

Genomprojekte und die Kartierung des menschlichen Genoms

Peter N. Robinson

Institut für medizinische Genetik
Charité Universitätsmedizin Berlin

14. Juni 2008

Rekombination

- Prinzip der Kartierung: Bestimmen, wie oft zwei Loci durch die meiotische Rekombination getrennt werden
- Die Rekombinationshäufigkeit ist ein Maß für den genetischen Abstand
- z.B. wenn zwei Marker auf unterschiedlichen Chromosomen liegen, segregieren sie unabhängig voneinander

Crossing-over (Rückblick)

- Liegen zwei Marker auf demselben Chromosom, könnte man erwarten, dass sie immer zusammen segregieren
- Dies lässt das meiotische Crossing-Over außer acht

Urheberrechtlich geschütztes Bild entfernt

Rekombinaten und Nichtrekombinanten

Urheberrechtlich geschütztes Bild entfernt

- In Generation III können wir zwischen Nachkommen unterscheiden, bei denen das Spermium des Vaters rekombinant oder nicht rekombinant war.
- Da die Mutter in Generation II homozygot ist, lässt sich nicht feststellen, ob Nachkommen aus rekombinanten oder nicht rekombinanten Oozyten hervorgegangen sind.

- Eine Rekombination wird nur selten zwei Loci trennen, die auf einem Chromosom dicht beieinander liegen
- Deshalb werden Gruppen von Allelen, die auf demselben kurzen Chromosomenabschnitt liegen, eher zusammen als Block übertragen → [Haplotyp](#)
- Haplotypen kann man in Stammbäumen und auch in Populationen verfolgen
- Je weiter die beiden Loci auf einem Chromosomen voneinander getrennt sind, umso größer ist die Wahrscheinlichkeit, dass sie durch eine Rekombination voneinander getrennt werden.

centiMorgan

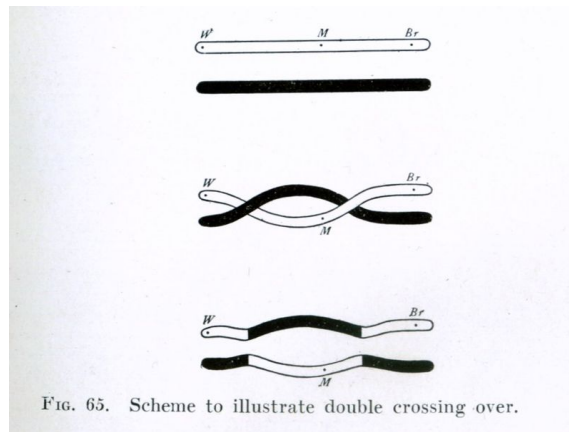
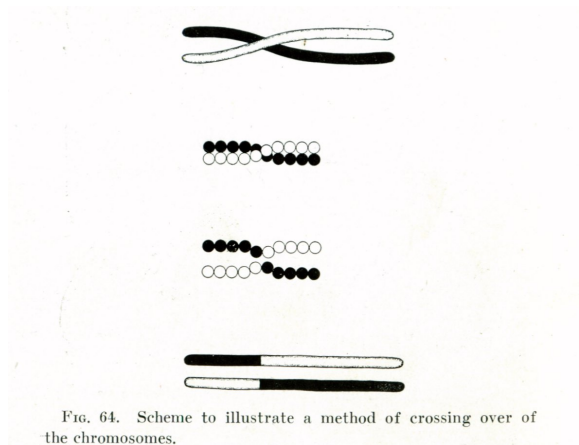
- Zwei Loci, die ein Prozent Rekombination zeigen, sind nach der Definition einer genetischen Karte ein [Centimorgan](#) (cM) voneinander getrennt
- Ein cM \approx 1.000.000 Nukleotide
- Nicht dasselbe wie physikalische Abstände



Thomas Hunt Morgan (1866-1945)

[Wikipedia Commons](#)

1fache und 2fache Crossing-Over



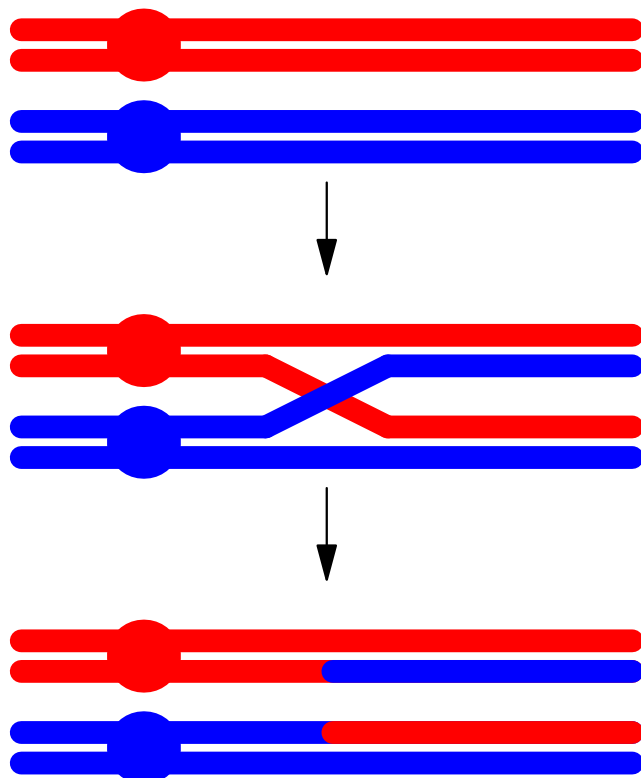
Thomas Hunt Morgan's A Critique of the Theory of Evolution

(1916), Wikipedia Commons

- Bei größeren Abständen muß mit mehrfachen Crossovers zwischen zwei Loci gerechnet werden
- Dieses Schema von Morgan ist zu simpel, da insgesamt vier Chromatiden (b.w)

Rekombination

- Bei einem einzigen Crossing-Over entstehen zwei rekombinante und zwei nichtrekombinante Chromosomen
- Daher entspricht die Rekombinationshäufigkeit θ zwischen zwei Loci genau der **halben** Wahrscheinlichkeit, dass sich ein crossing-over zwischen den Loci ereignet.



Rekombinationshäufigkeit ist nie größer als 50%

Auch bei mehreren Crossing-over-Ereignissen kommt es zu maximal 50% Rekombinanten

Urheberrechtlich geschütztes Bild entfernt

Kartierungsfunktion

- Da Rekombinationshäufigkeiten nie 50% übersteigen, kann man sie nicht einfach addieren, um die physikalische Entfernung zu schätzen
- z.B., wenn auf einer Karte eine Folge von Loci A,B,C,D,E,F... in Abständen von 10 cM liegen, dann ist zwar Locus A zwar 60 cM von Locus F entfernt, aber die Rekombinationshäufigkeit zwischen A und F beträgt nicht 60%
- Die Beziehung zwischen genetischem Abstand und Rekombinationsanteil wird stattdessen durch eine *Kartierungsfunktion* beschrieben.

Kartierungsfunktionen (1)

- Die einfachste (und erste) Kartierungsfunktion stammt von Morgan:

$$\theta = m$$

- θ = Rekombinationsfraktion, m = genetischer Abstand (map distance)
- Gilt nur, wenn die Wahrscheinlichkeit für eine Doppel-Crossing-Over vernachlässigbar klein ist, etwa $\theta < 0.10$

Kartierungsfunktionen (2)

- Tritt ein Doppel-Crossing-over zwischen Loci A und B auf, wird keine Rekombination zwischen den loci *beobachtet*.
- Kartierungsfunktionen passen den Anteil der beobachteten Rekombinanten an, in dem einfache Crossing-over-Ereignisse 1mal, 2fache Crossing-over-Ereignisse 2mal gezählt werden
- Sei m die genetische Entfernung (beobachtete Rekombinanten)
- Dann ist $2m$ die Anzahl der Crossing-Over-Ereignisse
- Haldanes Funktion modelliert die Wahrscheinlichkeit eines Doppel-Crossing-Overs durch eine Poisson-Verteilung mit $\lambda = 2m$. Die Wahrscheinlichkeit, dass kein Doppel-Crossover auftritt ist dann e^{-2m} .

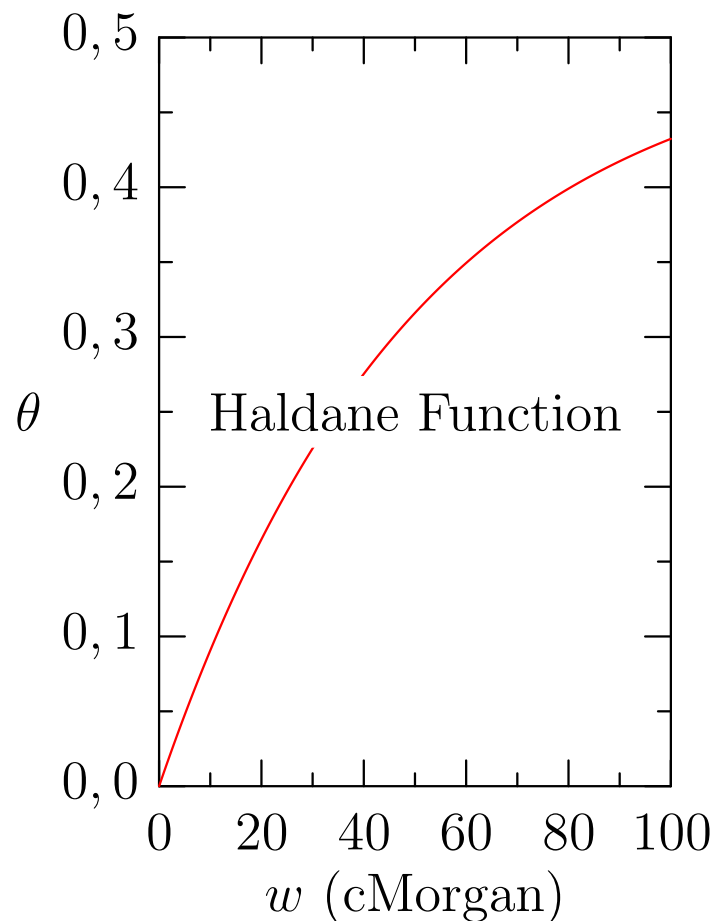
Haldane-Funktion

- m : Abstand auf der Genkarte in Morgan
- θ : *Beobachtete* Rekombinationshäufigkeit

$$\theta = \frac{1}{2} [1 - e^{-2w}]$$

So variiert die Rekombinationsfraktion zwischen 0 und 0,5.

Haldane-Funktion



Gesamtlänge einer genetischen Karte

Urheberrechtlich geschütztes Bild entfernt

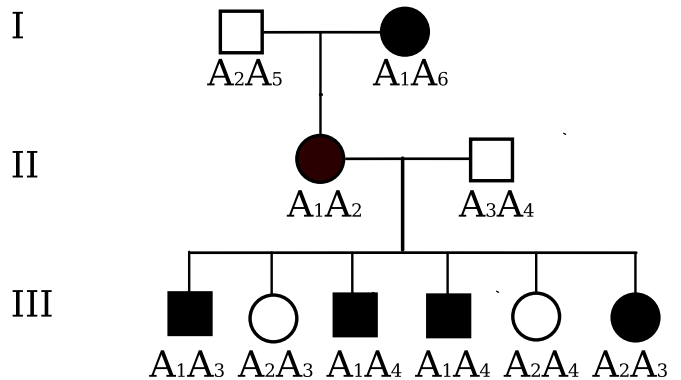
- Jedes Crossing-over erzeugt 2 rekombinante und 2 nichtrekombinante Chromatiden, so dass ein Crossing-over 50 cM zur genetischen Karte beiträgt
- Männliche Meiose: durchschnittlich 50,6 Chiasmata pro Zelle, weiblich 70,3 Chiasmata → genetische Kartenlänge 2,6 Morgan (m) bzw. 4,3 Morgan (w)

Physikalische und genetische Karten

- Faustregel: $1 \text{ cm} \approx 1 \text{ Mb} = 10^6 \text{ Nukleotide}$
- Variiert stark (ca. 0,3 cM pro Mb bis über 3 cM pro Mb)

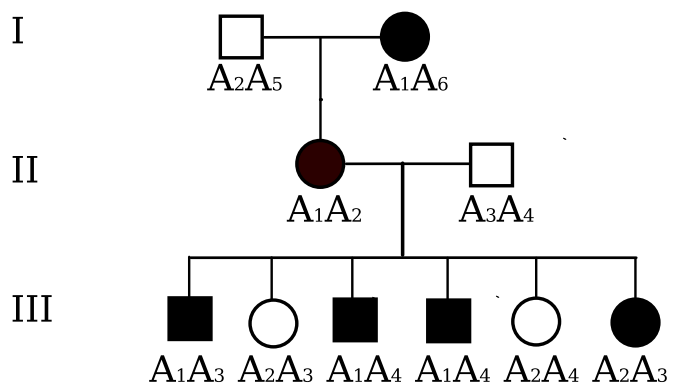
Stammbaumanalyse

- Blau: erkrankt
- A_1, A_2, \dots : Allele eines Markers
- Eine **Kopplung** eines Markers an den Phänotyp kann ein Hinweis sein, dass das entsprechende Krankheitsgen in der Nähe des Markers gelegen ist.



Stammbaumanalyse

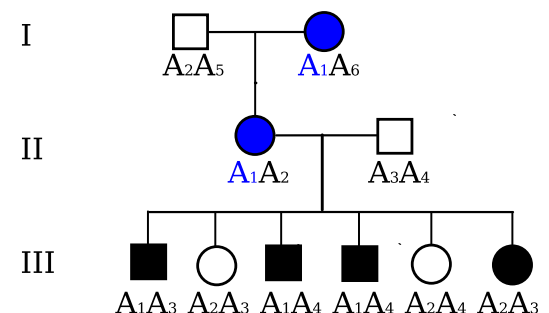
- 5/6 Kinder von II.1 sind nichtrekombinant, 1/6 rekombinant
- Zufall oder Kopplung?
- **lod-Wert-Analyse**: Standard-Methode, um die statistische Signifikanz bei der Stammbaumanalyse zu prüfen.



- logarithm of the odds
- Der LOD-Score ergibt sich aus der Summe der dekadischen Logarithmen der "Wahrscheinlichkeit" der Chance (odds) auf Kopplung.
- Es werden zwei Modelle miteinander verglichen
 - ① Kopplung: Wahrscheinlichkeit ergibt sich aus Stammbaum und θ
 - ② Keine Kopplung: Wahrscheinlichkeit immer $(1/2)^n$, wo n die Anzahl der beobachteten Meiosen ist.

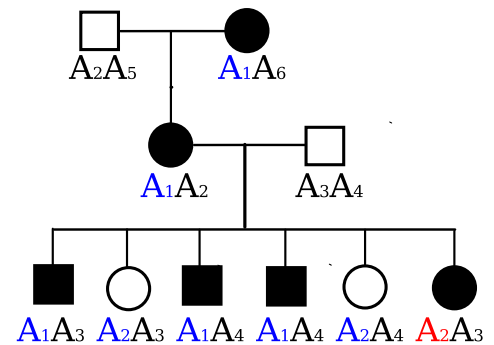
Iod: Schritt für Schritt (1)

- Phase bestimmen
- II_1 erbt Allel A_1 und Krankheit von I_2 .
- \Rightarrow Phase: Kopplung von Krankheitsgen an Markerloкус A



Iod: Schritt für Schritt (2)

- Anzahl der rekombinanten und nicht rekombinanten Personen im Stammbaum bestimmen
- → 5 nichtrekombinante, 1 rekombinante Person
 - ▶ Drei Erkrankte in Generation III erben Allel A_1 (nichtrekombinant)
 - ▶ Zwei Gesunde in Generation III erben Allel A_2 (nichtrekombinant)
 - ▶ Eine erkrankte Person in Generation III erbt Allel A_2 (rekombinant)



Iod: Schritt für Schritt (3)

- Rekombinationshäufigkeit $\theta \in [0, 0.5]$ wählen

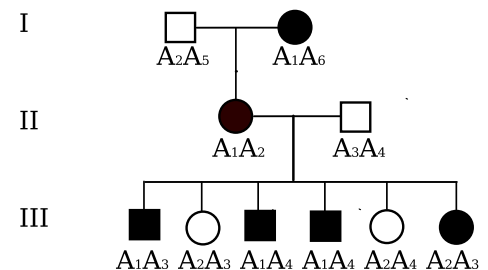
Iod: Schritt für Schritt (4)

- Wahrscheinlichkeit der Kopplung bestimmen

→ bei n Meiosen mit k Rekombinanten

$$\theta^k (1 - \theta)^{n-1}$$

→ In diesem Fall $\theta^1 (1 - \theta)^{6-1}$



Iod: Schritt für Schritt (5)

- Wahrscheinlichkeit des Stammbaums in dem Fall bestimmen, dass keine Kopplung vorliegt
- Die Rekombinationswahrscheinlichkeit zwischen zwei nicht gekoppelten Genen ist grundsätzlich $\theta = 0,5$.

$$\theta^k (1 - \theta)^{n-k} = (1/2)^k (1 - 1/2)^{n-k} \quad (1)$$

$$= (1/2)^k (1/2)^{n-k} \quad (2)$$

$$= (1/2)^n \quad (3)$$

- $(1/2)^n$ gilt ungeachtet der Anzahl rekombinanter Individuen!

Iod: Schritt für Schritt (6)

- Iod-Score betimmen
- Die allgemeine Formel lautet:

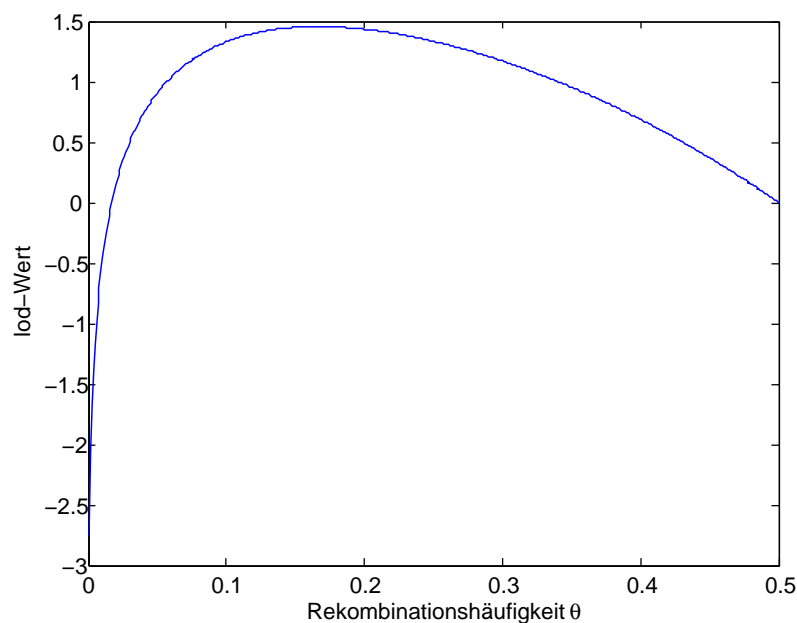
$$Z(\theta) = \log \frac{\theta^k (1 - \theta)^{n-k}}{(1/2)^n}$$

- In unserem Beispiel haben wir $k = 1$, $n = 6$, $\theta = 0.1$, folglich

$$Z(\theta = 0.1) = \log \frac{0.1^1 (1 - 0.1)^{6-1}}{(1/2)^6} = 1.3295$$

Iod: Schritt für Schritt (7)

- Für mehrere Werte von θ wiederholen
- Maximum bestimmen (Schätzer für die genetische Entfernung zum Marker!)



```

recomb=1;
nonrecomb=5;
n=recomb+nonrecomb;

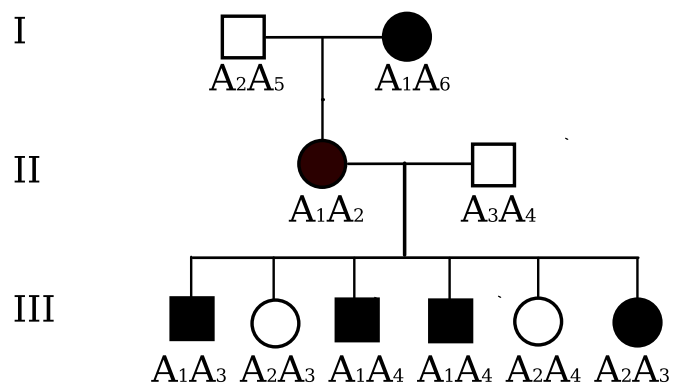
theta=0.001:0.001:0.5;
z=log((theta.^recomb.*(1-theta).^nonrecomb)./(0.5^n));

plot(theta,z);

```

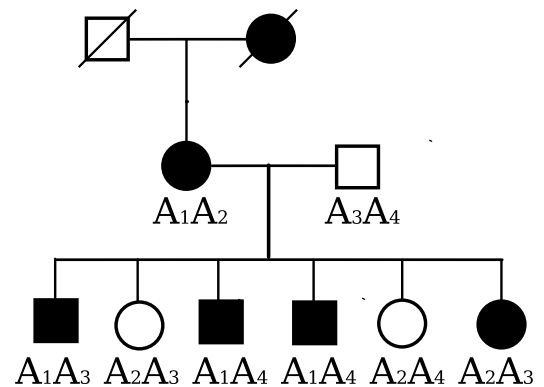
Phasendefiniert

- II_1 hat eindeutig das Allel A_1 von I_2 geerbt
- *phasendefiniert* heißt, man weiss, welche Allele von welchem Elternteil geerbt wurde.



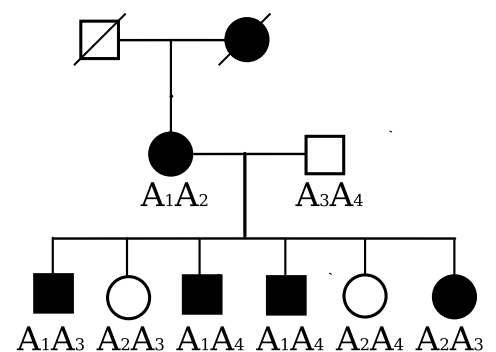
Nicht phasendefiniert

- In diesem Stammbaum können wir nicht sagen, die Mutter II₁ A₁ oder A₂ zusammen mit der Krankheit geerbt hat. Die Phase ist nicht definiert.



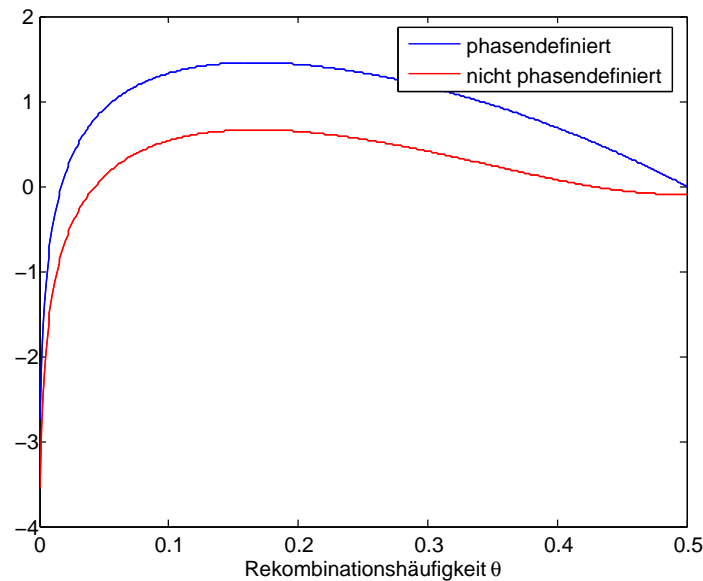
Nicht phasendefiniert

- Falls A₁ mit der Krankheit geerbt worden wäre, lägen 5 nicht rekombinante und 1 rekombinante Meiose vor
- Falls A₂ mit der Krankheit geerbt worden wäre, lägen 1 nicht rekombinante und 5 rekombinante Meiose vor
- Bei gleicher *a priori* Wahrscheinlichkeiten (wie hier) für jede Phase, muss man die entsprechenden Wahrscheinlichkeiten mitteln



Nicht phasendefiniert

$$Z(\theta) = \log \left[\frac{1}{2} \frac{[(1-\theta)^5 \theta]}{(1/2)^n} + \frac{1}{2} \frac{[(1-\theta) \theta^5]}{(1/2)^n} \right]$$



Weitere Komplikationen...

Urheberrechtlich geschütztes Bild entfernt

- Ist bei den Verwandten III₆ und III₇ das Allel A_1 durch gemeinsame Herkunft identisch?
- Oder gab es unter den Großeltern zwei Kopien des Alleles A_1 , wovon nur eines auf dem selber Haplotyp wie das Krankheitsgen lag?
- Wahrscheinlichkeiten hängen von der Populationshäufigkeit des Allels A_1 ab.

lod-Werte +3 und -2

- lod-Wert von +3, d.h., 1000:1, wird als Signifikanzschwelle für eine Kopplung angenommen
- entspricht etwa $p < 0,05$
- bei $z < -2$ ist eine Kopplung auszuschließen
- 1) Kopplung bei $\theta = 0$, 2) Hinweis auf Kopplung bei $\theta \approx 0,23$, 3) Ausschluss einer Kopplung für $\theta < 0,12$, keine weitere Aussage möglich 4) Keine Aussage

Urheberrechtlich geschütztes Bild entfernt

Multiples Testen

- Bei Krankheitsstudien untersucht man Familien so lange auf verschiedene Marker, bis positive lod-Werte vorliegen
- Beim lod-Wert von 3 besteht eine Chance von 0,05, dass es sich um ein falsch positives Ergebnis handelt
- In genomweiten Studien werden mehrere Marker untersucht. Dabei steigt die Wahrscheinlichkeit der falschen Schlussfolgerung
- In der Praxis sind lod-Werte unter 5 als vorläufige Hinweise auf eine Kopplung zu betrachten, unabhängig davon, ob sie mit einem oder mehreren Markern ermittelt wurden.

- Die Multimarker kartierung ist effizienter als die Zweimarkerkartierung
- z.B. kann man Probleme durch Meiosen mit unbekannter Phase durch geeignete Marker vermeiden.
- Algorithmen wesentlich komplizierter als die dargestellte lod-Wert-Methode für Zweimarkerkartierung
- vgl. <http://www.nslij-genetics.org/soft/> für Liste von Programmen für die Kopplungsanalyse

Urheberrechtlich geschütztes Bild entfernt

Autozygotie

- Wenn homozygote Marker in einer Person in ihrer Abstammung übereinstimmen, d.h., das mütterliche und das väterliche Allel kommen beide vom selben Vorfahren, dann redet man von Autozygotie (Identity by descent)
- Blutsverwandte Personen, die an einer autosomal rezessiven Krankheit leiden, sind wahrscheinlich für die mit dem Krankheitslocus gekoppelten Marker autozygot
- z.B. sind die Eltern Kusine und Vetter zweiten Grades, dann ist zu erwarten, dass ihre Gene aufgrund der gemeinsamen Vorfahren zu 1/32 übereinstimmen, ihre Kinder wären zu 1/64 autozygot

Urheberrechtlich geschütztes Bild entfernt

- Die Marker AFMa052yb5 und D2S158 sind bei allen Erkrankten homozygot → Hinweis auf Kopplung an Genort für angeborene Taubheit in dieser Familie

Kopplungsungleichgewicht

- Nijmegen-Chromosomenbruchsyndroms (NBS)
- Seltene AR Krankheit, gekennzeichnet durch Chromosomenbrüche, verringertes Wachstum, Mikrozephalie, Immunschwäche, Krebsanfälligkeit
- Mit Kopplungsanalysen konnte das Gen auf Chromosom 8p21 lokalisiert werden
- Kopplungsintervall umfasste noch immer 8 Mb (zuviel, um alle Gene zu sequenzieren)
- Annahme: Auch nichtverwandte NBS-Patienten haben die Mutation, sowie einen diese umgebenden Chromosomenabschnitt ([Haplotyp](#)) von einem gemeinsamen Vorfahren geerbt

Urheberrechtlich geschütztes Bild entfernt

- Ursprünglicher Haplotyp bei europäischen Patienten mit NBS. Nur bei Loci 11 und 12 gibt es keine rekombinanten Allele → Hinweis auf Lokalisation des NBS-Gens

Parametrische Modelle

- Die lod-Wert-Analyse erfordert ein genaues genetisches Modell
 - ① Vererbungsmodus
 - ② Genhäufigkeiten in Population
 - ③ Penetranz
- Daher reden wir von der **parametrischen** Kopplungsanalyse
- Dies ist vor allem für mendelnde (monogenetische) Krankheiten angebracht
- Nichtparametrische Analysen für komplexe Krankheiten (nächstes Mal)

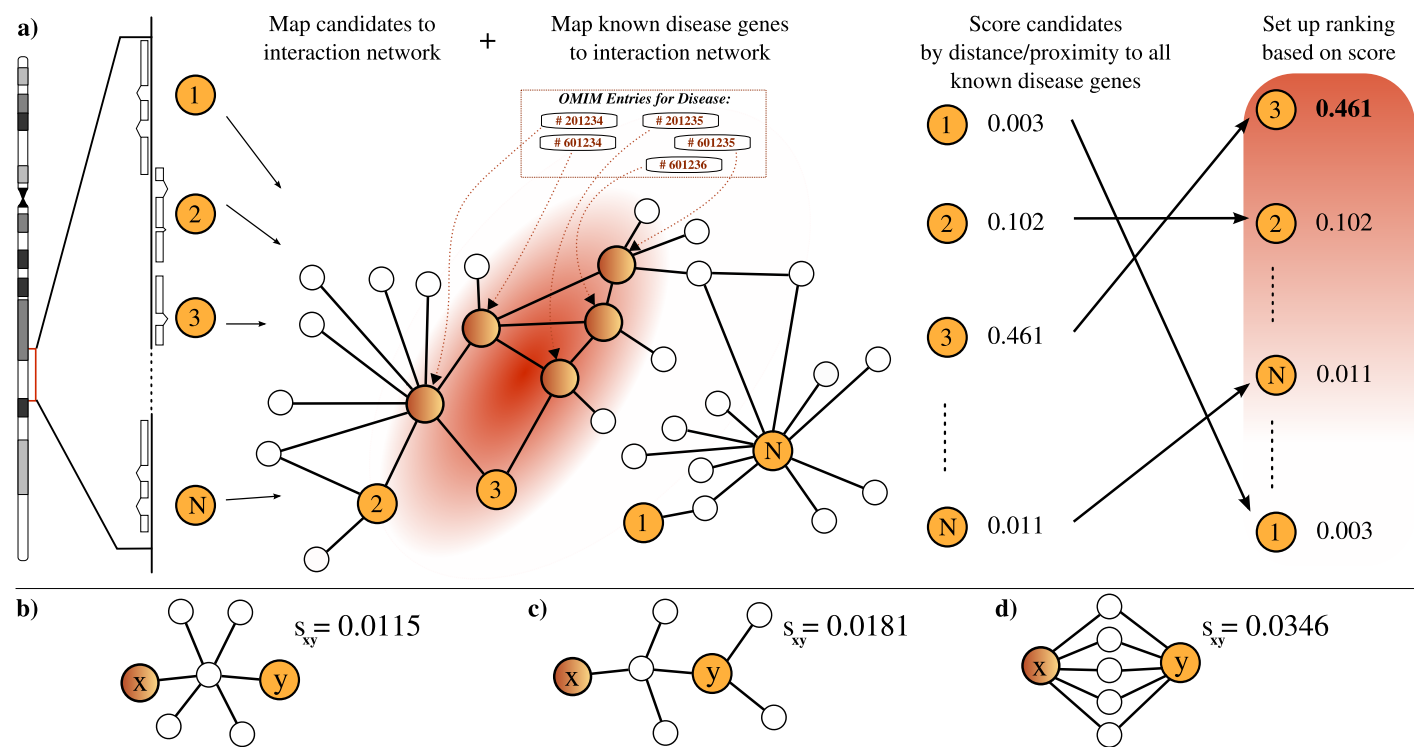
Prioritisierung

- (Mai 2008): ca. 2800 genetisch bedingte Krankheiten mit identifizierten Genen
- ca. 3600 genetisch bedingte Krankheiten, bei denen das Krankheitsgen noch unbekannt ist
- Identifizierung von Krankheitsgenen verbessert die medizinischen Betreuungsmöglichkeiten für betroffene Familien sowie unser Verständnis von Genfunktionen und zellulärer (Patho)physiologie
- Daher ein wichtiges Ziel vieler Forschungsgruppen weltweit

Prioritisierung

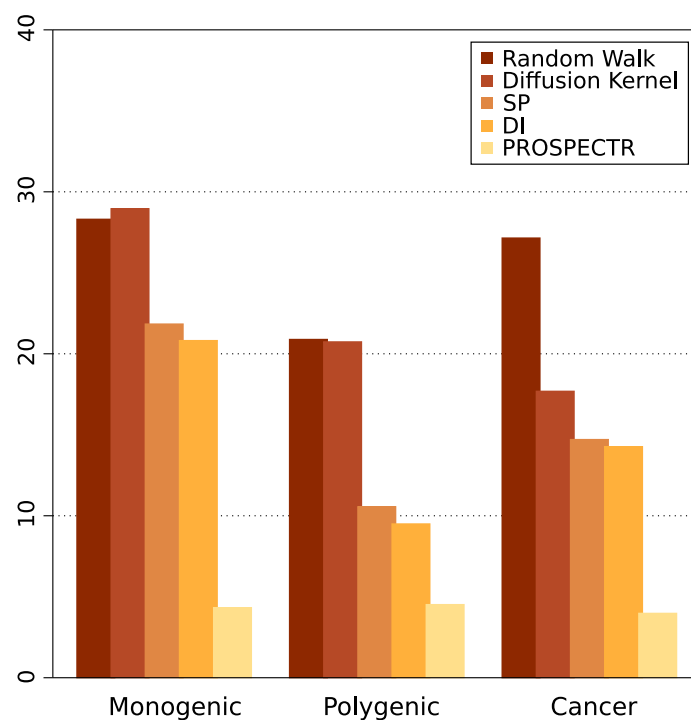
- Ein typisches Kopplungsintervall enthält 100–300 Gene
- Teuer, zeitaufwändig
- Viele bioinformatische Methoden entwickelt worden, um eine Prioritätsliste innerhalb der Gene des Kopplungsintervalls zu erstellen
- s. hierzu Strachan & Read: Kapitel 14

Beispiel: GeneWanderer



Köhler S et al. (2008) Walking the interactome for prioritization of candidate disease genes. Am J Hum Genet. 82:949–58.

Beispiel: GeneWanderer

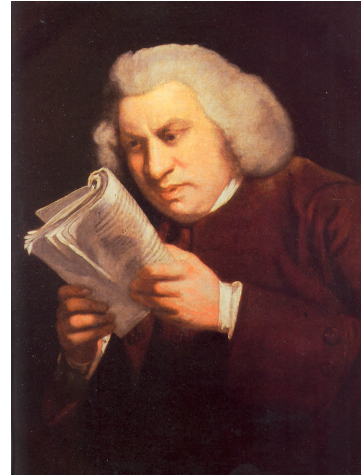


Köhler S et al. (2008) Walking the interactome for prioritization of candidate disease genes. Am J Hum Genet. 82:949–58.

The End of the Lecture as We Know It

- Strachan & Read Kapitel 13
- Diese Vorlesungsdias stehen unter der GNU-Lizenz für freie Dokumentation^b
- Kontakt:
peter.robinson@charite.de

^b<http://www.gnu.org/licenses/fdl.txt>



Lectures were once useful; but now, when all can read, and books are so numerous, lectures are unnecessary. If your attention fails, and you miss a part of a lecture, it is lost; you cannot go back as you do upon a book... People have nowadays got a strange opinion that everything should be taught by lectures. Now, I cannot see that lectures can do as much good as reading the books from which the lectures are taken. I know nothing that can be best taught by lectures, except where experiments are to be shown. You may teach chymistry by lectures. You might teach making shoes by lectures!

Samuel Johnson, quoted in Boswell's Life of Johnson (1791).