

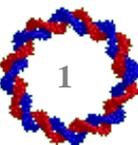
Welche Themen werden behandelt?

- **Gene**

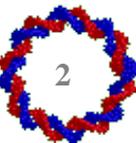
- Anfang und Ende
- Detektion von Genen
- Konservierung als Maß für Funktionalität
- Verschiedene Arten des Sequenzvergleichs

- **Genome**

- Organisation
- Chromosomen
- Zentromere und Telomere



Genanfang



Definition

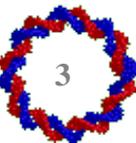
Ein Gen ist:

Eine lokalisierbare Region einer genomischen Sequenz, die einer vererbaren Einheit entspricht.

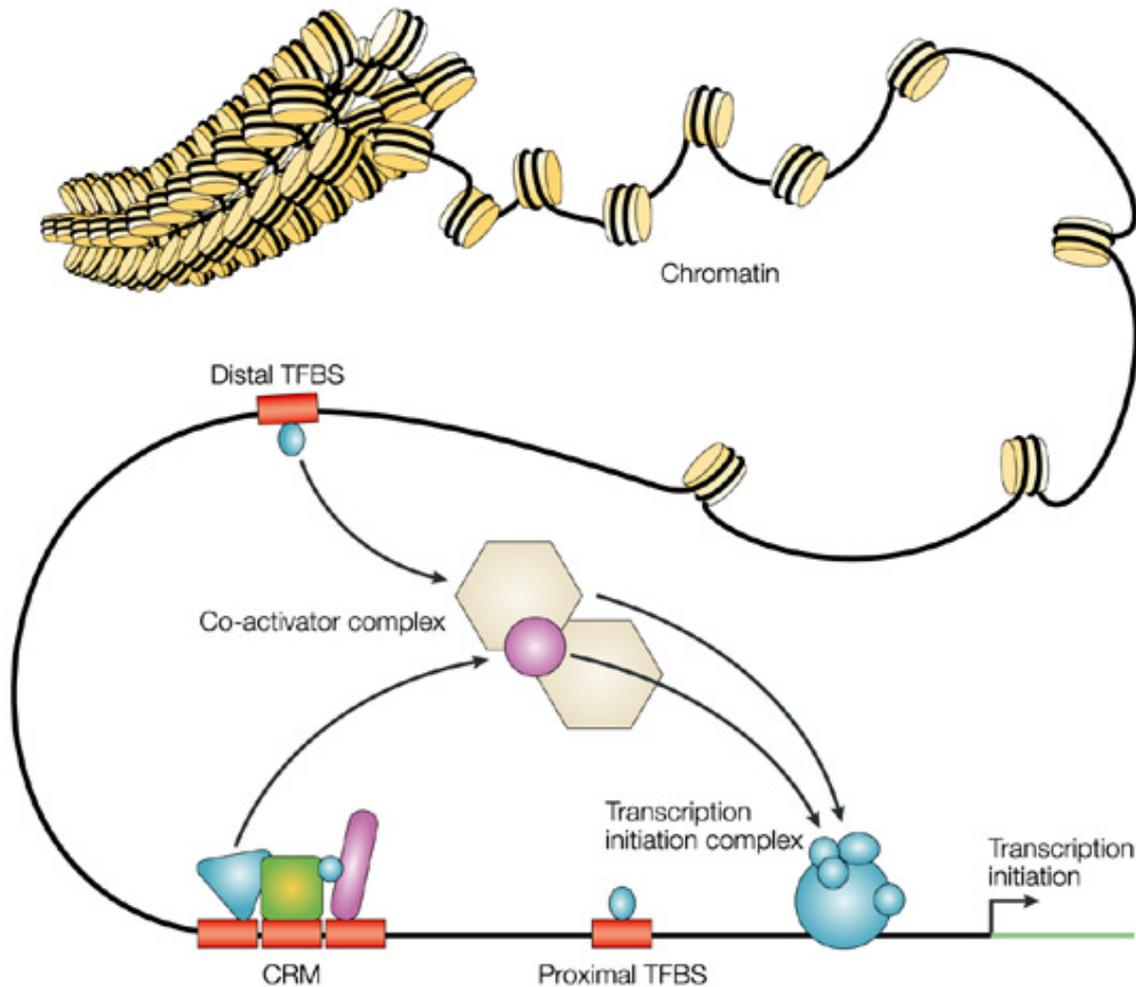
Es besteht aus mehreren Teilen:

regulatorische Regionen

transkribierte Regionen



Transkriptionsstart



TFBS = Transkriptionsfaktorbindungstelle
CRM = *cis* regulatorische Module

Transkriptionsstart

Definition

- experimentell festgelegt (durch cDNA)
- heute:
 - Sequenzvergleiche von ESTs und genomischer DNA
- früher:
 - S1-Nuclease Experimente
 - Primer Extension Methode



Promotor

- Bindungsort für RNA-Polymerase und Transkriptionsfaktoren (TF)
- Initiator (Haushaltsgene)
- GC Box (Haushaltsgene)
- TATA-Box ca. 30 Nucleotide 5' vom Transkriptionsstart (regulierte Gene)

Aufbau ist genspezifisch!!!

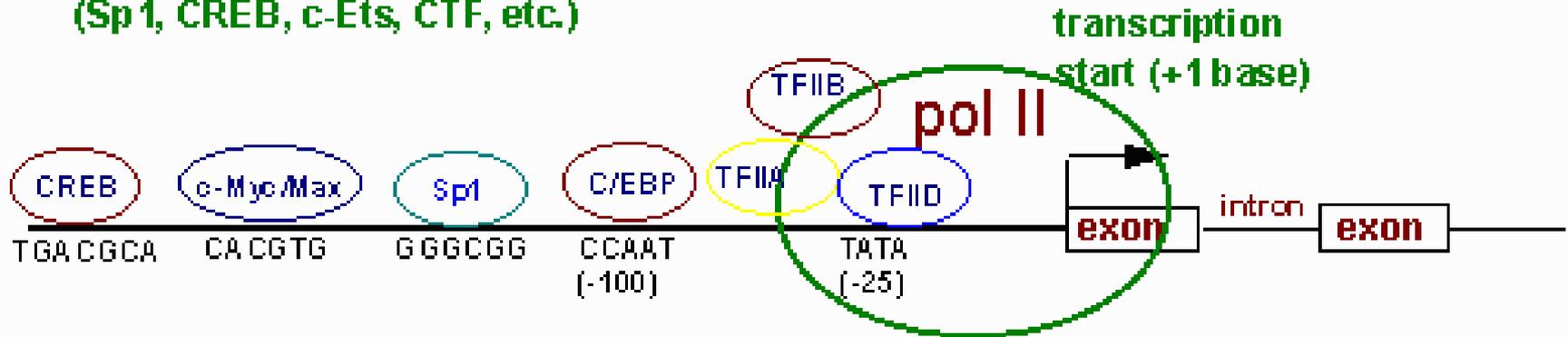


Promotor

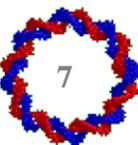
Many factor sites

(e.g. c-Fos, c-Jun, c-Myc/Max)

(Sp1, CREB, c-Ets, CTF, etc.)



- Promotoren enthalten mehrere TF - Bindungsstellen
- Reihenfolge der Bindungsstellen kann in verschiedenen Arten unterschiedlich sein
- Jede Bindungsstelle besteht nur aus wenigen Basen



Falsch positive Promotor-Vorhersagen

Humangenom

Programme sagen Bindungsstellen alle
500 bis 5000 Basenpaare vorher

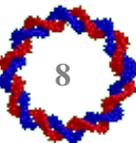
Beispiel: MyoD (muskelspezifischer Transkriptionsfaktor)

1 Bindungsstelle pro 500 Basen

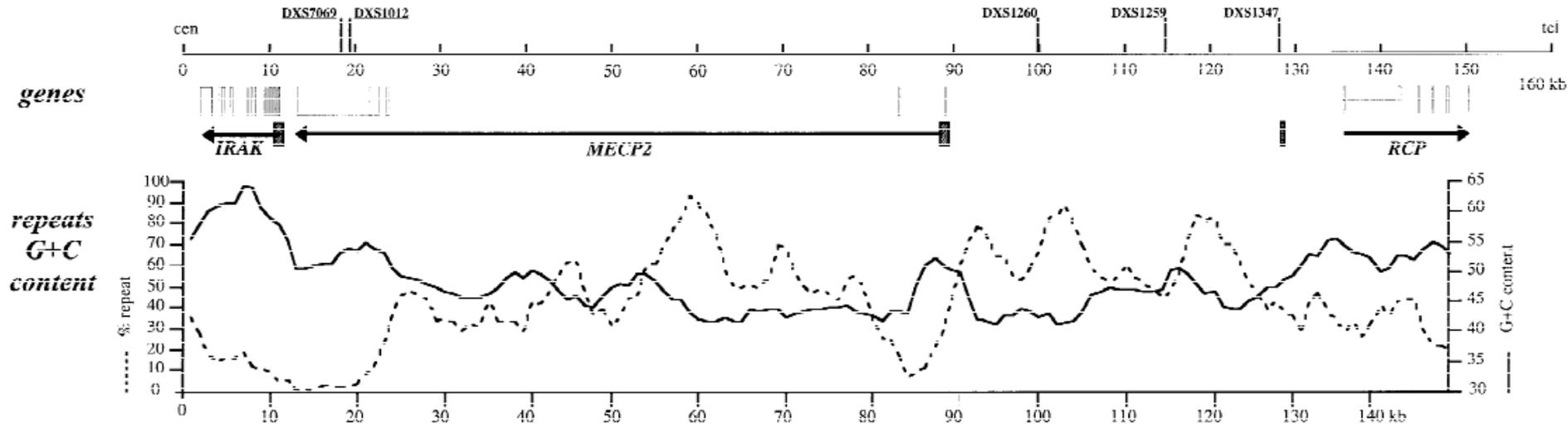
vorhergesagt: 10^6 Bindungsstellen

wahrscheinlich: 10^3 Bindungsstellen

1000 mal mehr falsche Vorhersagen als richtige!



CpG – Inseln



- humanes X-Chromosom, Region: Xq28
- schwarze Striche: Gene
- schwarze Kästen: CpG-Inseln
- weiße Kästen: Exons

CpG - Nucleotide

Definition

- C und G Nukleotidpaare (CpG)
- gehäuft an einer Stelle = CpG-Inseln

Vorkommen

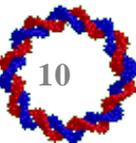
- in den meisten Genomen sehr selten
- im Menschen nur 30 % des erwarteten Anteils in Bezug auf die Nukleotidzusammensetzung

Definition für CpG-Inseln (Mensch)

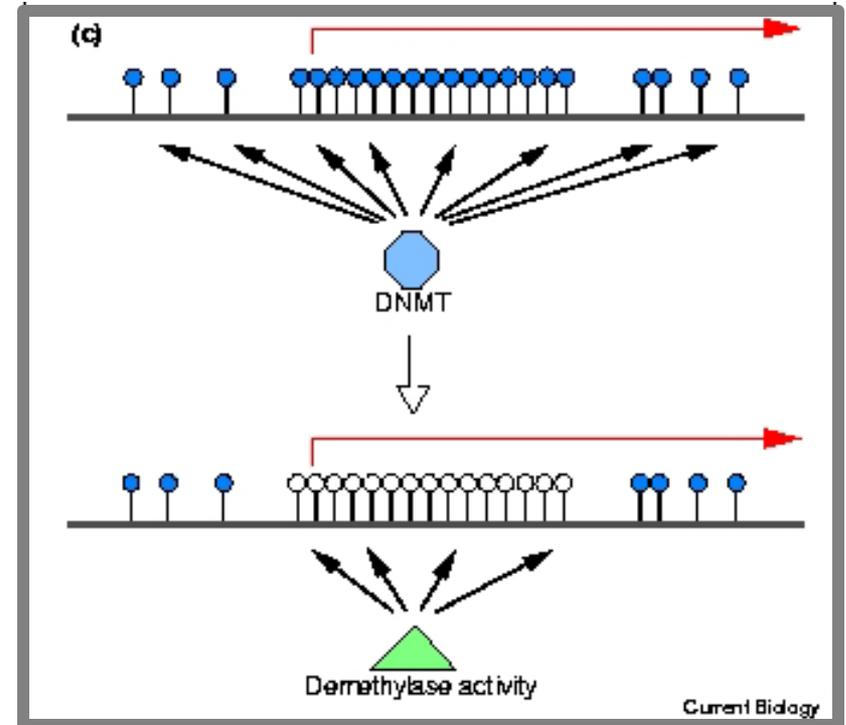
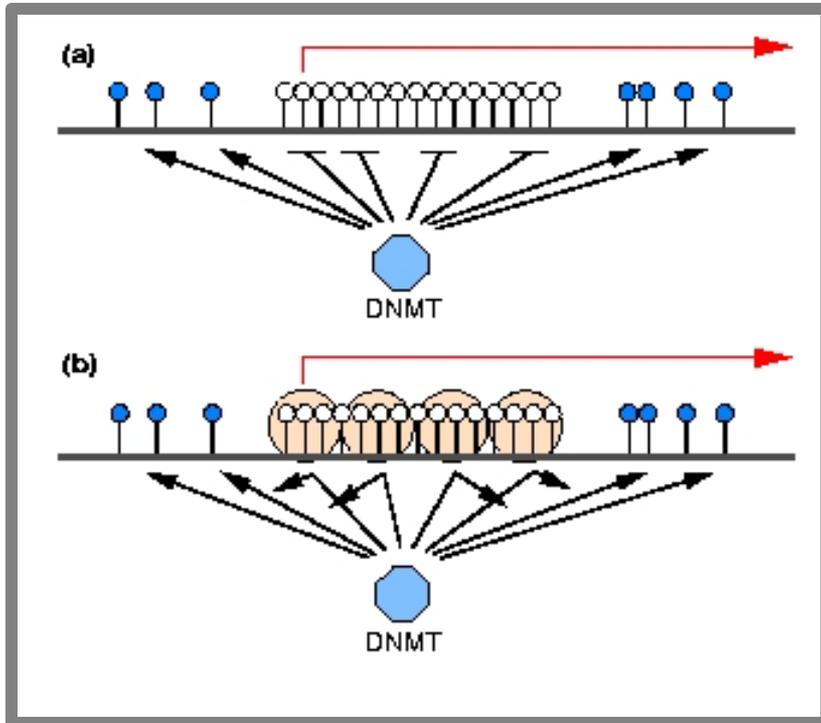
- >200 bp; > 50 % GC; CpG beobachtet/erwartet > 0.6

Bedeutung

- Regulation der Transkription durch Methylierung



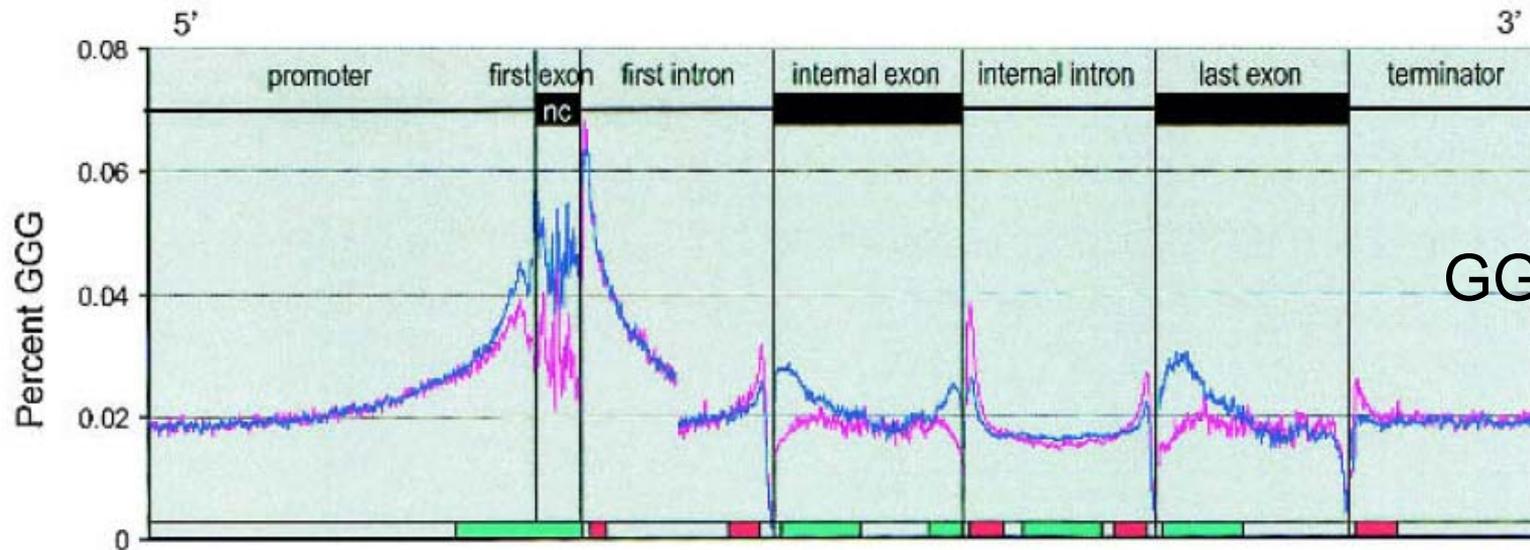
Erhalt des demethylierten Status (Modelle)



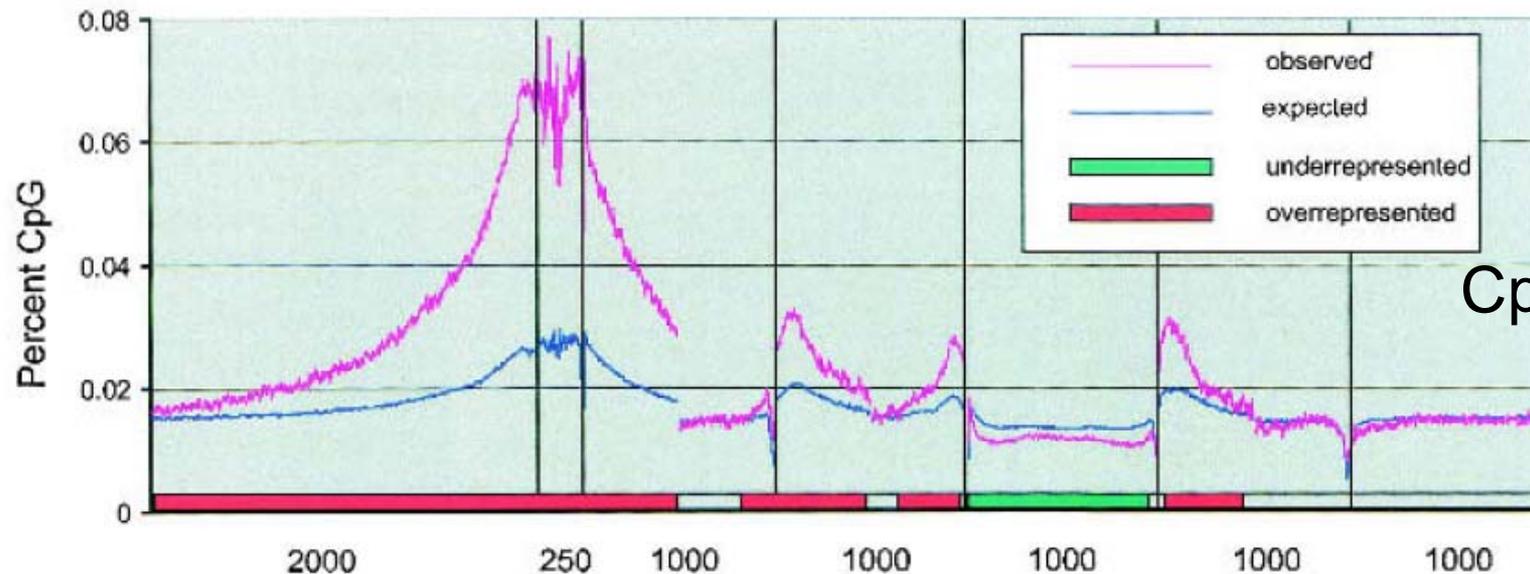
Methylierung blockiert durch
(a) Sequenzmotive
(b) Proteinfaktoren

(c) selektive Demethylierung

Ein typisches humanes Gen

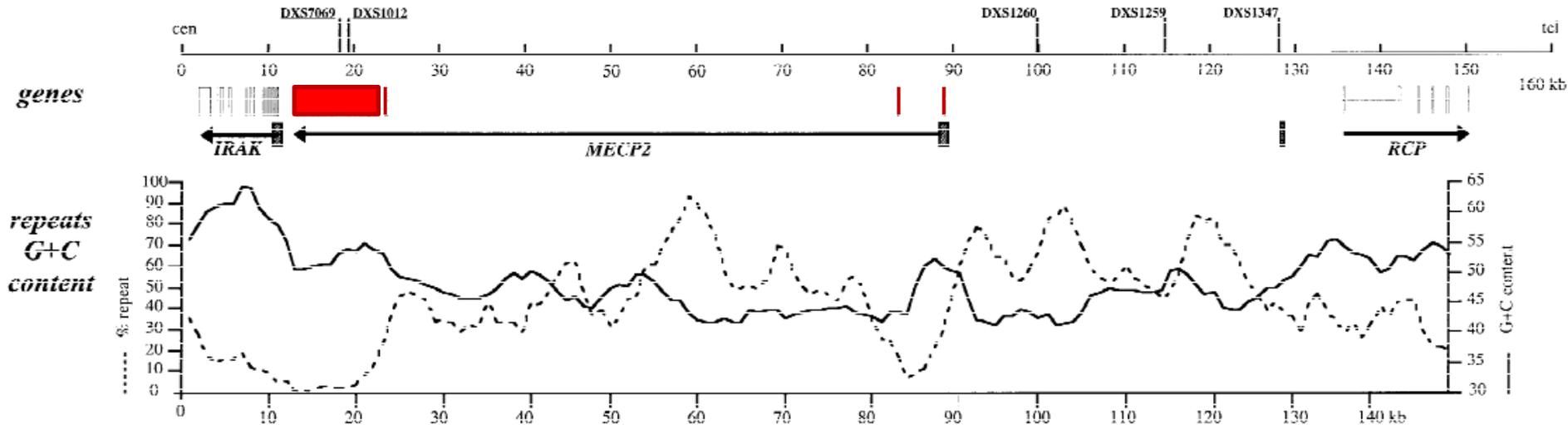


GGG Verteilung



CpG Verteilung

MECP2 - Gen



- humanes X-Chromosom, Region: Xq28
- schwarze Striche: Gene
- schwarze Kästen: CpG-Inseln
- weiße Kästen: Exons, rot für *MECP2*

Genende

Polyadenylierungssignale

Signale

- **AAUAAA**, AUUAAA und weitere Variationen

Lage im Gen

- 3' von der codierenden Sequenz,
- 10 -30 Nucleotide nach AAUAAA Signal wird die RNA geschnitten und polyadenyliert

Häufigkeit im Gen

- mehrere Signale per Gen können vorhanden sein (bei 29 % aller Gene des Menschen)
- das am weitesten 3' liegende Signal wird meist genutzt



Polyadenylierungssignale

```
AAUAAA  
AUUAAA  
AGUAAA  
UAUAAA  
CAUAAA  
GAUAAA  
AAUAUA  
AAUACA  
AAUAGA  
ACUAAA  
AAGAAA  
AAUGAA
```

NNUANA

Figure 4 The 12 putative human polyadenylation signals and their 90% consensus sequence (N = any nucleotide). The consensus does not take into account the relative frequency of signals. Positions conserved in more than 90% of the variants are highlighted.

Polyadenylierungssignale

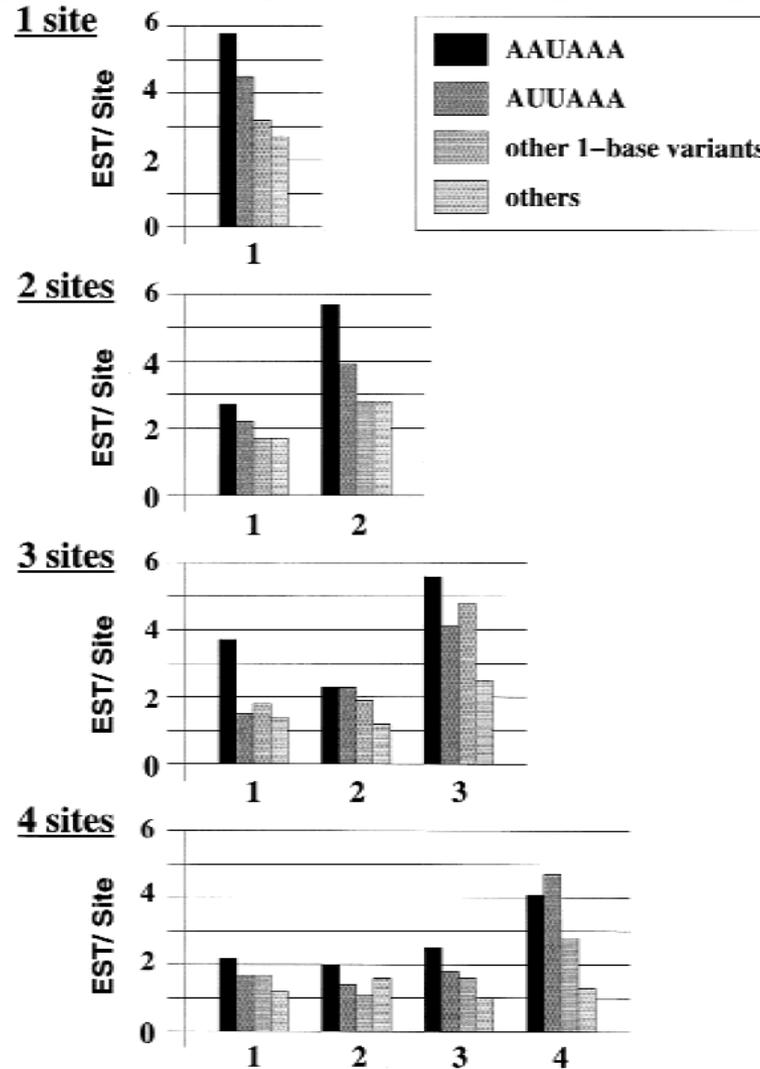
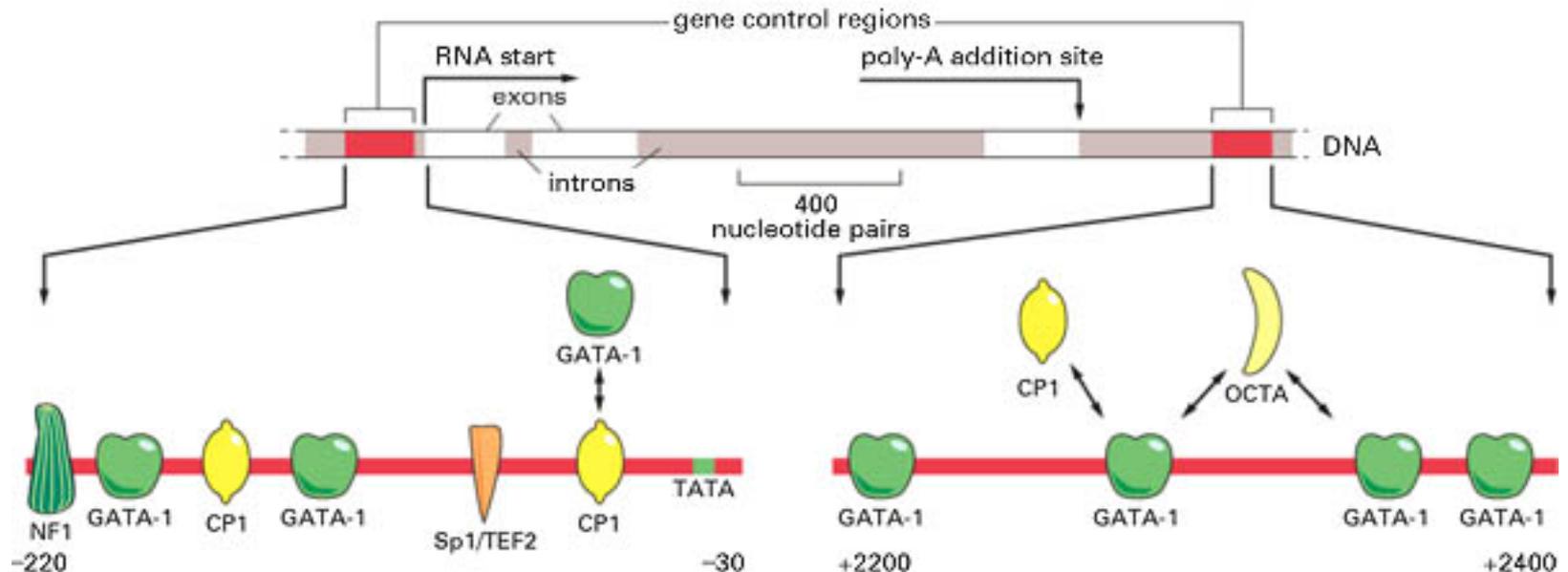


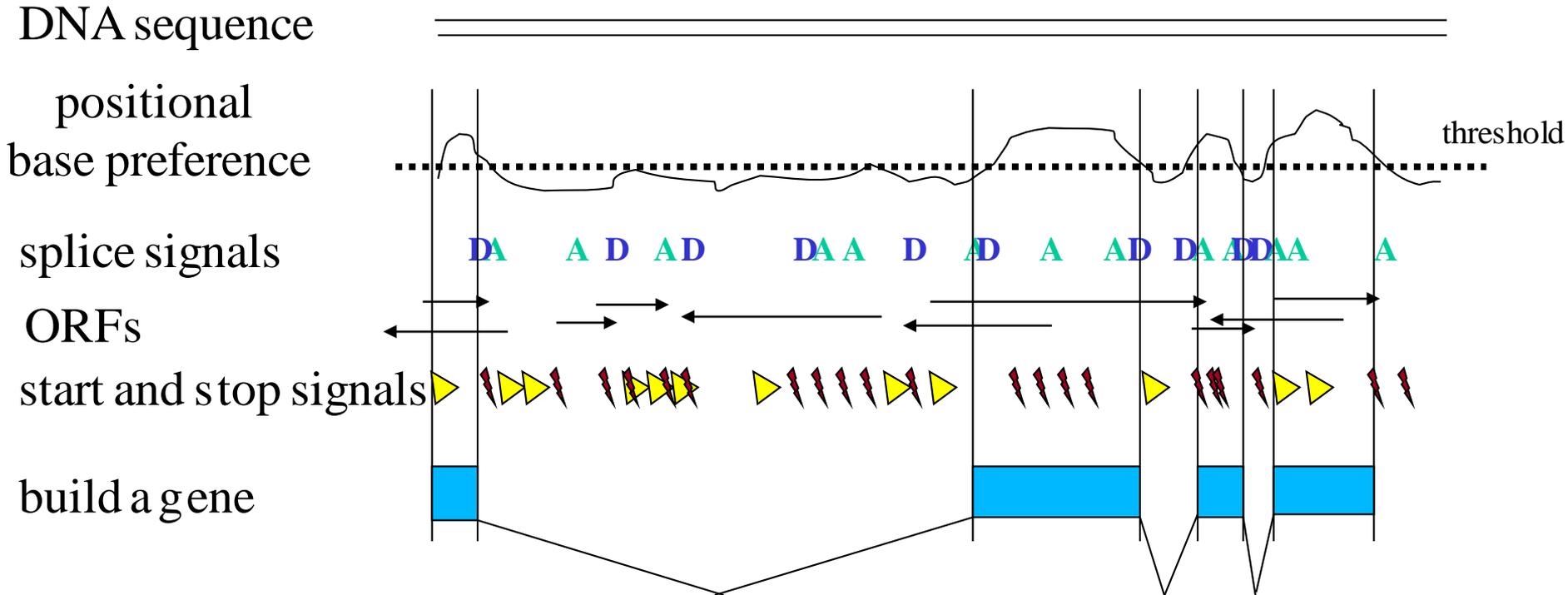
Figure 3 Number of revealing ESTs per poly(A) site, in function of the position of sites in the UTR. From top to bottom: mRNAs with a single poly(A) site identified, mRNAs with two poly(A) sites identified, mRNAs with two poly(A) sites identified, mRNAs with three poly(A) sites identified, mRNAs with three poly(A) sites identified, and mRNAs with four poly(A) sites identified. Poly(A) sites are numbered from 5' to 3'.

Genkontrollregionen



- Transkriptionsbeeinflussende Sequenzen können über den gesamten Genlokus verteilt sein
- Einzelne Motive können mehrfach vorhanden sein

Gendetektion



Danach:

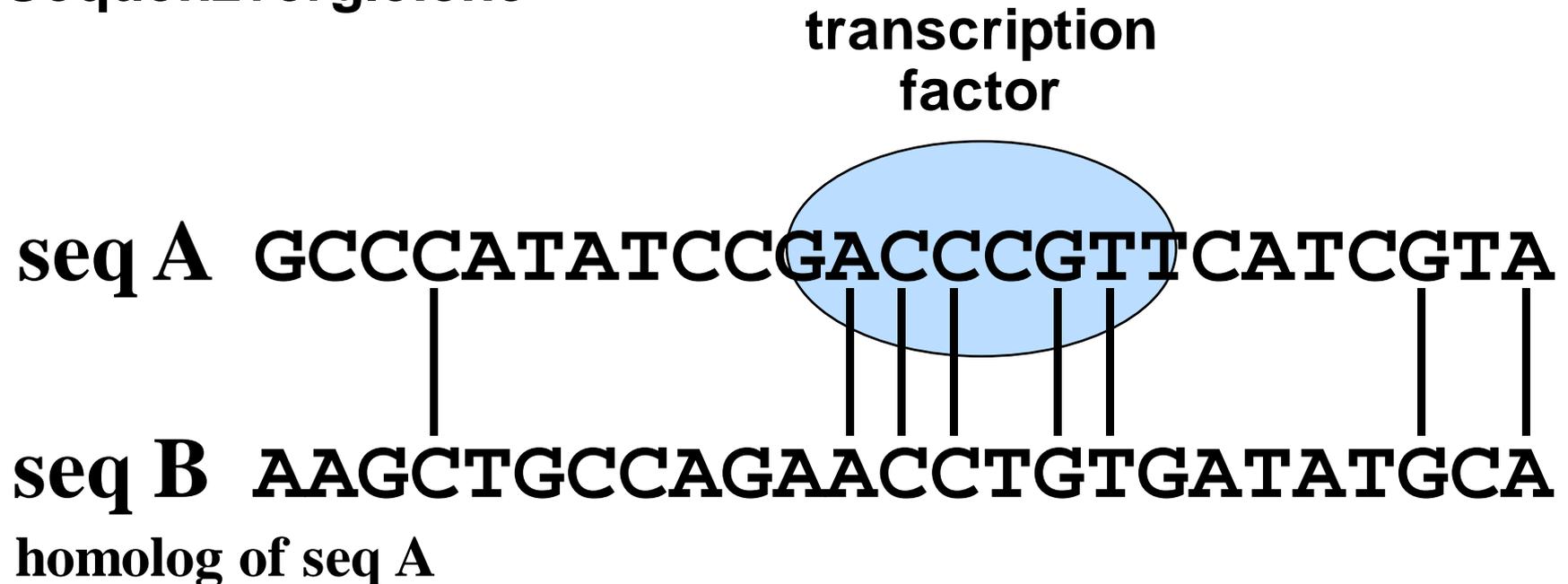
1. Ähnlichkeitsabgleich mit potentiell orthologen Genen
2. Suche nach regulatorischen Elementen

Konservierung als Maß für Funktionalität

Wie können wir Sequenzen mit Funktion finden?

De novo in einer Sequenz: nur experimentell!

Sequenzvergleiche



Problem: Die gesuchten Sequenzabschnitte sind kurz.
→ Allgemeine Regeln existieren nicht!

Konservierung als Marker für Funktion

Theorie:

Nicht translatierte Sequenzen, die wichtige, konservierte, Funktionen vermitteln, unterliegen einem negativen Selektionsdruck.

Konsequenz:

Sequenzen im 5' Bereich eines Gens im selben Organismus oder verschiedener Organismen sind kaum verändert.

Probleme:

Konservierung kann nur wenige Basen betreffen.

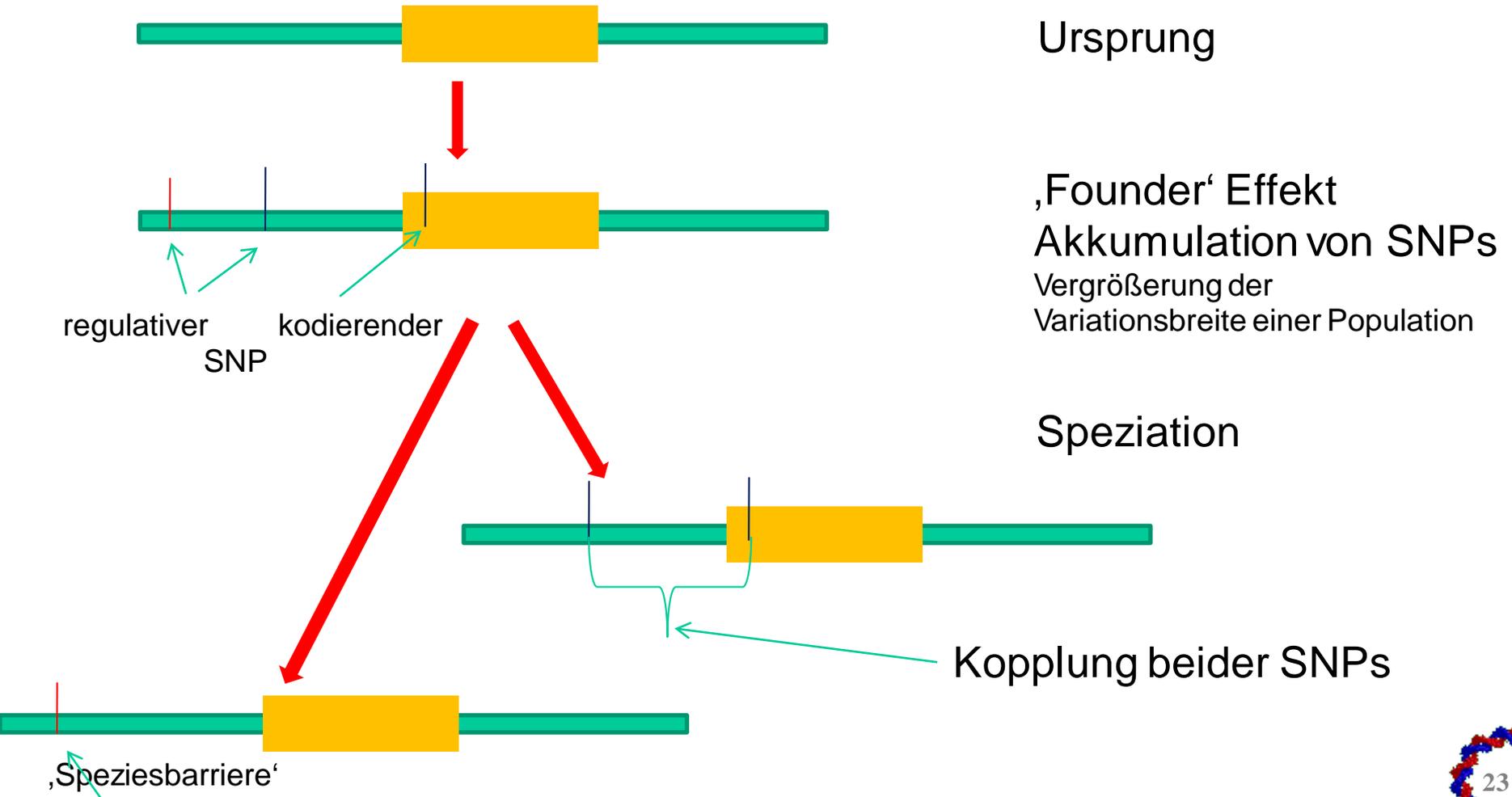
Es gibt auch „zufällige“ Konservierung.

Welche Gene werden verglichen? Orthologe/Paraloge?

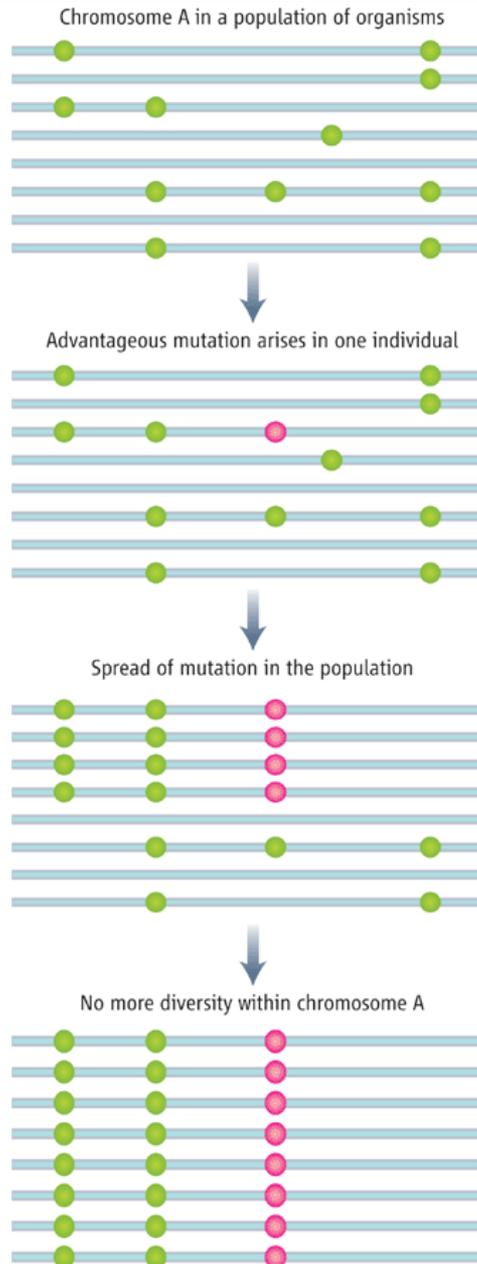


Evolutionär konservierte Sequenzen

Veränderungen in einer Population



Zufällige Konservierung durch Genetic Hitchhiking



Haplotypen können sich über große Regionen erstrecken. Sie können zur Detektion von positiver Selektion eingesetzt werden.

grün=Mutationen ohne Selektionsvorteil
Rot =Mutation mit Selektionsvorteil

Ortholog vs Paralog

Definition:

ortholog

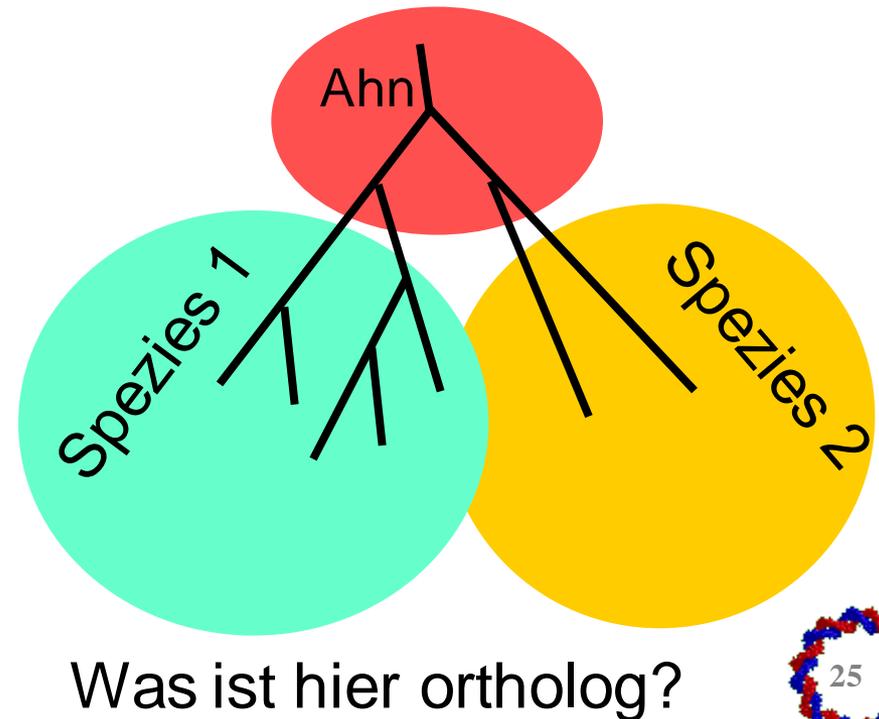
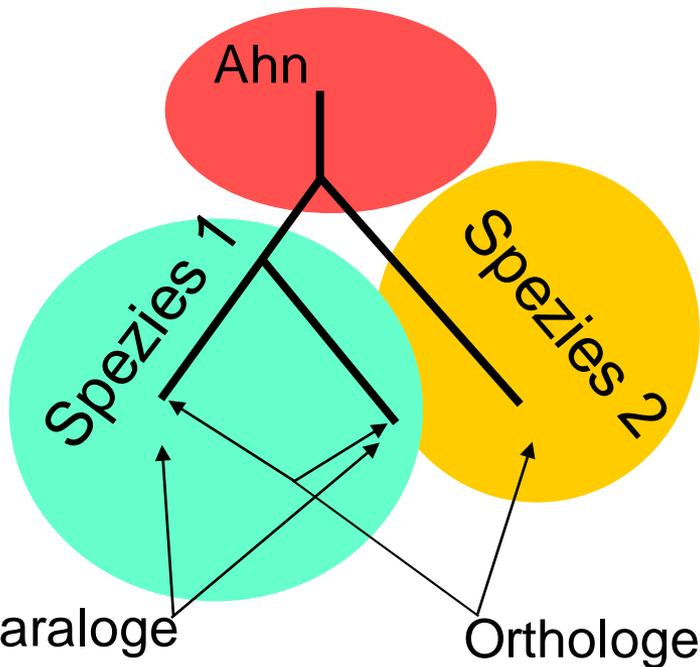
Entstehung durch Speziation

paralog

Entstehung durch Duplikation innerhalb einer Spezies

1:1 Beziehung

Genfamilien



Evolutionär konservierte Sequenzen

Das Zuordnungsproblem

Vergleich codierender Sequenzen, mögliche Definitionen

- ortholog: - sehr ähnlich über die gesamte Länge
- bester Treffer in einem Genom
- paralog: - geringe Ähnlichkeit, aber ebenfalls
über die gesamte Länge
- „zweitbester“ Treffer
- partielle Ähnlichkeit: - Definition von Domänen

Probleme:

- 1] Wo liegt der optimale Schwellenwert, um zwischen Orthologen und Paralogen unterscheiden zu können?
- 2] Gab es differentielle Genverluste, so dass keine orthologen Paare mehr auffindbar sind?
- 3] Genfamilien



Evolutionär konservierte Sequenzen

Pragmatische Bewertung

Was ist der Schwellenwert zwischen ortho- und paralog?

Kriterien:

- in beiden Genomen ist das Gen nur einmal vorhanden
- nachgewiesene gleiche Funktion
- bei Genfamilien und unsicheren Zuordnungen hilft die Berechnung einer Phylogenie

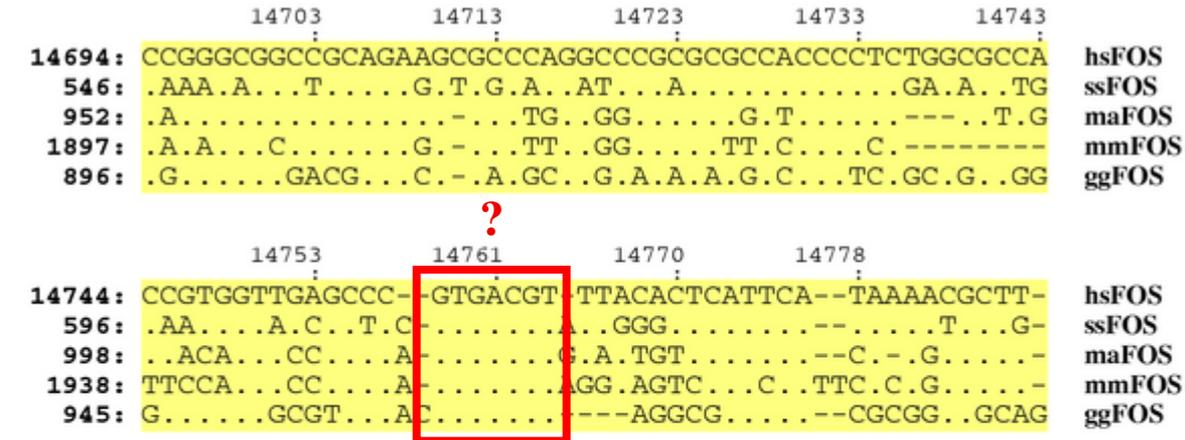
Was ist zufällig konserviert?

Der Anhalter-Effekt (benachbarte Sequenzen werden in die Selektion einbezogen) verhindert die genaue Lokalisation von funktionellen Basen. Deswegen werden Regionen definiert, nicht einzelne Positionen.



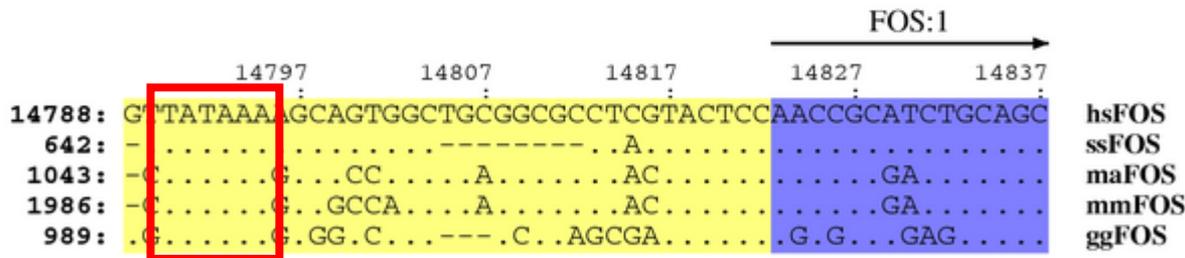
Multiple Alignments

c-Fos Promotor



 CpG Insel

 mRNA



TATA box



Transkriptionsstart
(Homologie-basierte
Bestimmung)

Promotormotive höherer Ordnung

new motif

maIL10	-438	CAAttaTTTctcaatcc	CAGTGTATTCTGGAATGGTCTATT	IgttcgCGTCAct	CTGACCT	-378
C		CAATTATTTCTCAATCC	CA TGTATTCTGGAATGG	C ATTTGT C	CGTCACT	TGACCT
hsIL10	-455	CAAttaTTTctcaatcc	CATTGTATTCTGGAATGGGCAATT	tgtccaCGTCActg	TGACCT	-395
hsIL10	-455	CAAttaTTTctcaatcc	CATTGTATTCTGGAATGGGCAATT	tgtccaCGTCActg	TGACCT	-395
C		CAATTATTTCTCAATCC	A T T TTCTGGAAT G	C ATTT	TCCACGTCA	T TGACCT
mmIL10	-463	CAAttaTTTctcaatcc	TAATATGTTCTGGAATAGCCCATTT	TATccacGTCATT	TATGACCT	-406
mmIL10	-463	CAAttaTTTctcaatcc	TAATATGTTCTGGAATAGCCCATTT	TATccacGTCATT	TATGACCT	-406
B		CAATTATTTCTCAATCC	A T T TTCTGGAAT G	C ATTT	T C CGTCA	T TGACCT
maIL10	-438	CAAttaTTTctcaatcc	CAGTGTATTCTGGAATGGTCTATT	IgttcgcGTCAct	CTGACCT	-378

C/EBP3 CRE4

Sections of two DBA blocks with predicted TFBS. - Block A

single motifs occur together > may constitute a complex binding motif

data: IL-10 gene – human (hs), mouse (mm), Marmota (ma)
software: DNA Block Aligner (DBA)



Promotor- Alignments

Fragen:

Sind die Promotoren aus demselben oder verschiedenen Organismen? (Paralogie vs. Orthologie)

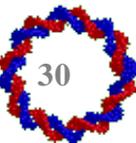
Haben die verglichenen, potentiell regulatorischen Elemente irgendetwas miteinander zu tun?

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?



Verschiedene Arten des Sequenzvergleichs

Evolutionär konservierte Sequenzen

Methoden der Detektion

Sequenzkonservierung ist korreliert mit konservierter Funktion

phylogenetic footprinting:

entfernt verwandte Spezies (Ratte<>Huhn)
(nicht-codierende) Regionen konserviert
zwischen Spezies

phylogenetic shadowing:

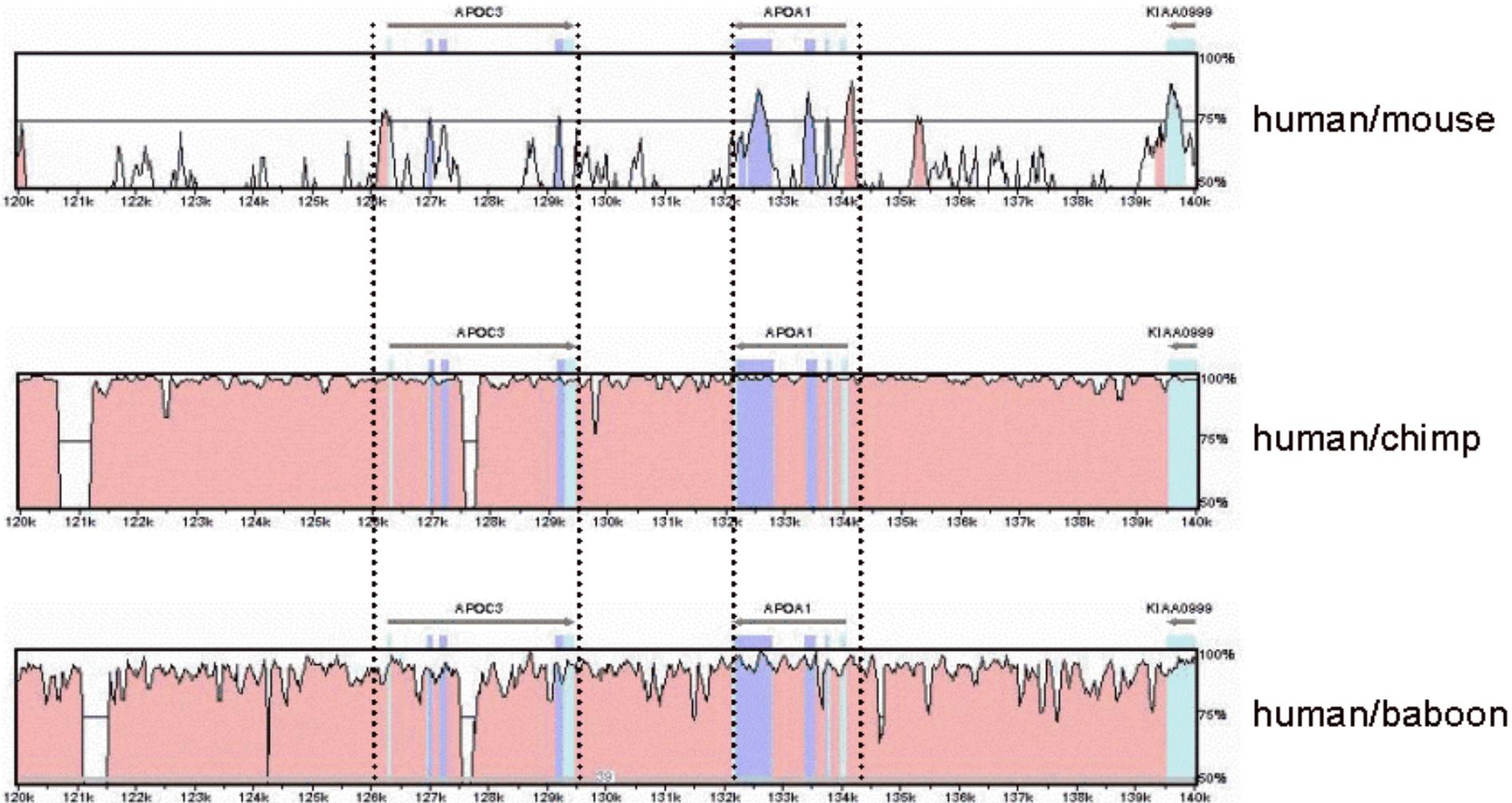
eng verwandte Spezies (Mensch<>Schimpanse)
Unterschiede in (nicht-kodierenden) Regionen



Unterscheidung zufällig/funktional konserviert??

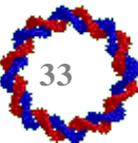


Phylogenetic Footprinting

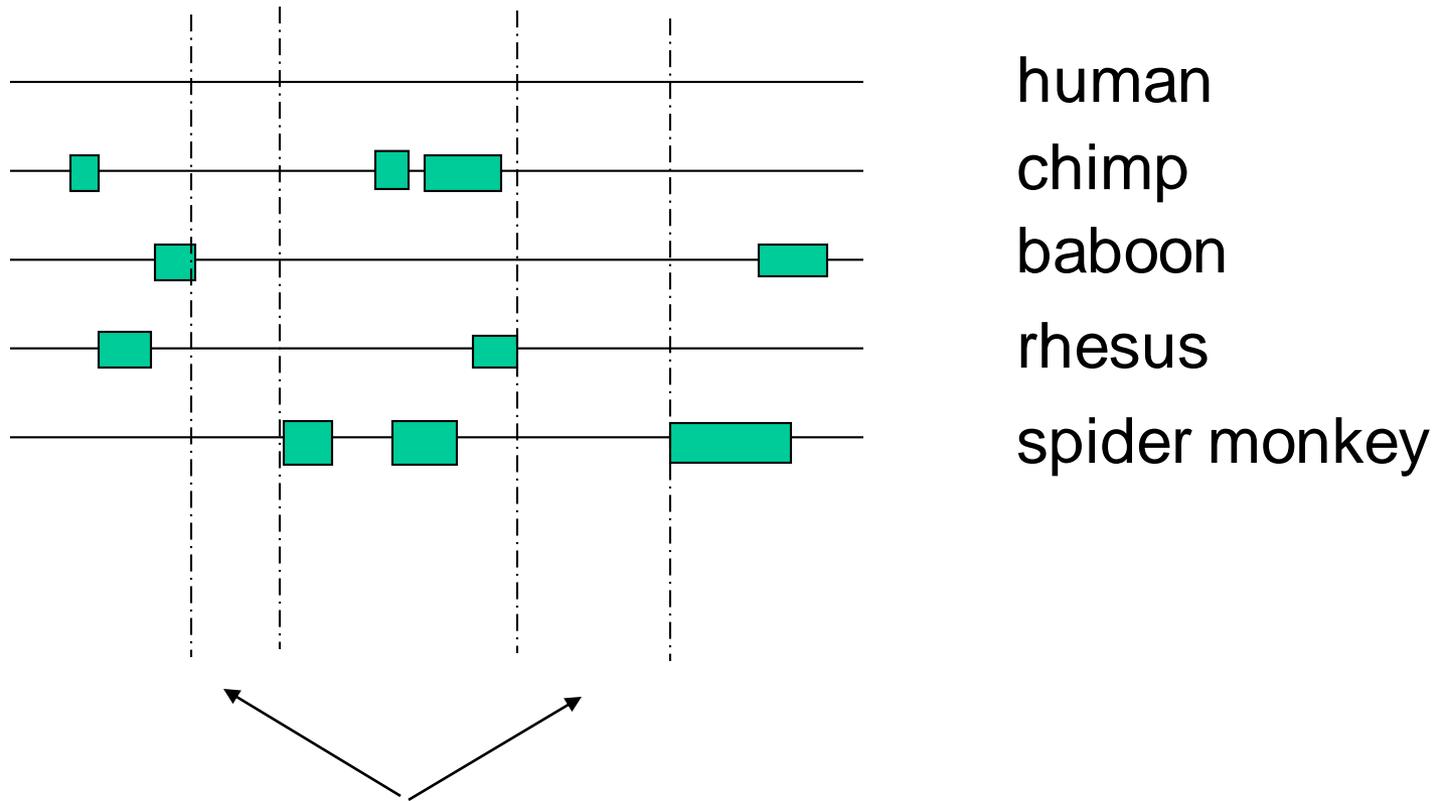


from: www.science.doe.gov

Vergleich von zwei Sequenzen!

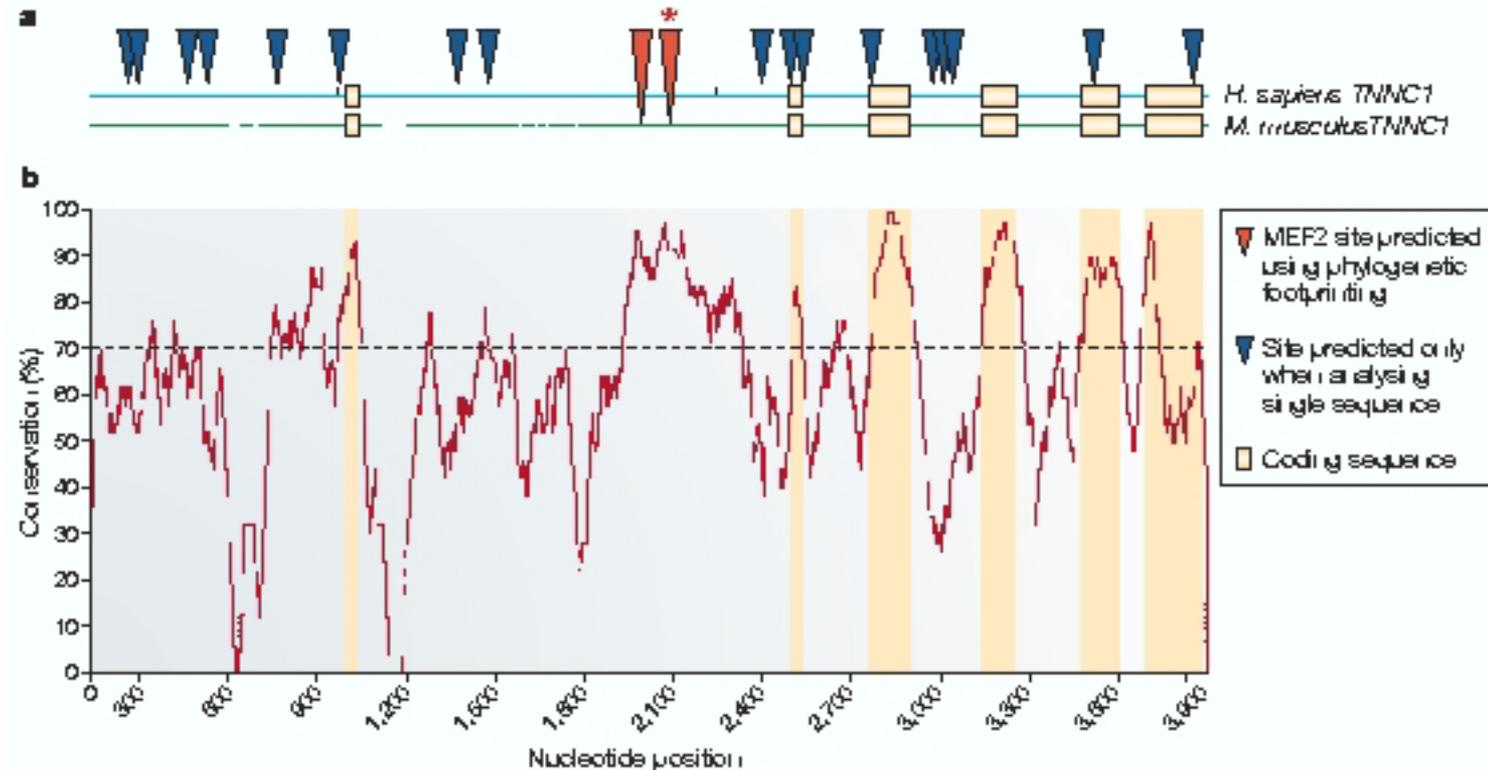


Phylogenetic Shadowing



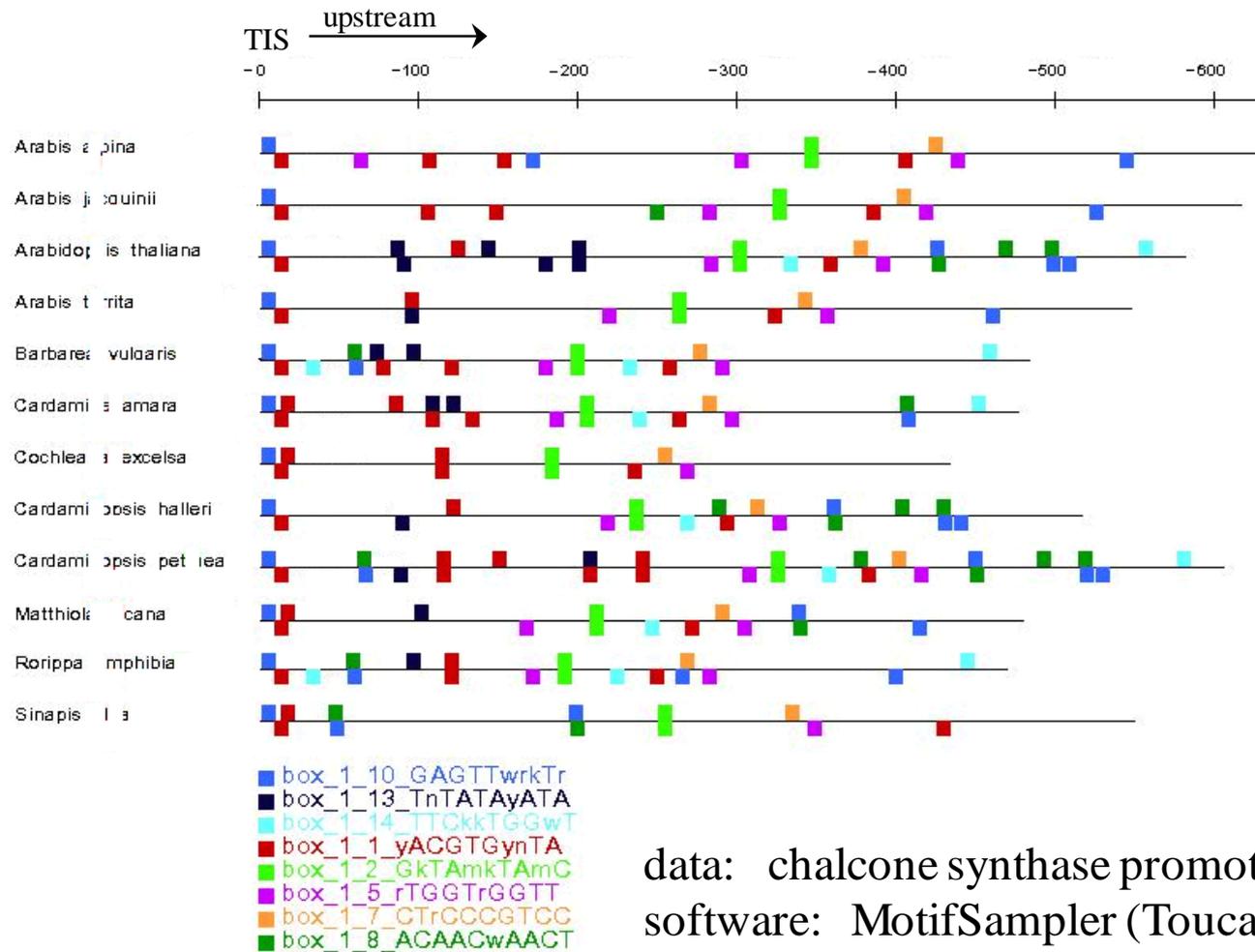
Zwei Bereiche sind konserviert in allen Spezies

Promotor-Detektion durch Footprinting



from: Nature Reviews Genetics 2004, 5 276-287

Promotoren orthologer Gene



Jenseits der Gene

Das Genom

Genom-Informationen

räumliche Zusammensetzung

Nukleotid Ungleichgewichte
,Tupel‘ Verteilungen

Faltungspotential

Erkennungsstellen für spezifische Proteine

,nuclear attachment sites‘

Zentromere, Telomere

Restriktionsenzymbindungsstellen

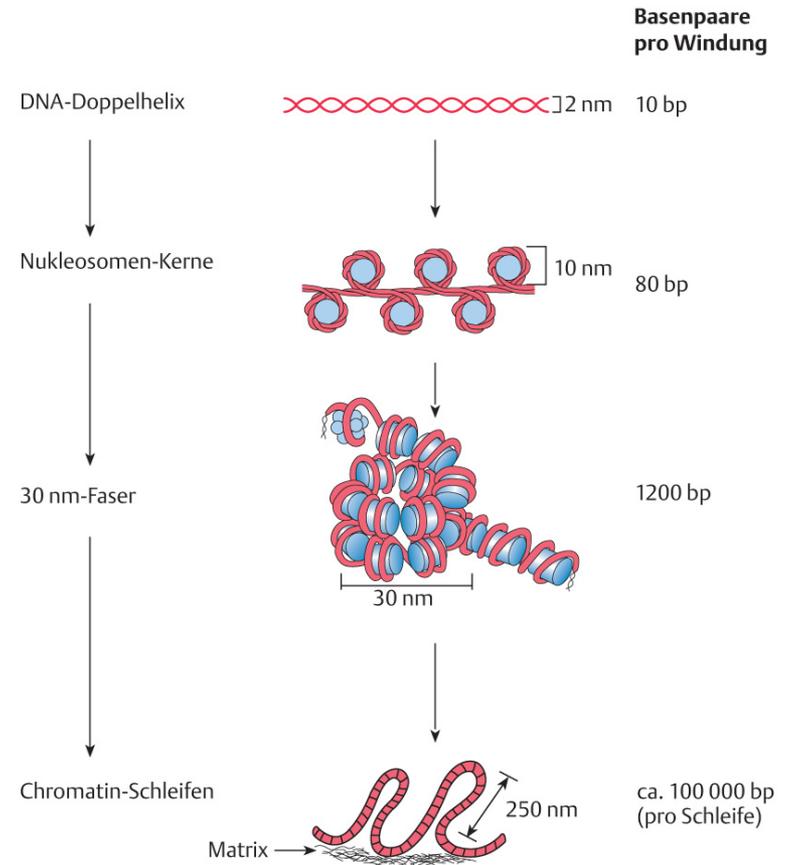
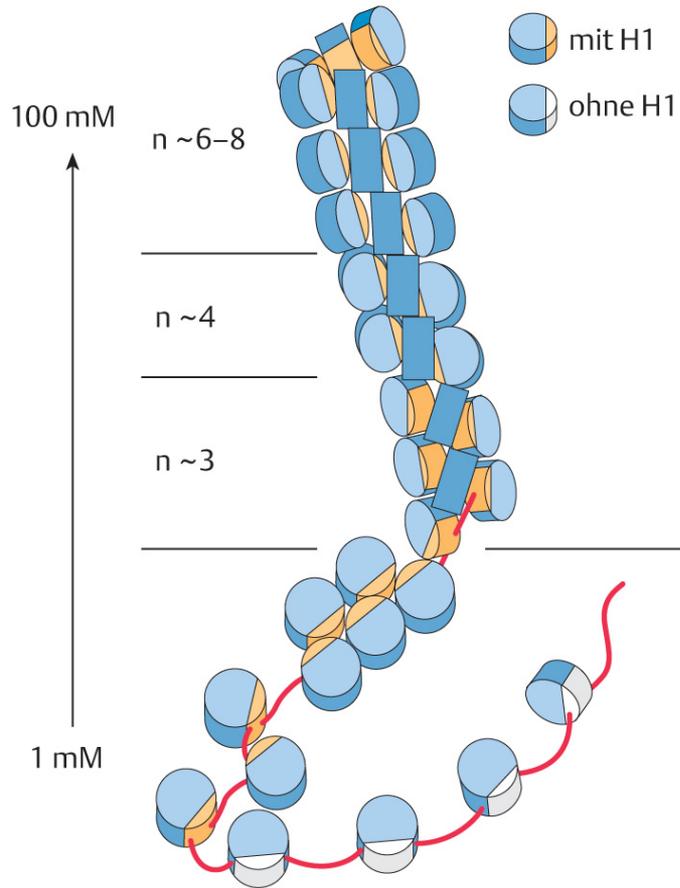
etc.



Genom-Organisation

	Prokaryonten	Eukaryonten
Chromosomen		
Anzahl	meistens 1	viele
Gestalt	zirkulär manchmal linear	linear
Größe	0.5-~10 Mb	<100 kb - einige 100 Mb
Plasmide		
Anzahl	0 ->40	wenige
Gestalt	zirkulär und linear	linear und zirkulär
Größe	1 kb - mehrere Mb	bis zu 20 kb

Chromosomenorganisation

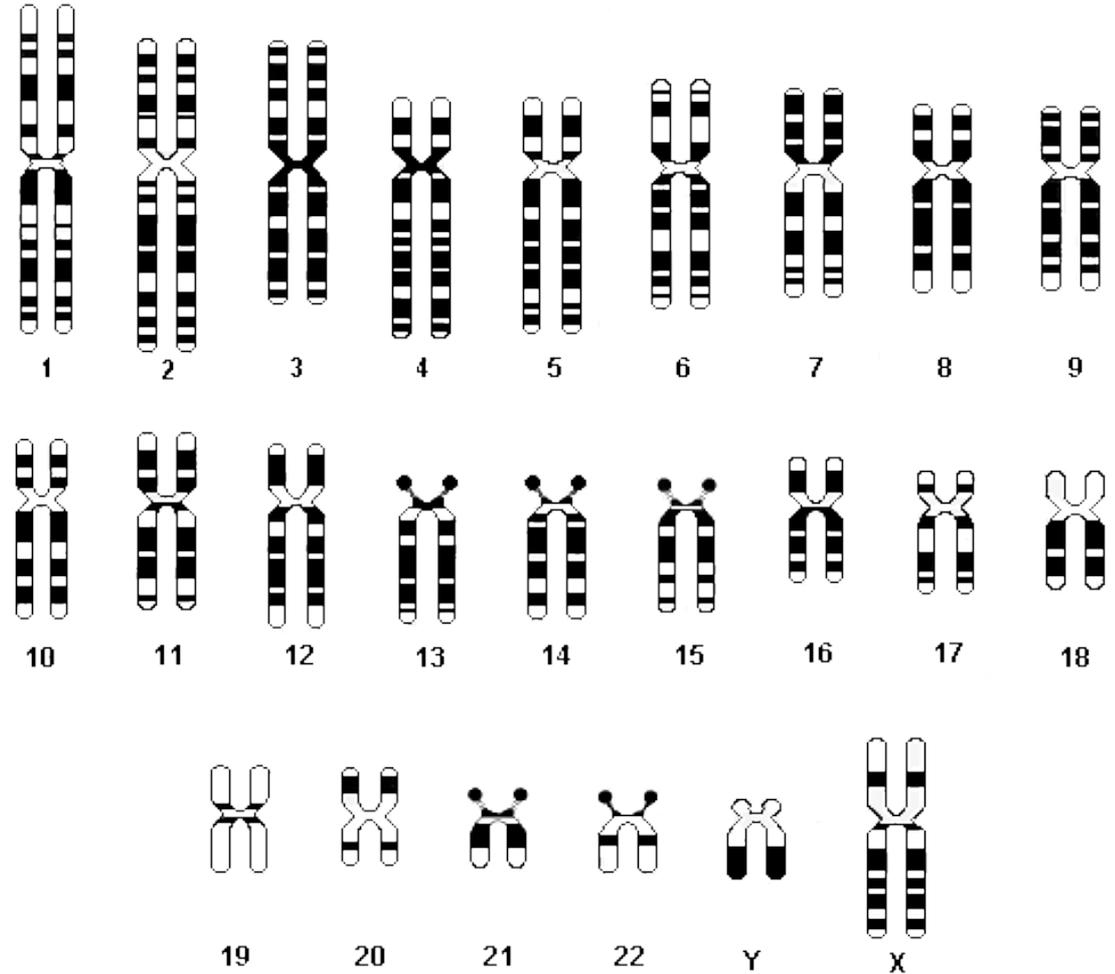
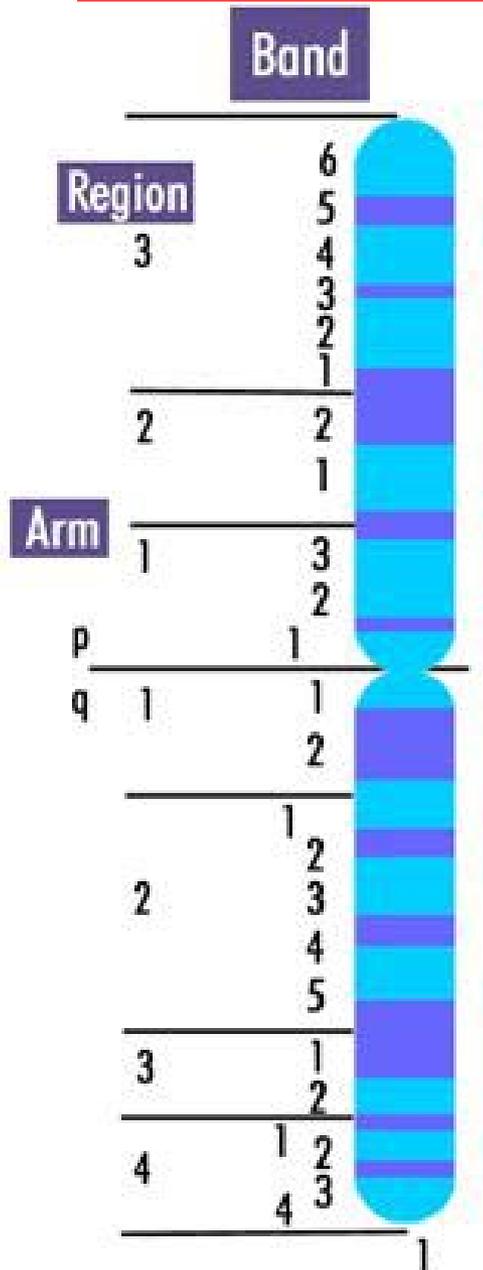


Geschichte der Bandenfärbung

Stain or Banding Technique	Investigator	
Q-banding	Caspersson, Zech, Johansson	1970
G-banding (by trypsin)	Seabright	1971
G-banding (by acetic-saline)	Sumner, Evans, Buckland	1971
C-banding	Arrighi, Hsu	1971
R-banding (by heat and Giemsa)	Dutrillaux, Lejeune	1971
G-11 stain	Bobrow, Madan, Pearson	1972
Antibody bands	Dev, et al	1972
R-banding (by fluorescence)	Bobrow, Madan	1973
In vitro bands (by actinomycinD)	Shafer	1973
Replication banding	Dutrillaux	1973
T-banding	Latt	1973
Silver (NOR) stain	Howell, Denton, Diamond	1973
High resolution banding	Yunis	1975
DAPI/distamycin A stain	Schweizer, Ambros, Andrle	1978
Restriction endonuclease banding	Sahasrabudde, Pathak, Hsu	1978



Chromosome Banding

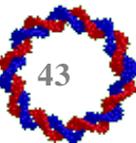


dunkel:
hell:

R Banden
G Bande

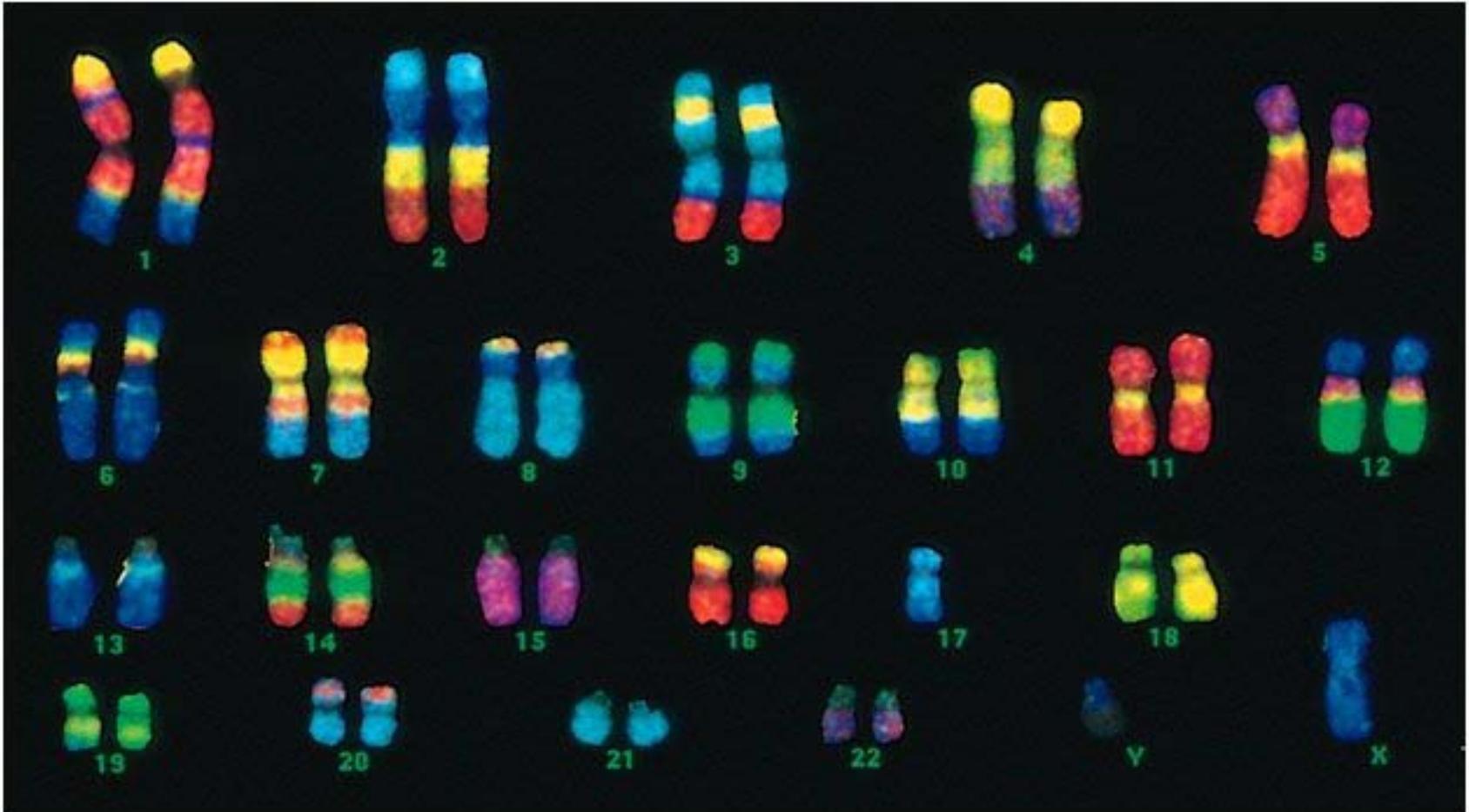
Chromosome Painting

- labelling of chromosome/ region specific probes
- collection of uniquely hybridizing probes
- different labels / combinations of labels are used
- simultaneous hybridization of all chromosomes
- detection by fluorescence microscopy



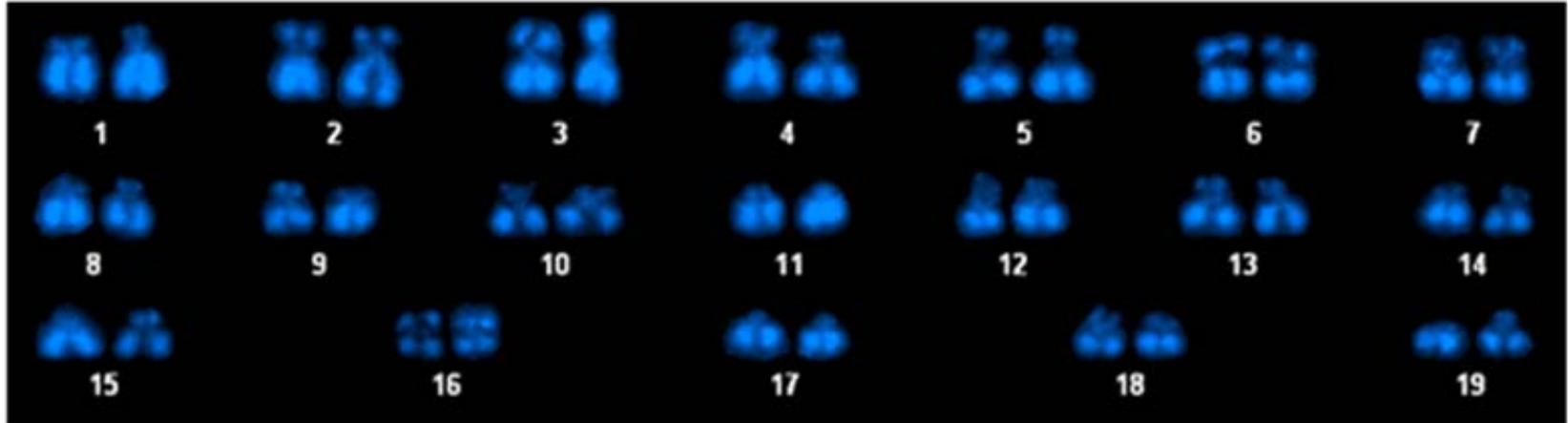
Bandenfärbungen

Karyotyp des Menschen

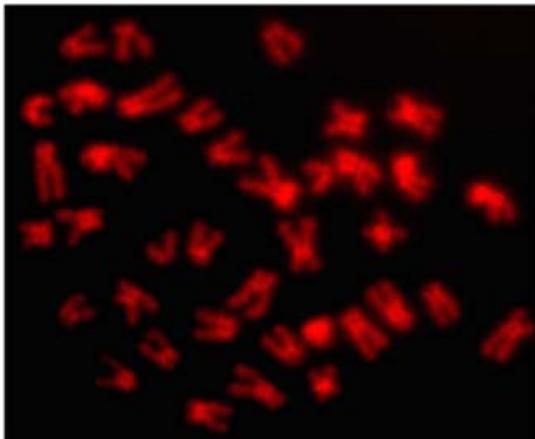


Bandenfärbungen

Karyotyp des Fisches *Nothobranchius*



DAPI (4'6'-diamidino-2-phenylindol) - Färbung zeigt weniger stark gefärbte Heterochromatinbereiche



Distamycin A/mithramycin
Färbung zeigt
G+C reiche Strukturen

Informationsgehalt von Chromosomen

Telomerregionen

Zentromerregionen

Organisatorregionen (rRNA Gene)

Strukturabnormitäten (Diagnostik von Aberrationen)

Translokationen

Verluste

Duplikationen

Heterochromatin



Zusammenfassung Teil II

Genanfänge, Genenden

Gen-Vergleiche

- Orthologe Gene, konservierte Sequenzen
Promotoren
- „gene control“ Regionen, Detektion von
Promotormotiven
- Gendetektion

Genome

- Chromosomen, Banding



Fragen zum zweiten Teil

- Was ist Orthologie?
- Wie kann ein Promotor aufgebaut sein?
- Wie hoch ist momentan die Fehlerrate bei Promotorberechnungen?
- Was ist genomic Footprinting/Shadowing?
- Welche allgemeinen Genomstrukturen kennen Sie?