

# Genetische Kartierung und lod-Score-Analyse

Peter N. Robinson

Institut für medizinische Genetik  
Charité Universitätsmedizin Berlin

9. Dezember 2014

# Outline

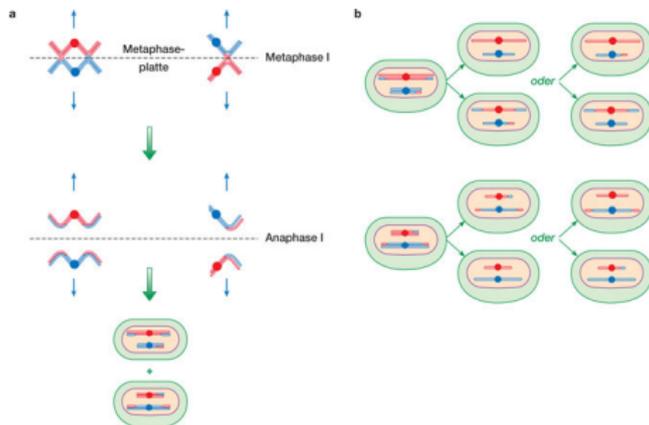
- 1 Rekombination
- 2 Zweimarkerkartierung
- 3 Feinkartierung
- 4 Bioinformatische Vorhersage von Krankheitsgenen

# 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

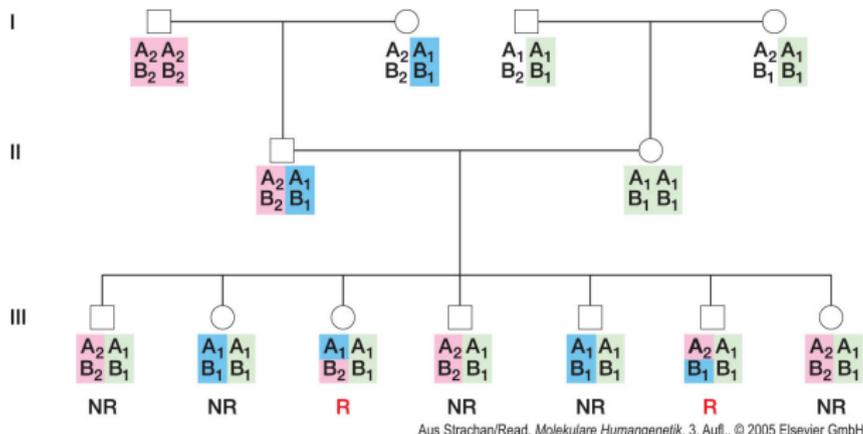
# Crossing-over (Rückblick)

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



Aus Strachen/Reed, Molekulare Humangenetik, 3. Aufl., © 2005 Elsevier GmbH

# Rekombinaten und Nichtrekombinanten



- 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.

# Rekombinationswahrscheinlichkeit

- 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

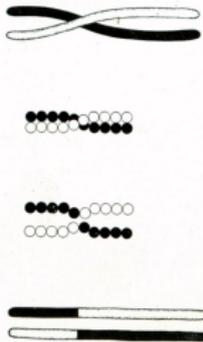


FIG. 64. Scheme to illustrate a method of crossing over of the chromosomes.

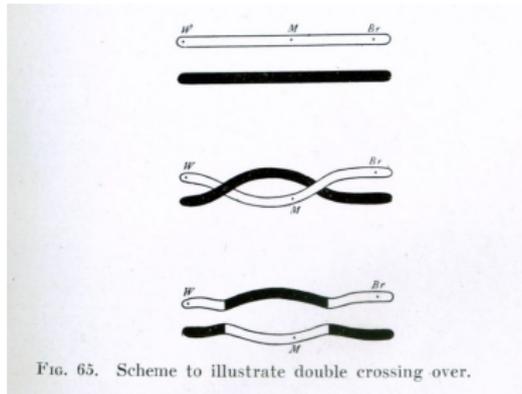


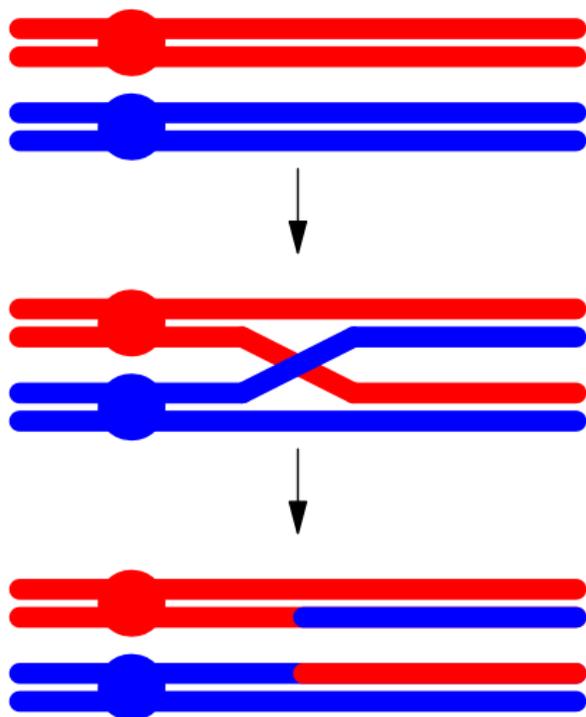
FIG. 65. Scheme to illustrate double crossing over.

Thomas Hunt Morgan's A Critique of the Theory of Evolution  
(1916), Wikipedia Commons

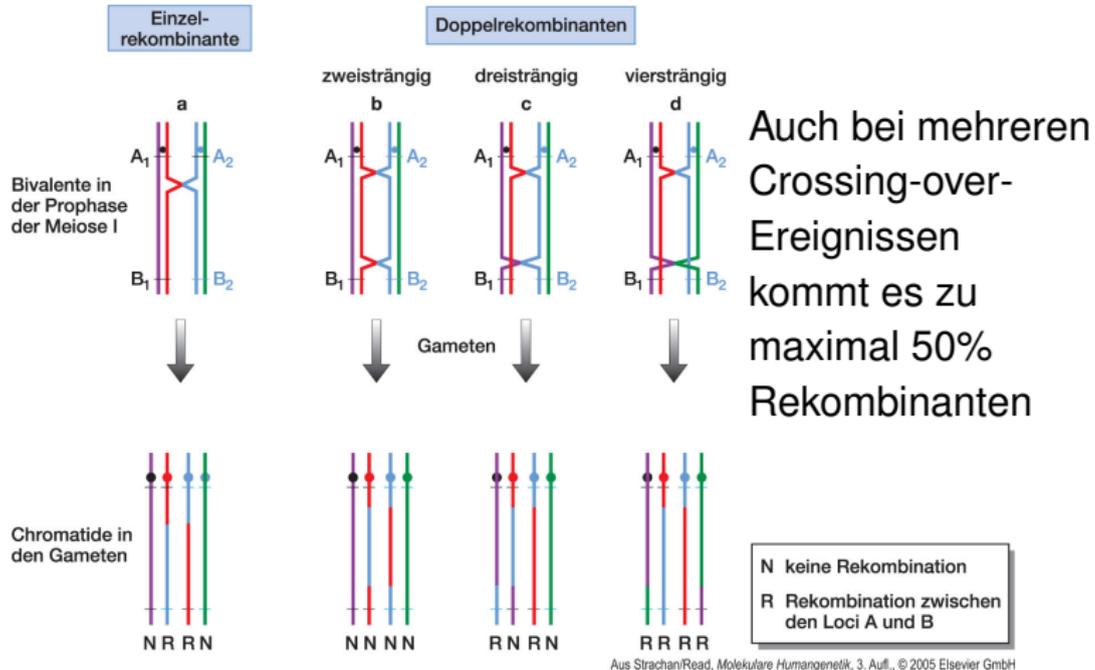
- Bei größeren Abständen muß mit mehrfachen Crossovers zwischen zwei Loci gerechnet werden
- Dieses Schema von Morgan ist zu simpel, da insgesamt vier Chromatiden (b.w)

# Rekombination

- Bei einem einzigen Crossing-Over entstehen zwei rekombinante und zwei nichtrekombinante Chromosomen
- Daher entspricht die Rekombinationshäufigkeit  $\theta$  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

**Abb. 13.2 Einzel und Doppelrekombinanten.** Jedes Crossing-over betrifft zwei oder vier Chromatiden der beiden gepaarten homologen Chromosomen. Ein Chromosom trägt die Allele A<sub>1</sub> und B<sub>1</sub> an zwei verschiedenen Loci, ein anderes die Allele A<sub>2</sub> und B<sub>2</sub>. a) Ein einzelnes Crossing-over führt zu zwei rekombinanten und zwei nichtrekombinanten Chromatiden (50% Rekombinanten). b) Die drei Arten eines doppelten Crossing-over treten in einer zufälligen Verteilung auf, sodass doppelte Crossing-over durchschnittlich 50% Rekombinanten verursachen.

# 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**:

$$\theta = m$$

- $\theta$  = 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, indem 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  $e^{-2m}$ .<sup>1</sup>

---

$$^1 p(n=0) = \frac{\lambda^n e^{-\lambda}}{n!} = e^{-\lambda}.$$

# Haldane-Funktion

- $m$ : Abstand auf der Genkarte in Morgan
- $\theta$ : *Beobachtete* Rekombinationshäufigkeit

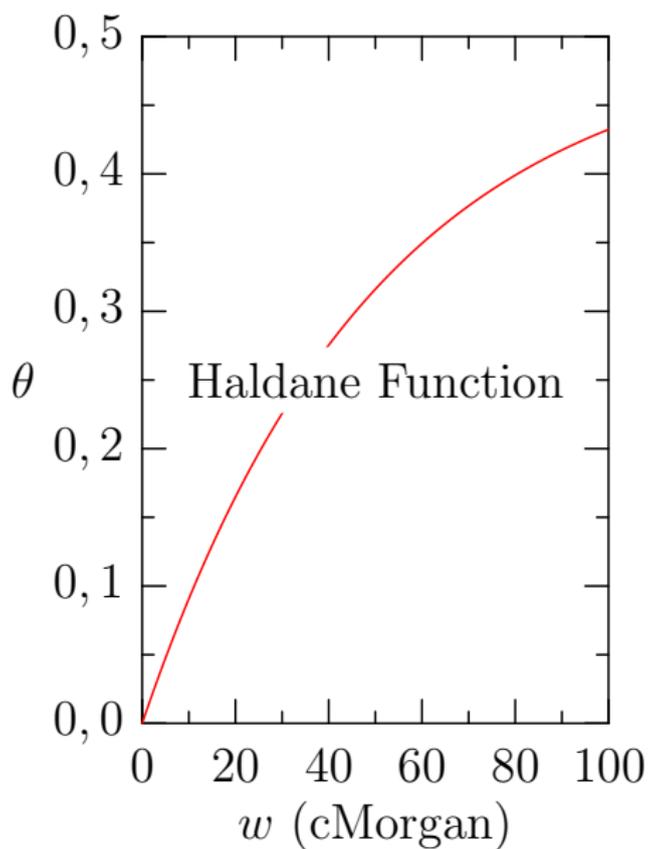
$$\theta = \frac{1}{2} [1 - e^{-2m}]$$

- So variiert die Rekombinationsfraktion zwischen 0 und 0,5.<sup>1</sup>

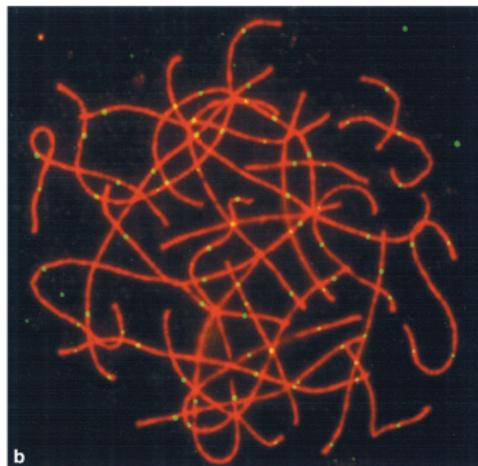
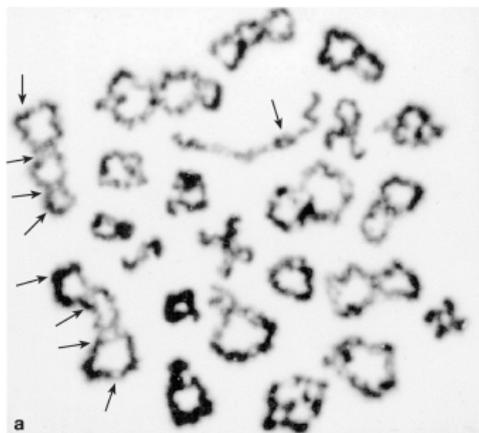
---

<sup>1</sup>Ist  $e^{-2m}$  die Wahrscheinlichkeit dass kein Crossing-Over auftritt, so ist  $[1 - e^{-2m}]$  die Wahrscheinlichkeit dass mindestens eins auftritt. Der Faktor  $1/2$  ist erforderlich, da für jedes Crossing-over lediglich 2 der 4 Stränge und somit im Schnitt jeder zweite Nachkomme betroffen ist.

# Haldane-Funktion



# Gesamtlänge einer genetischen Karte



- 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

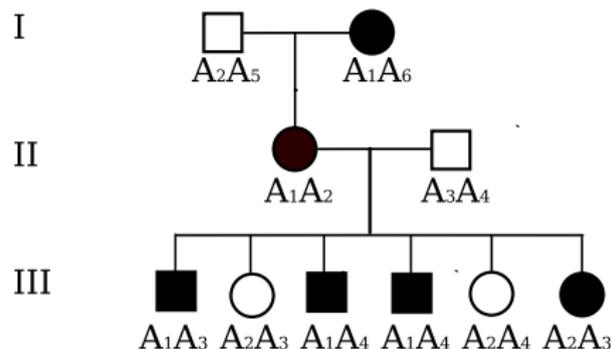
- Faustregel:  $1 \text{ cM} \approx 1 \text{ Mb} = 10^6 \text{ Nukleotide}$
- Variiert stark (ca. 0,3 cM pro Mb bis über 3 cM pro Mb)

# Outline

- 1 Rekombination
- 2 Zweimarkerkartierung**
- 3 Feinkartierung
- 4 Bioinformatische Vorhersage von Krankheitsgenen

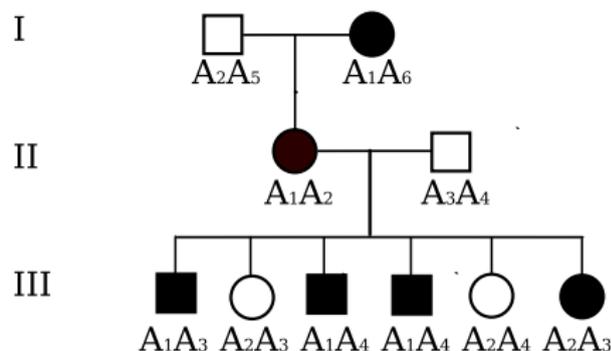
# Stammbaumanalyse

- Schwarz: 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.

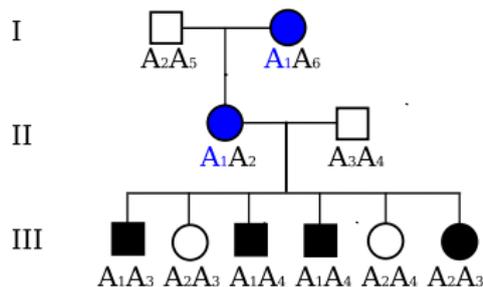


# Iod-Wert-Analyse

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

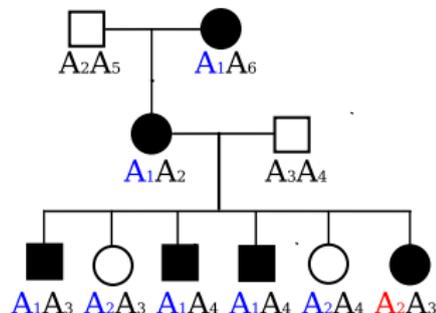
# Iod: Schritt für Schritt (1)

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



## 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)



## lod: Schritt für Schritt (3)

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

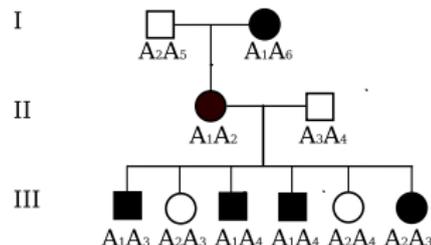
# Iod: Schritt für Schritt (4)

- 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!

## lod: Schritt für Schritt (6)

- lod-Score bestimmen
- Die allgemeine Formel lautet:

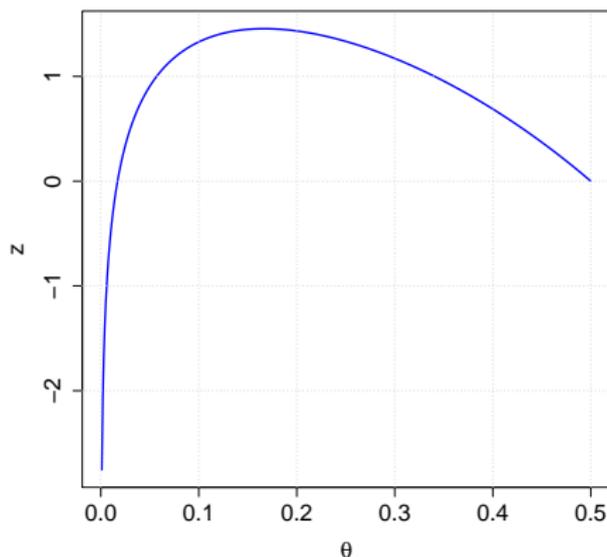
$$Z(\theta) = \log_{10} \frac{\theta^k (1 - \theta)^{n-k}}{(1/2)^n}$$

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

$$Z(\theta = 0.1) = \log_{10} \frac{0.1^1 (1 - 0.1)^{6-1}}{(1/2)^6} = 1.3295$$

## Iod: Schritt für Schritt (7)

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



# R

```
recomb <- 1
nonrecomb <- 5
n <- recomb + nonrecomb

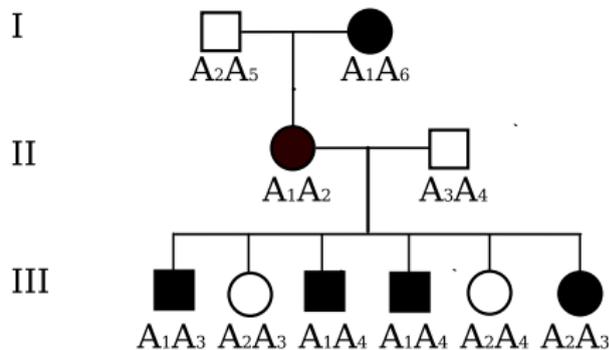
theta <- seq(from=0.001, to=0.5, by=0.001)

z <- log( (theta^recomb *(1-theta)^nonrecomb) / 0.5^n)

plot(theta, z, type='l',
      col="blue",
      lwd=2,
      cex.axis=1.5,
      cex.lab=1.5,
      xlab=expression(theta),
      ylab="z")
grid()
```

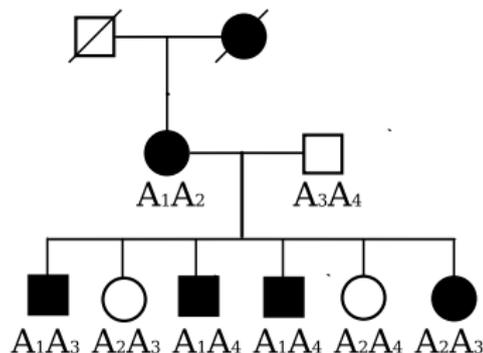
# Phasendefiniert

- $II_1$  hat eindeutig das Allel  $A_1$  von  $I_2$  geerbt
- *phasendefiniert*: man weiß, welches Allel von welchem Elternteil geerbt wurde.



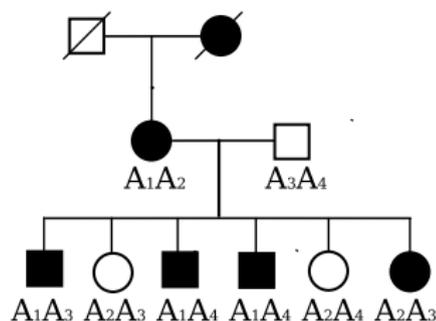
# Nicht phasendefiniert

- In diesem Stammbaum können wir nicht sagen, die Mutter II<sub>1</sub> A<sub>1</sub> oder A<sub>2</sub> zusammen mit der Krankheit geerbt hat. Die Phase ist nicht definiert.



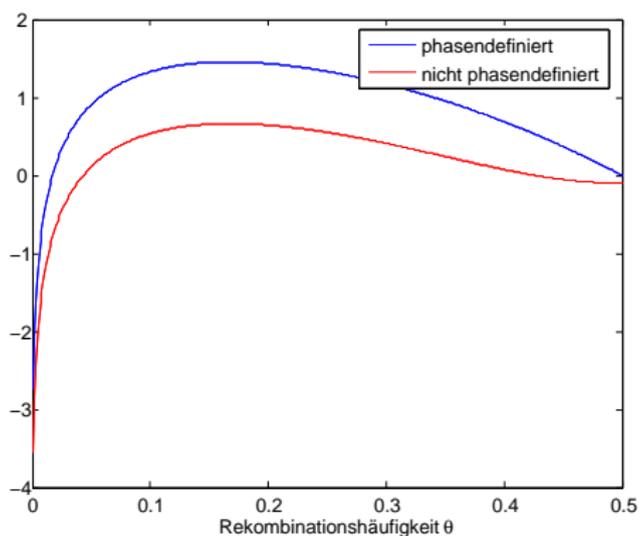
# Nicht phasendefiniert

- Falls  $A_1$  mit der Krankheit geerbt worden wäre, lägen 5 nicht rekombinante und 1 rekombinante Meiose vor
- Falls  $A_2$  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

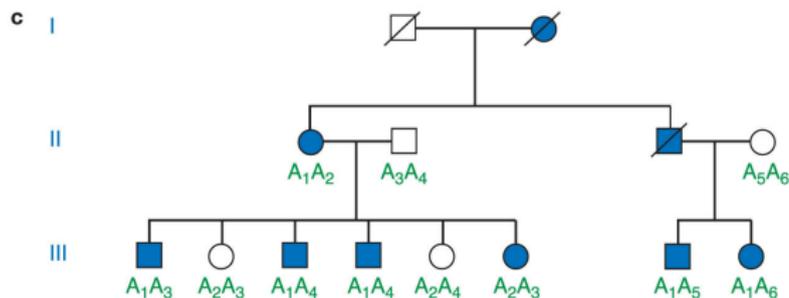


# Nicht phasendefiniert

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



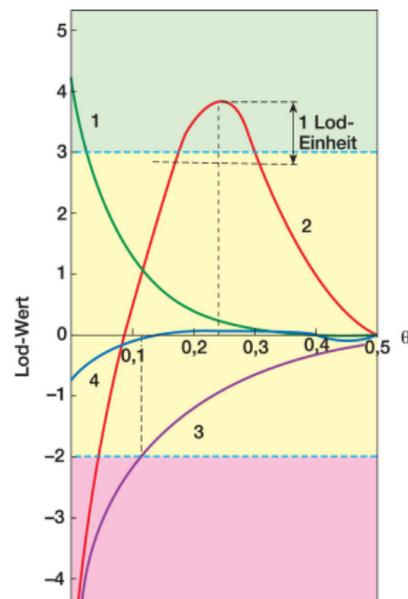
# Weitere Komplikationen...



- Ist bei den Verwandten III<sub>6</sub> und III<sub>7</sub> das Allel A<sub>1</sub> durch gemeinsame Herkunft identisch?
- Oder gab es unter den Großeltern zwei Kopien des Alleles A<sub>1</sub>, wovon nur eines auf dem selben Haplotyp wie das Krankheitsgen lag?
- Wahrscheinlichkeiten hängen von der Populationshäufigkeit des Alleles A<sub>1</sub> 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

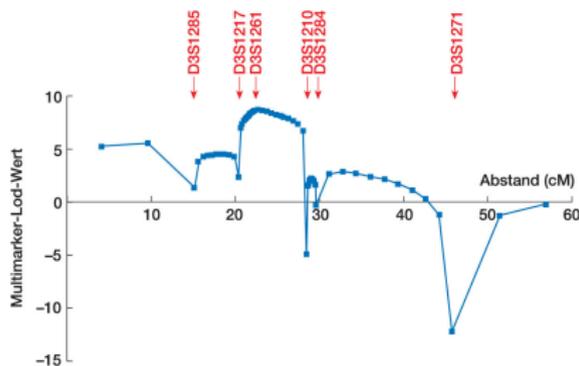


# Multiples Testen

- 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.

# Multimarkerkartierung

- Die Multimarkerkartierung 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.nslj-genetics.org/soft/> für Liste von Programmen für die Kopplungsanalyse



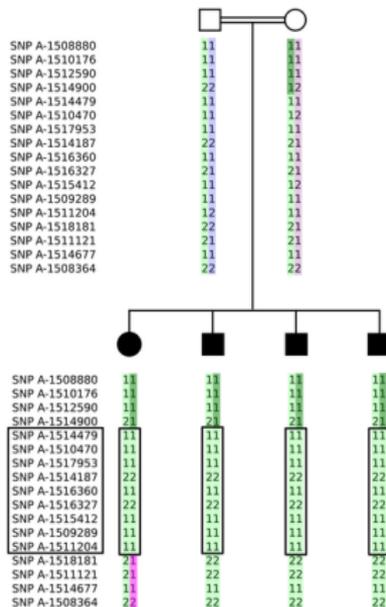
# Outline

- 1 Rekombination
- 2 Zweimarkerkartierung
- 3 Feinkartierung**
- 4 Bioinformatische Vorhersage von Krankheitsgenen

# 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

# Autozygotiekartierung: Beispiel



- Die eingekästelten Marker sind bei allen Erkrankten homozygot → Hinweis auf Kopplung an Genort für Ataxie in dieser Familie

# Homozygosity-Mapping

- Unter Inzucht (inbreeding) versteht man im Allgemeinen die Paarung relativ naher Blutsverwandter
- Das Maß für die Inzucht ist der Inzuchtkoeffizient.
- Der Inzuchtkoeffizient (abgekürzt IK, oft auch COI von engl. Coefficient of Inbreeding) gibt die Wahrscheinlichkeit an, dass an einem Locus beide Allele vom selben Vorfahr stammen.

# Inbreeding coefficient

Um den Inzuchtkoeffizienten bei  $i$  gemeinsamen Vorfahren zu berechnen, verwendet man die Formel nach Wright

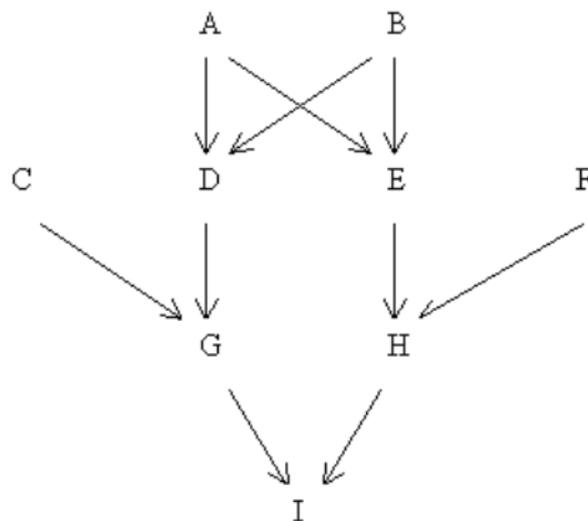
$$F_I = \sum_i \left(\frac{1}{2}\right)^{n_1+n_2+1} \quad (4)$$

Hierbei gilt

- $n_1$ : Anzahl der Generationen vom Vater zum gemeinsamen Ahnen
- $n_2$  = Anzahl der Generationen von der Mutter zum gemeinsamen Ahnen
- Haben diese gemeinsamen Vorfahren selber gemeinsame Vorfahren, ist ein weiterer Korrekturfaktor erforderlich

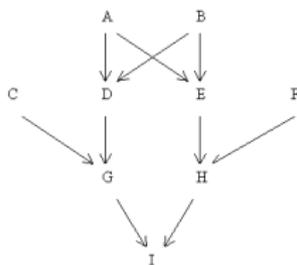
# Inbreeding coefficient

- $F$  ist daher die Wahrscheinlichkeit dass ein Genlocus “homozygot durch Abstammung” (homozygous by descent, HBD) ist
- $F$  gibt daher den erwarteten HBD-Anteil des Genoms an



Aber wie berechnen wir  $F_I$  für Person I?

# Inbreeding coefficient



- Die einzigen gemeinsamen Vorfahren der Eltern sind Individuen A und B
- Für A:  $n_1 = A \rightarrow D \rightarrow G = 2$ ,  $n_2 = A \rightarrow E \rightarrow H = 2$
- Für B:  $n_1 = B \rightarrow D \rightarrow G = 2$ ,  $n_2 = B \rightarrow E \rightarrow H = 2$

$$F_I = \sum_i \left(\frac{1}{2}\right)^{n_1+n_2+1} = \left(\frac{1}{2}\right)^5 + \left(\frac{1}{2}\right)^5 = \frac{1}{16} \quad (5)$$

# Homozygosity-Mapping

- Wir betrachten eine autosomal rezessive Mendelsche Erkrankung, bei der Mutationen eine Allelfrequenz von  $q$  in der Population aufweisen
- Ist ein Kind aus einer konsanguinen Beziehung erkrankt, dann liegt Homozygotität (HBD) im Bereich des Krankheitsgens mit großer Wahrscheinlichkeit vor
- Warum?  $Fq$  der betroffenen Personen sind HBD im Bereich des Krankheitsgens
- Bei  $(1 - F)q^2$  der Betroffenen liegt im Bereich des Krankheitsgenes keine Homozygotität vor, sondern zwei Mutationen aus der Population ( $q^2$ ) trafen per Zufall aufeinander
- Die Wahrscheinlichkeit einer HBD ist demnach

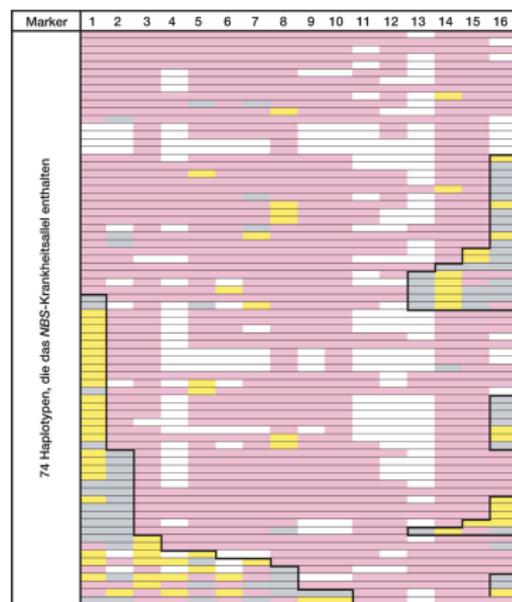
$$\alpha = \frac{Fq}{Fq + (1 - F)q^2} \approx 1 \quad (6)$$

# Homozygosity-Mapping

- Die Region um das gesuchte Krankheitsgen ist daher HBD mit einer Wahrscheinlichkeit von  $\alpha \approx 1$
- Ungekoppelte Regionen sind HDB mit einer Wahrscheinlichkeit von  $F \ll 1$
- Beobachten wir Homozygotie in einem hochpolymorphen (informativen) Marker, können wir daher das Verhältnis der Wahrscheinlichkeiten Kopplung vs. Keine-Kopplung als  $\alpha : F \approx 1 : F$  annehmen.
- Wie wir gesehen haben, beträgt  $F_I = \frac{1}{16}$  für Kusinenehen.
- Sind drei betroffenen Kinder aus einer Kusinenehe homozygot durch Abstammung (HBD) in einem bestimmten Genlocus, dann beträgt die Wahrscheinlichkeit einer Kopplung etwa  $(16 : 1)^3 = 4096 : 1$ , d.h., besser als lod=3.

# 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



Aus Strachan/Read, Molekulare Humangenetik, 3. Aufl., © 2005 Elsevier GmbH

- 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)

# Outline

- 1 Rekombination
- 2 Zweimarkerkartierung
- 3 Feinkartierung
- 4 Bioinformatische Vorhersage von Krankheitsgenen**

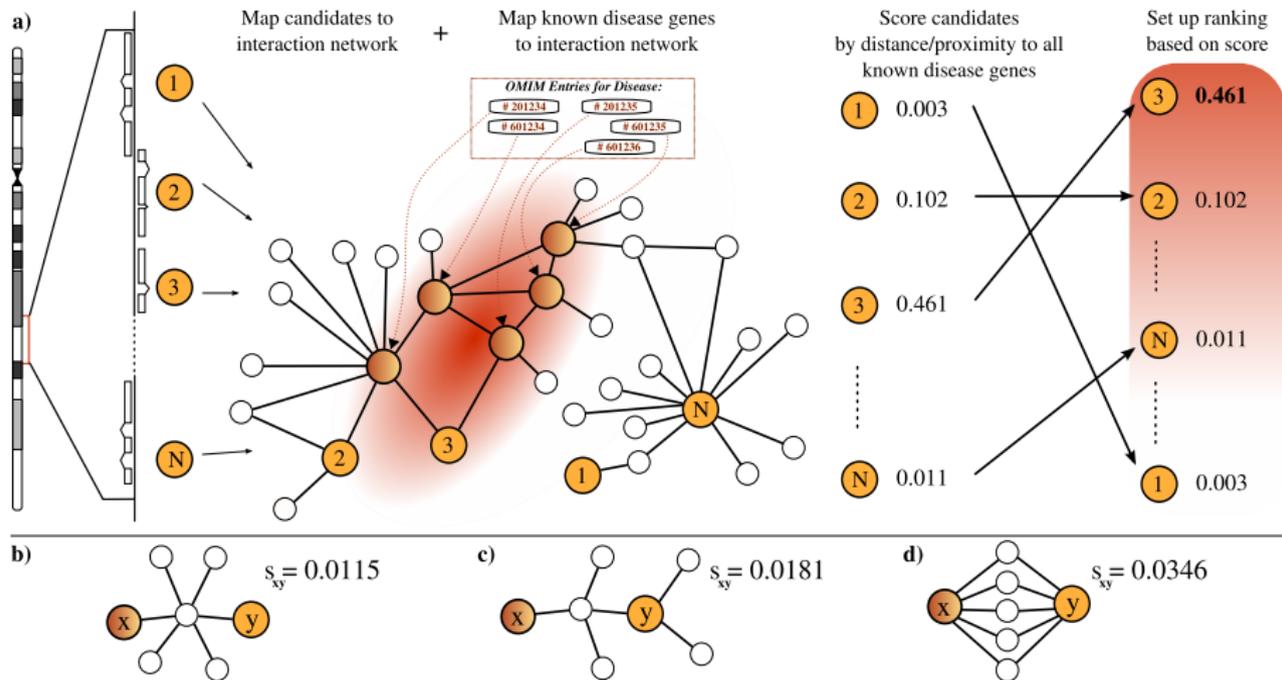
# Priorisierung

- (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

# Priorisierung

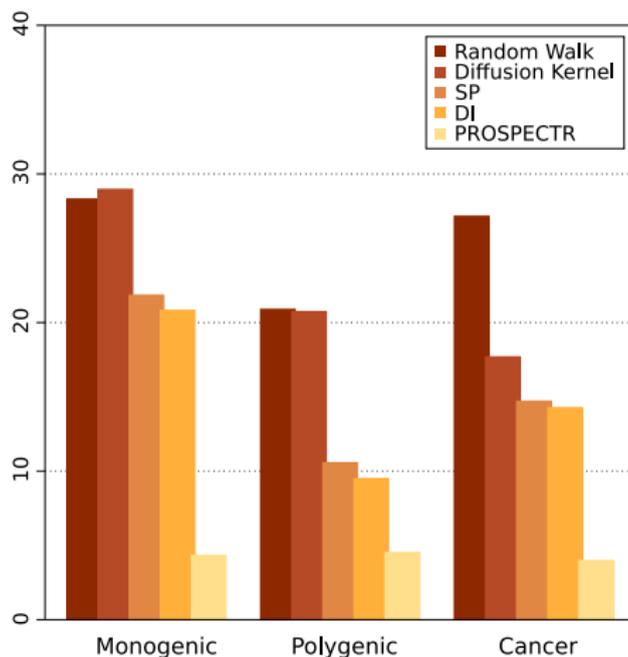
- Ein typisches Kopplungsintervall enthält 100–300 Gene
- Teuer, zeitaufwändig
- Viele bioinformatische Methoden entwickelt worden, um eine Prioritätsliste innerhalb der Gene des Kopplungsintervalls zu erstellen
- s. hierzu Strachan & Read: Kapitel 14

# 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
- Kontakt:  
peter.robinson@charite.de



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).