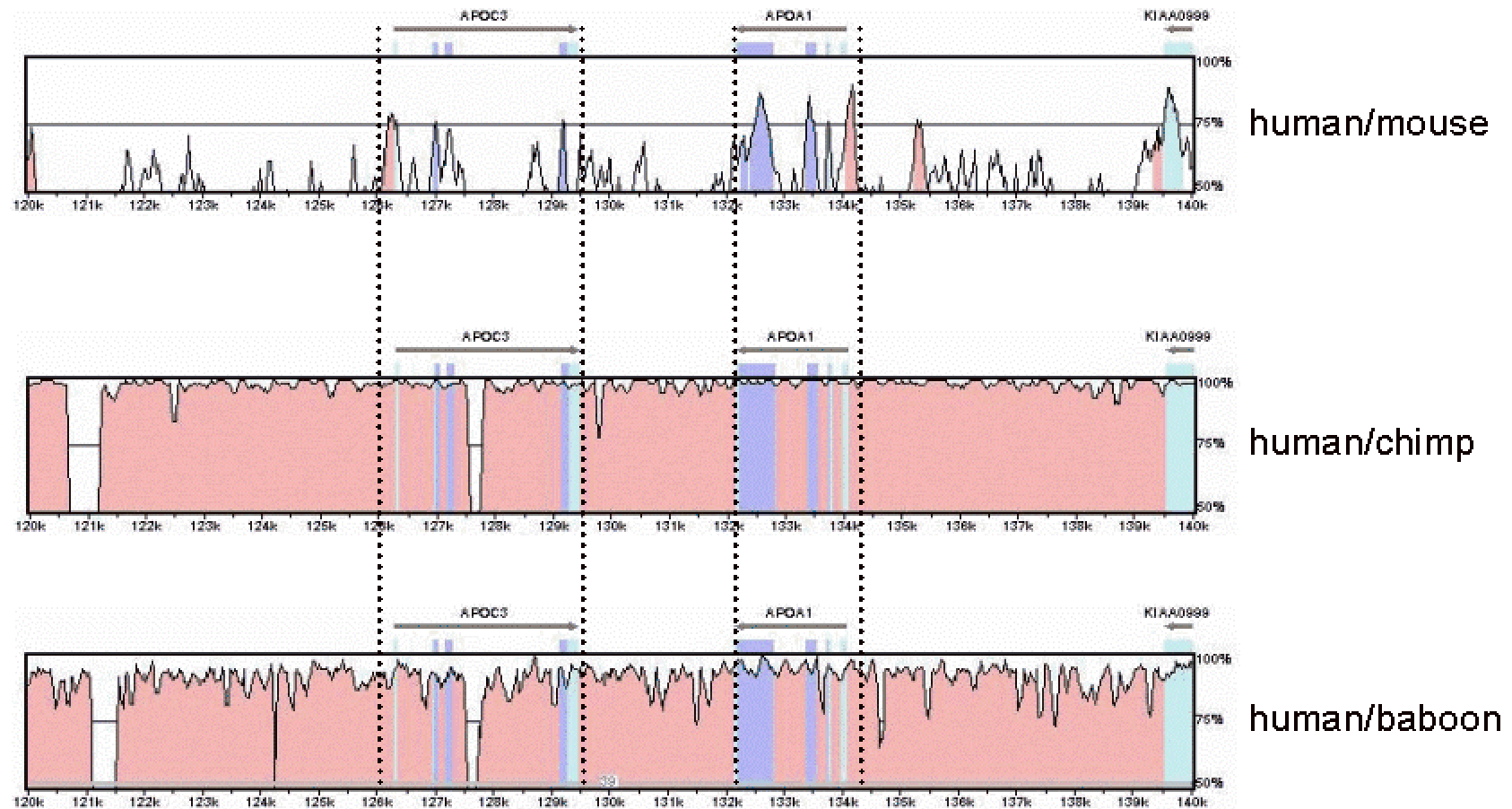
Promotoren

Inter-Spezies Vergleiche von Motiven



data: chalcone synthase promoter data set
software: MotifSampler (Toucan Suite)

Detektion Footprinting

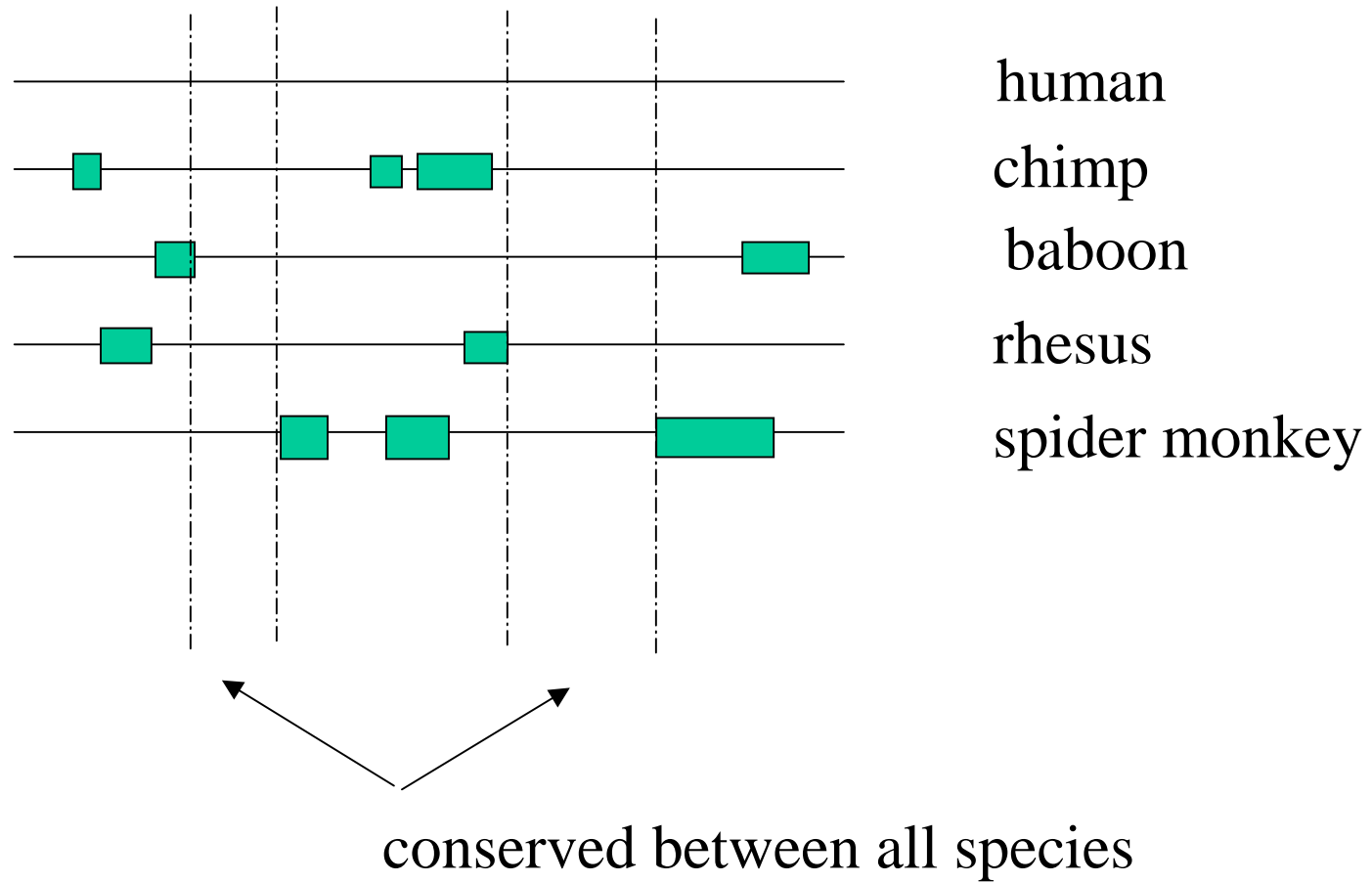


from: www.science.doe.gov

comparison between **2** sequences

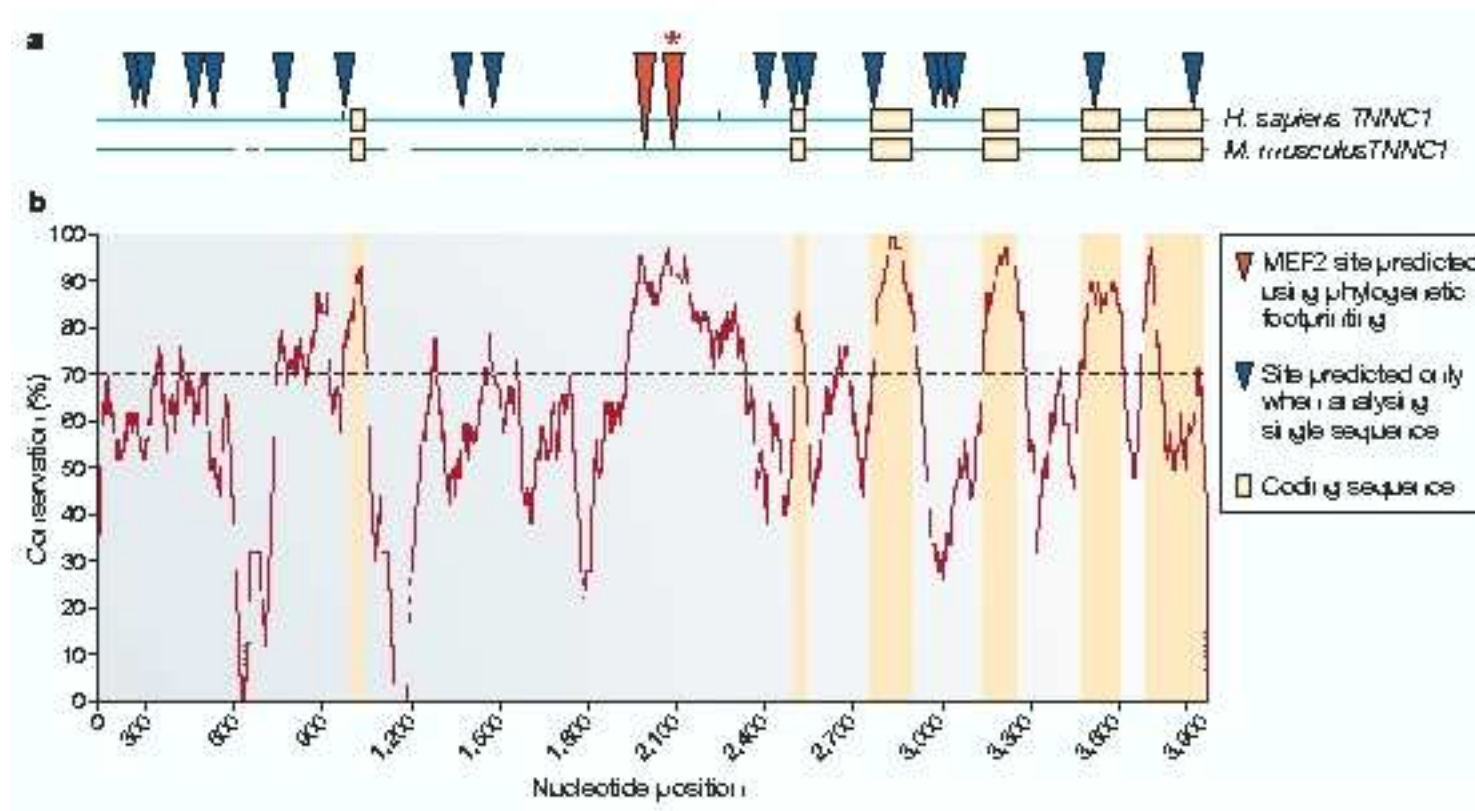
Detektion

Shadowing



Detektion

Beispiel Footprinting



from: Nature Reviews Genetics 2004, 5 276-287

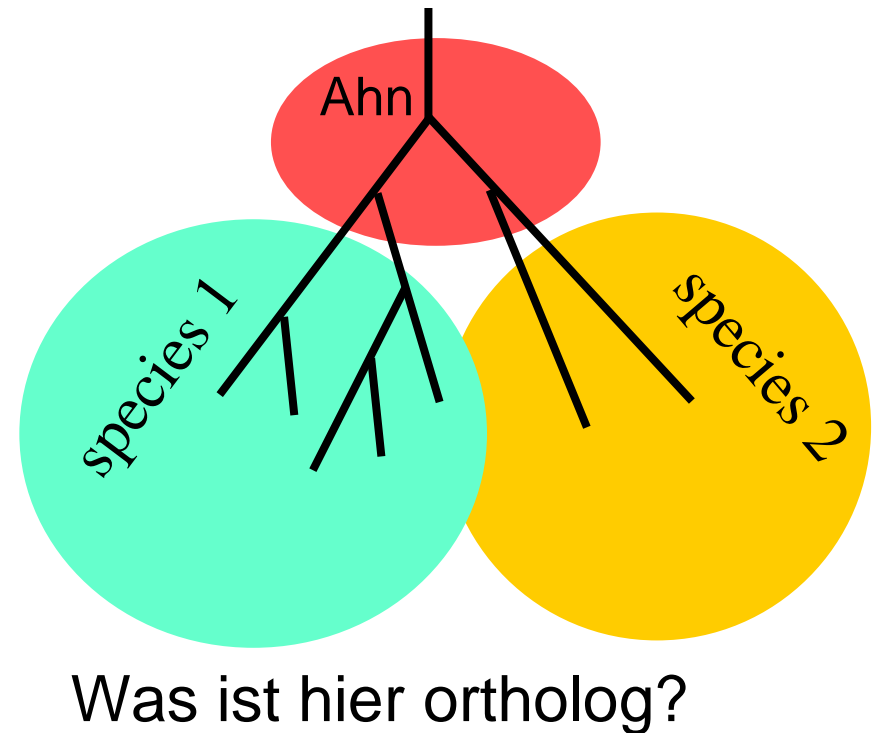
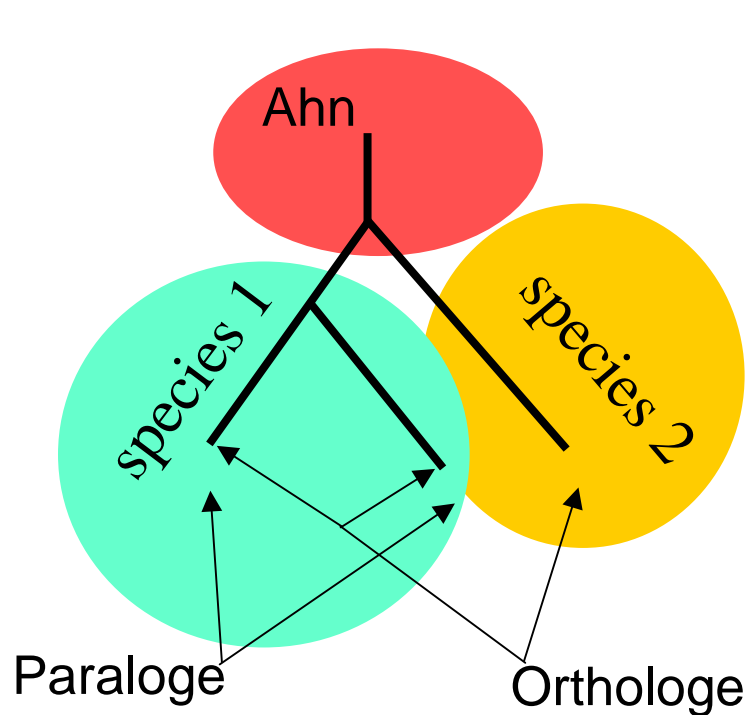
Detektion

Orthologie der Proteine ist Voraussetzung

Definition:

ortholog
paralog

Entstehung durch Speziation
Abstammung von duplizierten Genen



Detektion Alignment

```

      14703      14713      14723      14733      14743
14694: CCGGGCGGCCGCAGAAGCGCCAGGCCCGCGGCCACCCCTCTGGCGCCA
546:  .AAA.A...T....G.T.G.A..AT...A.....GA.A..TG
952:  .A.....TG..GG.....G.T.....T.G
1897: .A.A...C.....G...TT..GG.....TT.C...C.-----
896:  .G.....GACG...C.-.A.GC..G.A.A.A.G.C...TC.GC.G..GG

```

hsFOS
ssFOS
maFOS
mmFOS
ggFOS

CpG
island
mRNA

```

      14753      14761      14770      14778
14744: CCGTGGTTGAGCCC--GTGACGT--TTACACTCATTC--TAAAACGCTT-
596:  .AA...A.C..T.C-.....A..GGG.....T...G-
998:  .ACA...CC...A-.....G.A.TGT...-C.-G.....
1938: TTCCA...CC...A-.....AGG.AGTC...C..TTC.C.G.....
945:  G.....GCGT...A-.....AGGCG.....CGCGG..GCAG

```

hsFOS
ssFOS
maFOS
mmFOS
ggFOS

c-Fos promoter
data set

```

                                FOS:1
                                →
      14797      14807      14817      14827      14837
14788: G TATAAA AGCAGTGGCTGCGGCGCCTCGTACTCC AACC GCATCTGCAGC
642:  .....A.....A.....
1043: -C.....G...CC...A...AC.....GA.....
1986: -C.....G..GCCA...A.....AC.....GA.....
989:  .G.....G.GG.C...-.C..AGCGA.....G.G...GAG.....

```

hsFOS
ssFOS
maFOS
mmFOS
ggFOS

TATA box

transcription start
(homology-based assignment)

Detektion

Alignment - Die Wahl der Parameter

orthologes achttes Intron des Gens CG9935-RA
in *D. melanogaster* und *D. pseudoobscura*

M=5, X=-5, S=-5

```
mel GTAAGTTTGTAT-ATTTTTTTTTTTTGAAGTGA-CAAATAGC-A-CTTATAAATATACTTAG
pse GTTCGTTAACACATGAAATTCATCGCCTGAT-TGTTCA-CTATCTAACTAACGAAT-T--TTAG
** *** ** * ** * *** ** ** * * * * * ** * **
```

33 Positive

M=5, X=-6, S=-4

```
mel GTAAGTT-----TGTTTATTTTTTTT--T--TT-TTGAAGTGA-CAAATAGCACTTATA--A
pse GTTCGTTAACACATG-A-A-ATTCATCGCCTGATTTGTT-CACT-ATC---TA--AC-TA-ACGA
** *** ** * *** * * ** ** * * * * ** ** * * *
```

36 Positive

```
mel ATATRCTTAG
```

```
pse AT-T--TTAG
```

```
** * ****
```

M=Match
X=Mismatch
S=Strafe für Lücke

Detektion

IUPAC Beschreibung von Signalen

a	a; adenine
c	c; cytosine
g	g; guanine
t	t; thymine in DNA;
uracil	in RNA
m	a or c
r	a or g
w	a or t
s	c or g
y	c or t
k	g or t
v	a or c or g; not t
h	a or c or t; not g
d	a or g or t; not c
b	c or g or t; not a
n	a or c or g or t

Beispielmotiv mit IUPAC wobble

NNGTAWBTSRWM

IUPAC=International Union of Pure and Applied Chemistry

Detektion

Motive

Das Motiv:

ATN₁₋₃G₂CGTN_xTGA_{4,5}

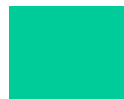
kann in viele mögliche Sequenzen übersetzt werden:

ATGGGCGTAGAGGAGACTTTATGAAAAA

ATCGGCGTTGAAAA

ATAAGGCGTGAGTGAAAAA

etc.



=stabile Motivanteile

Detektion

Matrizen

Bindemotive

kürzer als 20 Basen
degeneriert, d.h. nur Häufigkeiten von Basen an bestimmten
Motivpositionen können angegeben werden

Herstellung von Matrizen

finde Gene mit ähnlichem Transkriptionsprofil
korreliere Profil mit Funktion
bilde ein Alignment mit den potentiell funktionsgleichen
Promotormotiven
berechne die Häufigkeit einer Base an den Alignmentpositionen

Nutzung der Matrize

Scannen der Sequenz mithilfe der Matrize (Stepsize=1)

Detektion

„Positional Weight Matrix“ Berechnung

- Erstellung einer ‚position frequency matrix‘ PFM
Normalisierung ergibt Wahrscheinlichkeiten für eine Base an einer Position
- normalisierte PFM
- Berechnung der Wahrscheinlichkeit einer Sequenz,
ein Profil widerzuspiegeln
Das Produkt der Wahrscheinlichkeit jeder Base an der entsprechenden Position
- Für die Computeranalyse:
Umrechnung in Log Werte mit Korrekturfaktor
(Korrekturfaktor sehr unterschiedlich, z.B. Quadratwurzel der Anzahl an Positionen)
+Korrektur für die Nukleotidfrequenzen
=PWM
‚Score values‘: Summierung aller PWM Werte über die
Länge des Profils

Detektion

Positional Weight Matrix (PWM)

```
# scale = ln (frequency)
# frequency counts are based on
N(seq) = 21
```

#	A	C	G	T
1	1.50	1.25	2.25	1.50
2	1.70	0.92	2.01	1.87
3	2.01	0.92	1.50	2.01
4	-0.69	-0.69	-0.69	2.80
5	2.35	-0.69	1.87	-0.69
6	0.41	0.41	2.67	-0.69
7	-0.69	2.44	-0.69	1.70
8	0.41	0.41	1.50	2.35
9	0.41	2.25	0.41	1.70
10	2.80	-0.69	-0.69	-0.69
11	0.41	-0.69	2.67	0.41
12	0.92	0.41	-0.69	2.60
13	0.41	2.01	-0.69	2.14
14	-0.69	-0.69	2.80	-0.69
15	-0.69	-0.69	2.80	-0.69
16	0.92	0.92	-0.69	2.53
17	2.53	0.92	0.41	1.70
18	1.87	0.92	2.25	1.25
19	2.35	0.92	1.25	1.70

The smaller the differences in base values
the lower is the information content

The larger the differences in base values
the higher is the information content

Produced from an alignment of
21 sequences of A-Box motifs
in *D. discoideum*

Detektion

PWM Graphics

A-Box Motif aus tRNAs in *D. discoideum*



Detektion

Falsch positive Vorhersagen

Humangenom

Programme sagen Bindungsstellen alle
500 bis 5000 Basenpaare vorher

Beispiel: myoD (musklespezifischer Transkriptionsfaktor)

1 Bindungsstelle pro 500 Basen

vorhergesagt 10^6 Bindungsstellen

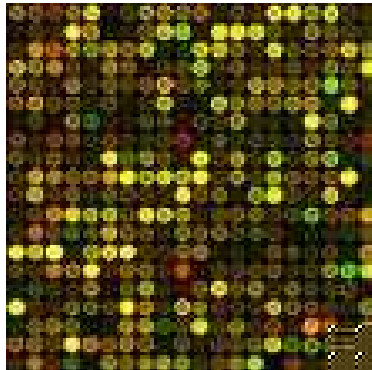
wahrscheinlich: 10^3 Bindungsstellen

1000 mal mehr falsche Vorhersagen als richtige!

Experimentelles

Transkriptomanalyse

Prinzip:



zwei unterschiedliche Zellzustände werden untersucht

Isolierung der mRNA

differenzielle Markierung mit Fluoreszenzfarbstoffen

Mischung in äquimolaren Mengen

Hybridisierung gegen ein Mikroarray, das das gesamte Genom repräsentiert

Ein Mikroarray kann mit DNA aus genomischen Sequenzen, cDNA-Banken und spezifisch hergestellten Oligonukleotiden bestückt sein.

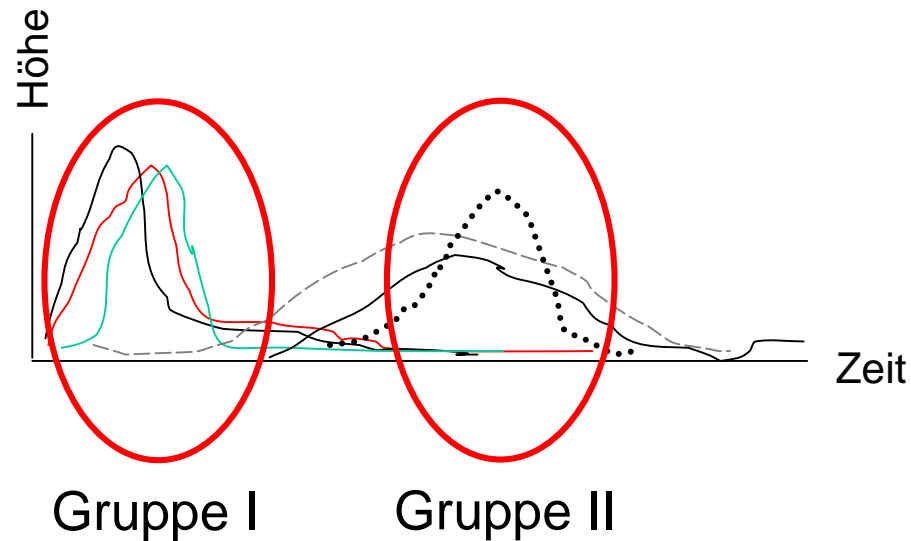
Entsprechend der Bestückung ergeben sich unterschiedliche Fehlermöglichkeiten!

Experimentelles

Promotordefinition durch Transkriptomanalyse

Innerhalb eines Organismus werden viele Gene ähnlich transkribiert:

z.B. zeitabhängiges
Profil



Diese ähnliche Transkription **kann** durch ähnliche Promotoren bedingt sein.
Ein Alignment kann hier zu der Konstruktion eines Motivs führen

Experimentelles

Transkriptomanalyse - Fehlerquellen

genomische DNA

enthält Intronen, Hybridisierungsstärke nicht kontrollierbar

cDNA

meist nur 5' Enden der Transkripte bekannt, unvollständige Transkripte wegen unvollständiger reverser Transkription
nicht alle Gene werden erfasst, da cDNA Banken unvollständig sind
alternatives Splicen führt zu Informationsverlust

Oligonukleotide

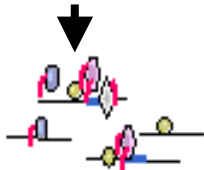
Sequenzierungsfehler der Ursprungs-DNA führen zu fehlerhaften Oligos. Teilweise differentielle Signale bei 2 verschiedenen gegen das selbe Gen gerichteten Oligos

Experimentelles

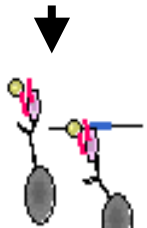
ChIP



Fixierung von Protein an DNA durch Formaldehyd in lebenden Zellen



Zell-Lyse und Aufbrechen der DNA in kurze Stücke durch Ultraschall

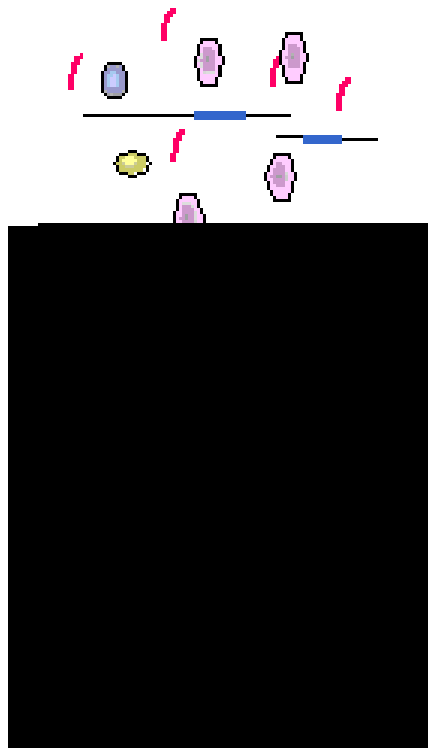


Immunopräzipitation mit Magnetpartikeln, die mit den Transkriptionsfaktor-spezifischen beladen sind

kritischster Schritt! Die Antikörper müssen spezifisch sein!

Experimentelles

ChIP Detektion



Aufhebung der Vernetzung von DNA und Protein, DNA Isolierung

quantitative PCR gegen spezifische und Kontrollregionen

Detektion und Quantifizierung gegenüber der Kontrolle
mit Hilfe von Agarosegelen

input DNA

immunpräzipitierte DNA

Durch ChIP wird DNA angereichert, die mit bestimmten Proteinen assoziiert ist

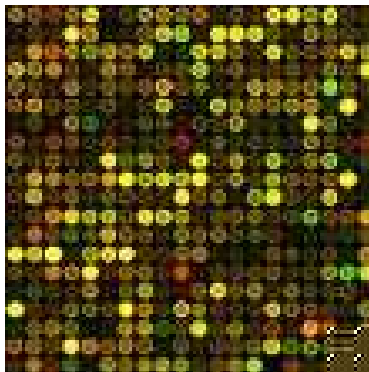
Experimentelles

ChIP-CHIP

Detektion der Immunpräzipitierten DNA kann auch global erfolgen.

Hier wird statt PCR ein Micro-Array eingesetzt.

Dieses enthält z.B. die Sequenzen von allen Promotoren eines Organismus



Die Quantifizierung erfolgt auch hier im
Verhältnis zu nicht angereicherter, markierter DNA

Experimentelles

Epistasis

Was braucht man zu einer Epistasis-Analyse?

- Eine Kollektion von Mutanten
- Alle möglichen Kombinationen von Doppelmutanten
- Eine Phänotypbeschreibung des Wildtyps sowie der Mutanten

Ist ein Transkriptionsprofil geeignet als Phänotypbeschreibung?

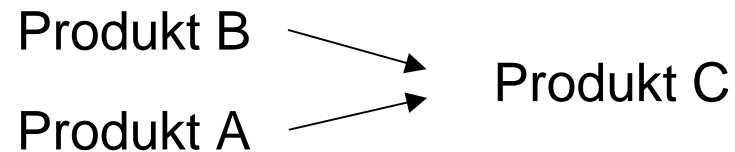
Experimentelles

Epistasis Definition

Epistasis beschreibt die Abfolge von Genprodukten voneinander in einer Wirkungskette:

Produkt A \longrightarrow Produkt B \longrightarrow Produkt C

A ist epistatisch zu B ist epistatisch zu C



Epistasis von B zu C und A zu C aber nicht von A zu B!

Symbole \longrightarrow positive Regulation
 —| negative Regulation

Experimentelles

Epistasis-Analyse

Um eine Wirkungskette mit ihren Abhängigkeiten bestimmen zu können, werden Mutanten der in der Kette wirkenden Genprodukte benutzt

Produkt A \longrightarrow Produkt B \longrightarrow Produkt C

wenn Phänotyp der Knockout Mutante A \neq Phänotyp der Knockout Mutante B und die Doppelmutante gleich B, dann ist B epistatisch zu A

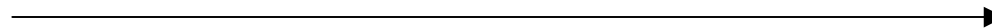
Produkt B \searrow
Produkt A \nearrow Produkt C

die Doppelmutante ist ein neuer Phänotyp > keine epistatische Beziehung zwischen A und B

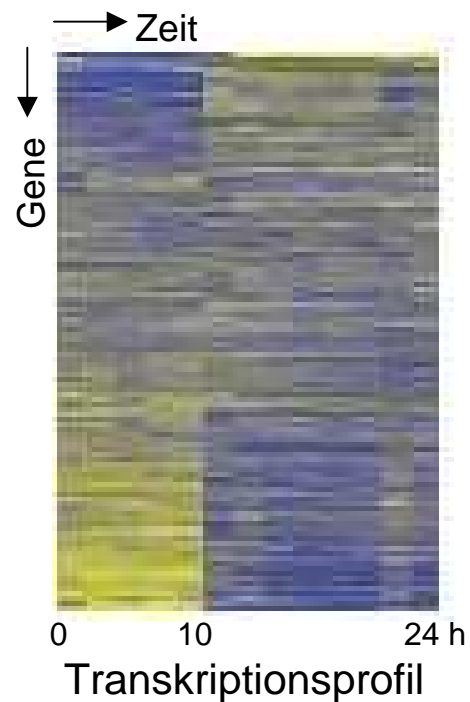
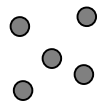
Experimentelles

Epistasis - Transkriptionsprofil

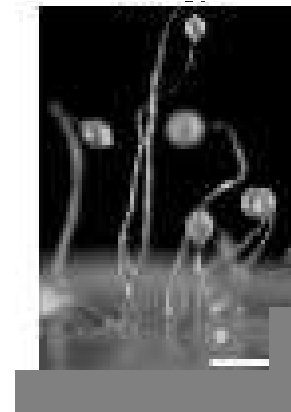
Entwicklung



Einzeller

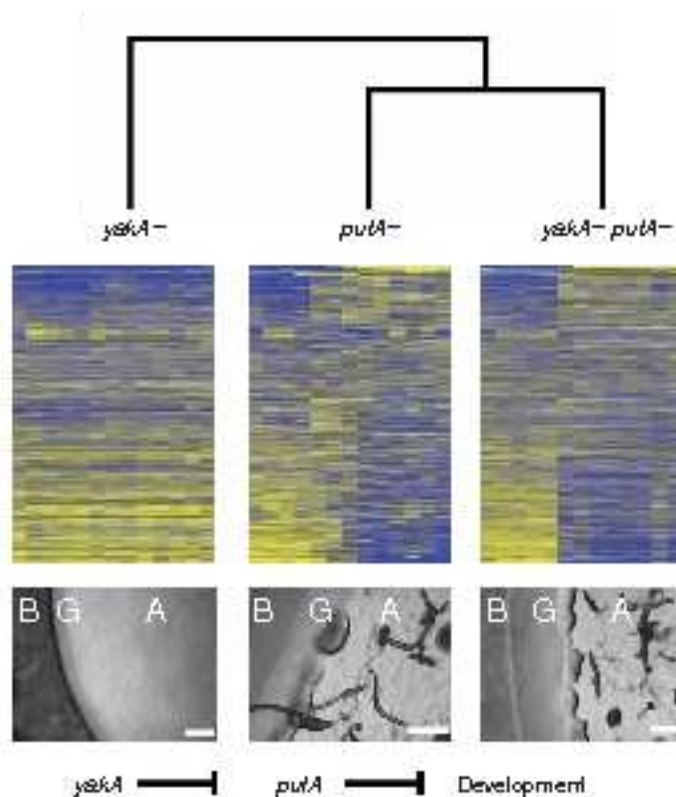


vielzelliger Fruchtkörper



Experimentelles

Epistasis-Analyse mit Transkriptionsprofilen



Wie ähnlich sind die Transkriptionsprofile zueinander?

Mutantenprofile

realer Phänotyp

pufA ist epistatisch zu *yakA*

Experimentelles

Epistasis - Transkriptphänotyp

Vorteile

Vorabwissen ist nicht nötig, Profiling reicht aus

Es muss kein sichtbarer Phänotyp existieren

Es gibt ein Mass für die Ähnlichkeit zwischen Phänotypen

Probleme

selten gibt es einen derart klaren Übergang zwischen zwei Zellzuständen

aus Transkriptionsprofilen allein kann keine Richtung der Abhängigkeit definiert werden