

Genetische Epidemiologie

**Prof. DI Dr. Andrea Berghold
Institut für Med. Informatik, Statistik
und Dokumentation**

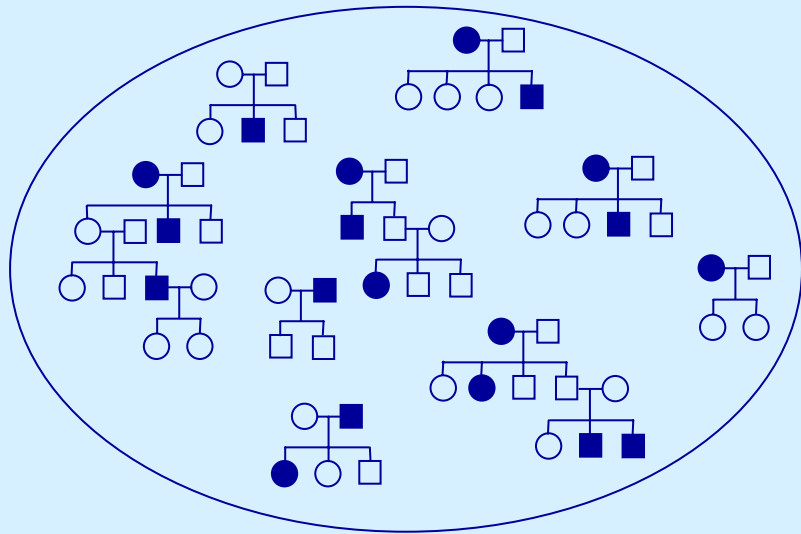
Definition

„A science that deals with etiology, distribution, and control of disease in groups of relatives, and with inherited causes of disease in populations.“

N.E. Morton and C.S. Chung (1978)

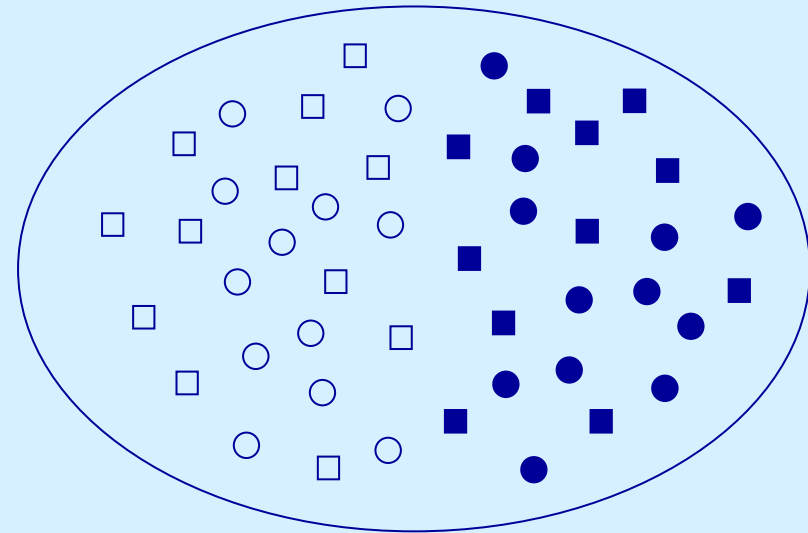
Untersuchte Population

- **Familienstudien**

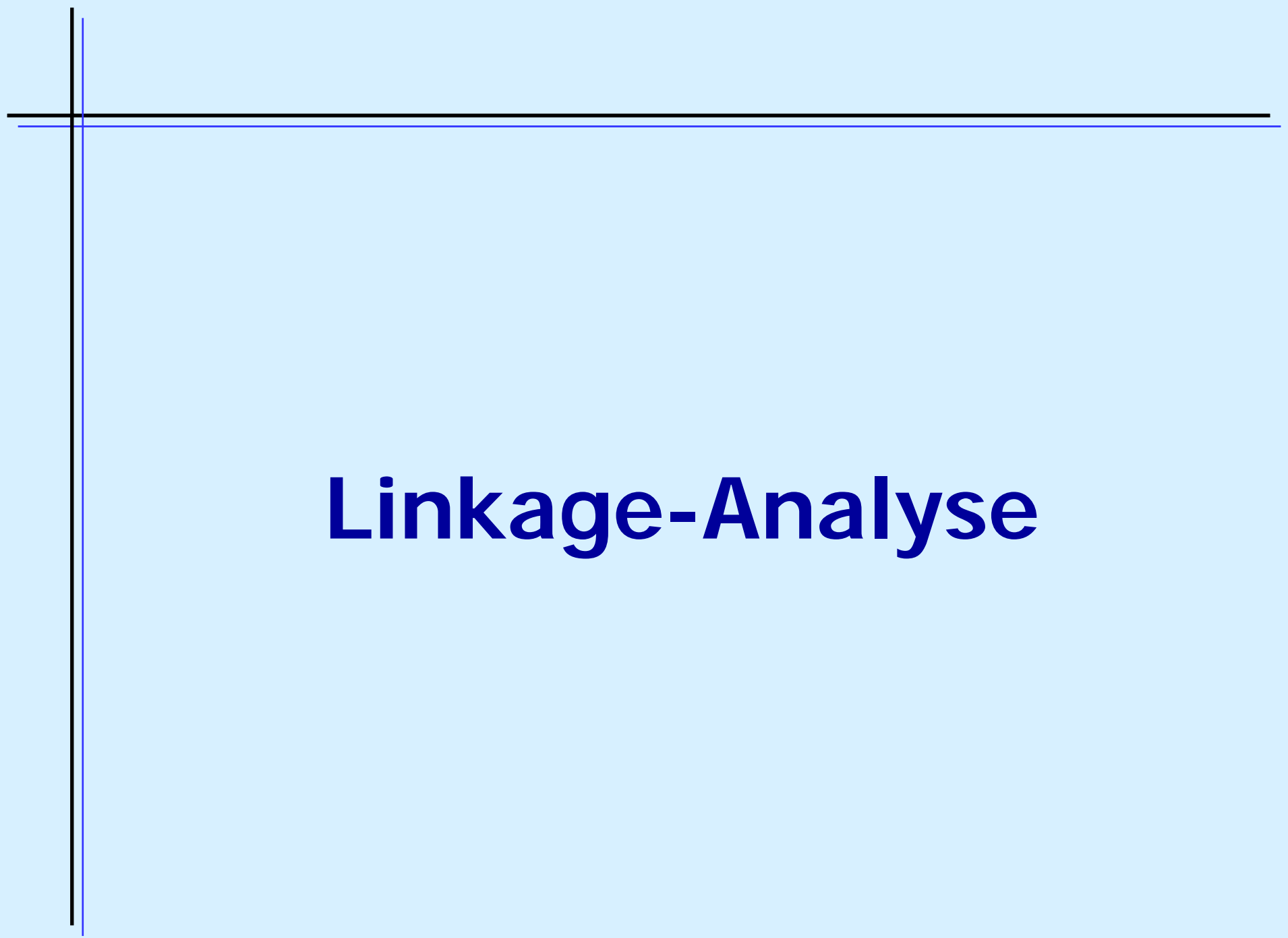


- Beobachtungseinheit: „Familie“
- z.B. Segregationsanalysen, Linkage-Analysen, Geschwister-Paar Studien, Zwillingsstudien

- **Populationsstudien**



- Beobachtungseinheit: Individuen
- z.B. epidemiologische Studien: Fall-Kontroll Studien, Kohortenstudien



Linkage-Analyse

- ◆ Ziele von Linkage-Analysen (Kopplungsanalyse)
- ◆ Prinzip von Linkage-Analysen
- ◆ LOD Score Analyse
- ◆ Interpretation

Ziele von Linkage-Analysen

- ◆ Identifikation und Lokalisation neuer Krankheitsgene
- ◆ Untersuchung genetischer Heterogenität von Erkrankungen

Was ist Linkage?

Die unabhängige Vererbung von Genen an verschiedenen Loci ist **nicht allgemein gültig!**

Gilt nur

- für Gene, die auf verschiedenen Chromosomen liegen
- für Gene, die auf dem gleichen Chromosom (in Syntänie) „weit genug“ voneinander entfernt liegen – keine Kopplung

Es gilt

Gene, die auf einem Chromosom nahe zusammen liegen, werden überzufällig häufig gemeinsam vererbt - **Kopplung** oder **Linkage**

Was ist Linkage?

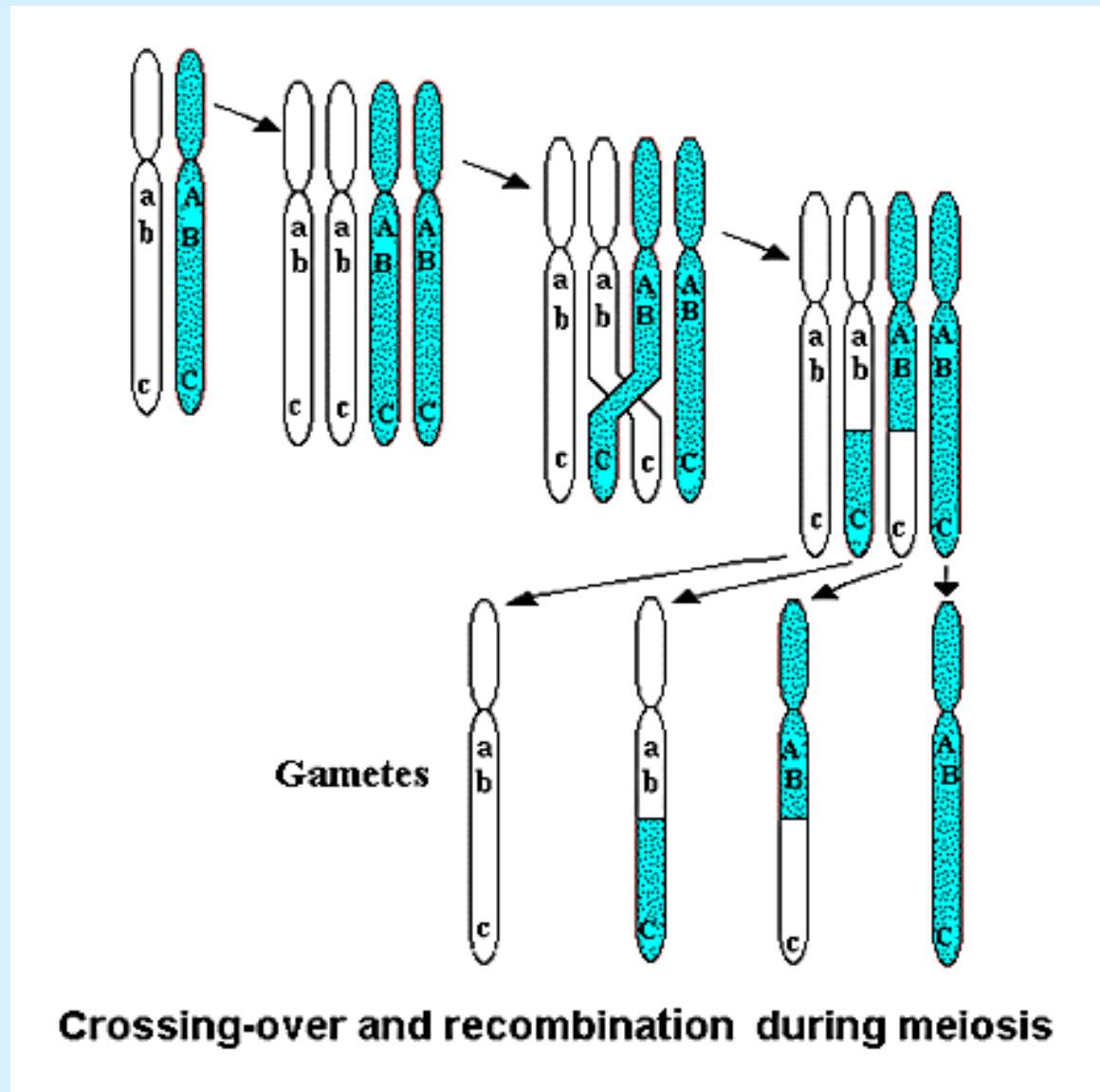
Gametenbildung - ideal

- ◆ Liegen zwei Gene auf unterschiedlichen Chromosomen, ist die Wahrscheinlichkeit gemeinsamer Vererbung = $\frac{1}{2}$
- ◆ Liegen zwei Gene auf demselben Chromosom, ist die Wahrscheinlichkeit gemeinsamer Vererbung im Idealfall = 1

Rekombination

- ◆ Bei der Gametenbildung werden Allele auf demselben Chromosom aber durch Rekombination nicht immer miteinander vererbt.
- ◆ Die relative Häufigkeit dieses „getrennten“ Vererbens nennt man Rekombinationsfrequenz (-fraktion) (θ).

Crossing-Over und Rekombination



Rekombination

Die Rekombinationsfrequenz (zwischen 2 Loci) ist definiert als die Wahrscheinlichkeit, dass eine Gamete rekombinant ist (bezogen auf die beiden Loci)

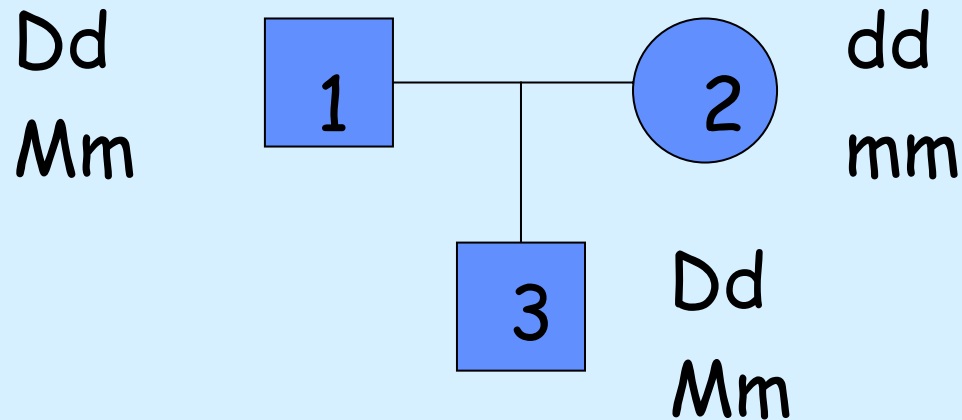
Rekombination

- ◆ Wahrscheinlichkeit θ , dass zwei Allele „getrennt“ vererbt werden, ist je größer, je weiter Genorte auf Chromosom voneinander entfernt liegen.
- ◆ Maximal = $\frac{1}{2}$ (wie wenn sie auf zwei verschiedenen Chromosomen liegen)
- ◆ Aufgrund von Rekombinationsfraktion θ werden Entfernungen zwischen Genorten geschätzt und die Genorte so auf den Chromosomen lokalisiert.

Markereigenschaften zur Verwendung in Kopplungsanalysen:

- ◆ **Polymorph:** d.h. Marker sollte bei 2 zufällig ausgewählten, nicht verwandten Personen mit großer Wahrscheinlichkeit verschieden sein (erleichtert Familienzuordnung seiner Herkunft)
- ◆ **Informativ:** d.h. ein Elternteil muss an der Stelle des Markers heterozygot sein
- ◆ z.B. HLA System, ABO Blutgruppensystem, Minisatelliten, Microsatelliten, SNPs

2 Loci: Linkage und Rekombination



M m

3 erzeugt
Gameten in
Anteilen von:

D	$(1-\theta)/2$	$\theta/2$	$1/2$
d	$\theta/2$	$(1-\theta)/2$	$1/2$
	$1/2$	$1/2$	

Rekombinationsfraktion/-frequenz (θ)

- ◆ $\theta = 1/2$: keine Kopplung,
freie Rekombination (Mendel)
- ◆ $\theta < 1/2$: gekoppelte Loci
- ◆ $\theta = 0$: sehr enge Kopplung
(keine Rekombination)
- ◆ In 3: D-M and d-m bezeichnet man als **parentale (nicht-rekombinante) Haplotypen**, D-m and d-M rekombinante Haplotypen

Linkage-Analyse

- ◆ Untersuchung, ob Krankheitsgen und gewisse Marker vermehrt gemeinsam vererbt werden - Bewertung der relativen Position von 2 oder mehreren Genorten = besteht Linkage?
- ◆ Familienstudie: Stichprobe von Stammbäumen von Familien
- ◆ LOD-Score Analyse – statistische Analyse dieser Familiendaten

Liegt Linkage vor?

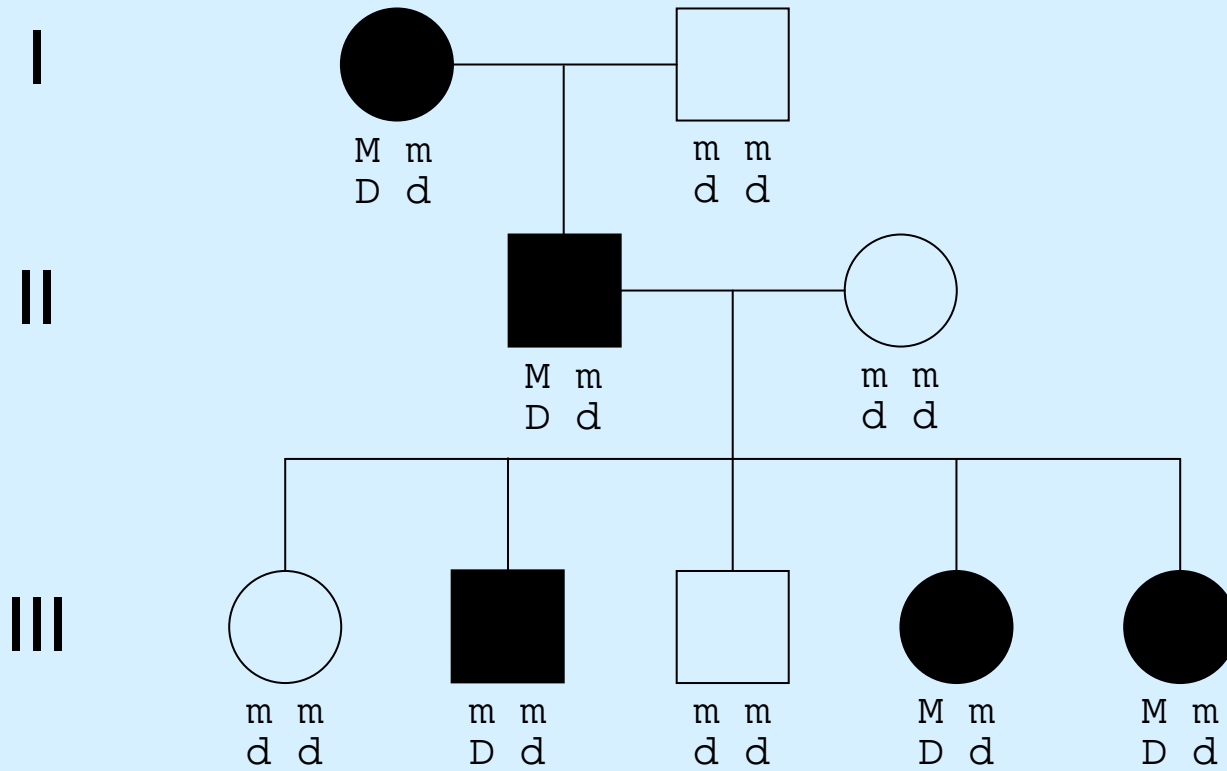
- ◆ Man ist also interessiert an der Rekombinationshäufigkeit θ zwischen Marker und vermuteten Krankheitsgen

Schätzung der Rekombinationsfraktion θ

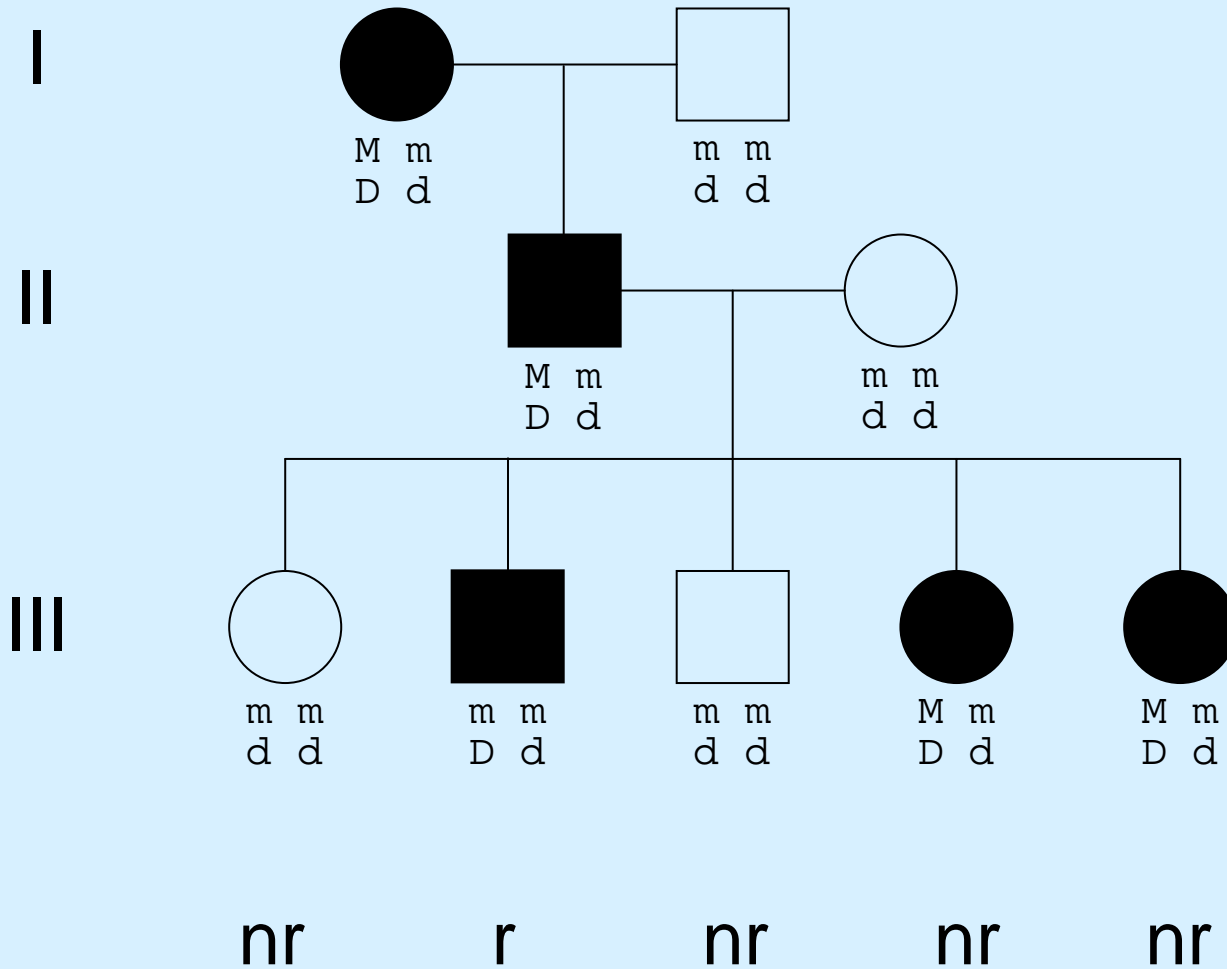
- ◆ Im speziellen möchte man wissen, ob $\theta < \frac{1}{2}$ ist (also ob Kopplung vorliegt)

Test der Hypothese $\theta = \frac{1}{2}$

Beispiel



Beispiel



LOD-Score Methode

- ◆ Zähle die Rekombinationsereignisse für eine gegebene Familie
- ◆ Likelihood Funktion für einen Stammbaum:

$$L(\theta) = \theta^r (1 - \theta)^{nr}$$

r: Zahl der Rekombinanten;
nr: Zahl der Nicht-Rekombinanten
 $n = r + nr$

LOD-Score Methode

◆ Bilde "Likelihood Ratio"

$$\text{LR} = \frac{\text{Likelihood Daten zu beobachten, wenn Loci bei einem gewissen } \theta \text{ gekoppelt sind}}{\text{Likelihood Daten zu beobachten, wenn Loci nicht gekoppelt sind } (\theta = 0,5)}$$

$$\text{LR} = \frac{L(\theta)}{L(\theta = 0,5)}$$

LOD-Score Methode

- ◆ Berechne den LOD-Score:

$$\text{LOD} = \log_{10} (L(\theta) / L(\theta = 0,5))$$

$$\text{LOD} = \log_{10} \left(\frac{(\theta)^r (1 - \theta)^{nr}}{(1/2)^n} \right)$$

$$= r \log_{10} \theta + nr \log_{10} (1 - \theta) + n \log_{10} 2$$

- ◆ Dasjenige θ , welches den LOD-Score maximiert (Maximum-Likelihood Schätzung), gibt die beste Schätzung für die Rekombinationsfraktion

LOD-Score Beispiel

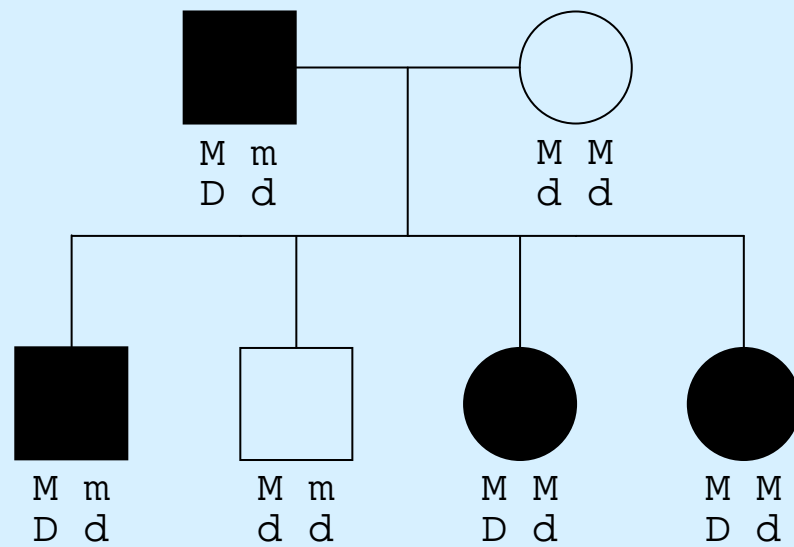
$$\text{LOD} = \log_{10} \left(\frac{\theta^1 (1 - \theta)^4}{(1/2)^5} \right)$$

$$\text{LOD} = 0,41$$

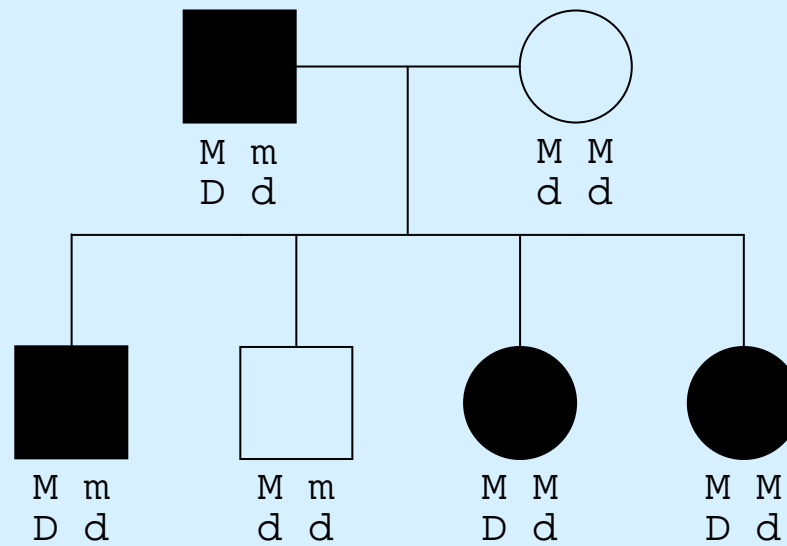
LOD-Score Methode - Interpretation

- ◆ LOD: Addition der LOD-Scores von unabhängigen Familien
- ◆ $\text{LOD} \geq 3$ Kopplung
- ◆ $\text{LOD} \leq -2$ keine Kopplung

Beispiel



Beispiel



Phase

$M-D|m-d$

$M-d|m-D$

r

nr

nr

nr

nr

r

r

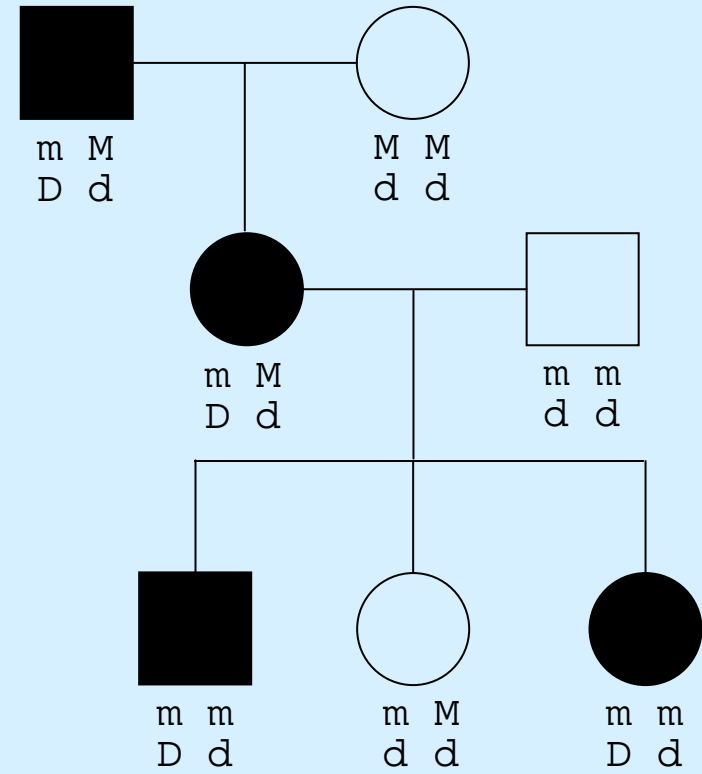
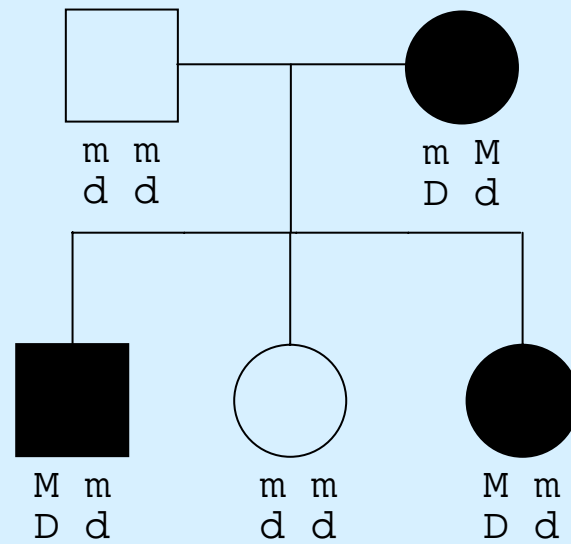
r

LOD Score Methode

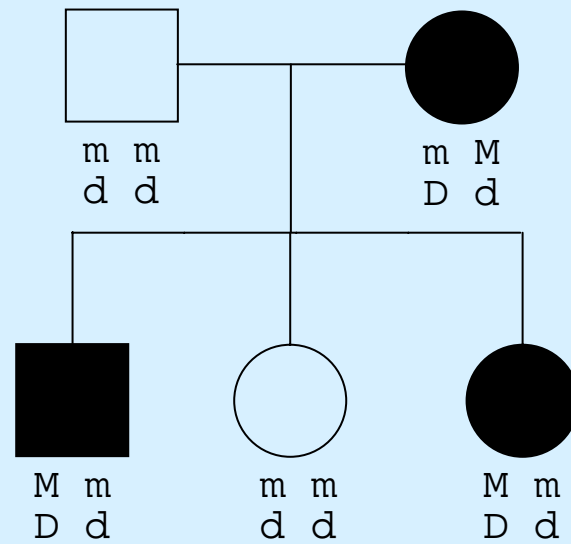
- ◆ Falls nicht entschieden werden kann, in welcher **Phase** sich die geerbten elterlichen Allele befinden, wird der Likelihood entsprechend gewichtet

$$L(\theta) = \frac{1}{2} \theta(1 - \theta)^3 + \frac{1}{2} \theta^3(1 - \theta)$$

Beispiel



Beispiel

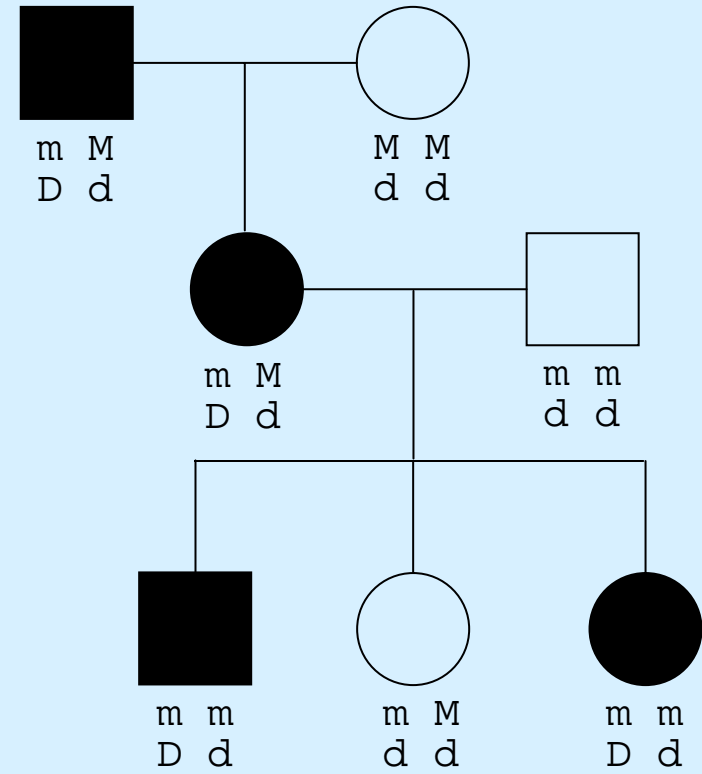


Phase
 $m-D|M-d$
 $M-D|m-d$

r
 nr

r
 nr

r
 nr



Phase
 $m-D|M-d$

nr

nr

nr

Beispiel Ergebnisse

	Lod Scores (Z) bei diversen Werten von θ									
Stammbaum	0.00	0.01	0.05	0.10	0.20	0.30	0.40	Z_{\max}	θ_{\max}	
Phase unbekannt	.602	.589	.533	.465	.318	.170	.049	.602	.00	
Phase bekannt	.903	.890	.837	.765	.612	.438	.237	.903	.00	

Frage

Die Linkage zwischen einem Polymorphismus und einer autosomal dominanten polyzystischen Nierenerkrankung wurde bei einer Stichprobe von britischen und holländischen Familien untersucht:

θ	0.00	0.01	0.10	0.20	0.30	0.40
Lod Scores (Z)	$-\infty$	23.4	24.6	19.5	12.85	5.5

$$Z_{\max} = 25.85 \text{ bei } \theta_{\max} = 0.05$$

- ◆ Wie würden Sie diese Daten interpretieren?

In einer Folgestudie wurde eine große Familie aus Sizilien mit polyzystischen Nierenerkrankungen untersucht:

θ	0.00	0.10	0.20	0.30	0.40
Lod Scores (Z)	$-\infty$	-8.34	-3.34	-1.05	-0.02

- ◆ Wie würden Sie diese Daten interpretieren?

Locus Heterogenität

Ist gegeben, wenn Mutationen in einem von mehreren Genen zu einem identischen Phänotyp führen können

z.B. Neurofibromatose:

- 2 Subgruppen (2 Loci)

Retinitis pigmentosa

- mehr als 60 genetische Loci

Multipoint Linkage Analyse

- ◆ Mehr als 2 Loci werden untersucht
 - Betrachte jeweils 2 Loci
 - Simultane Analyse der multiplen Loci

Allgemeinere Modelle

- ◆ Erweiterung der Methode für komplexere Vererbungsmodelle etc.
 - Elston-Stewart Algorithmus
 - Lander-Green Algorithmus
 - MCMC
- ◆ Aufwendige Berechnung bei großen Stammbäumen

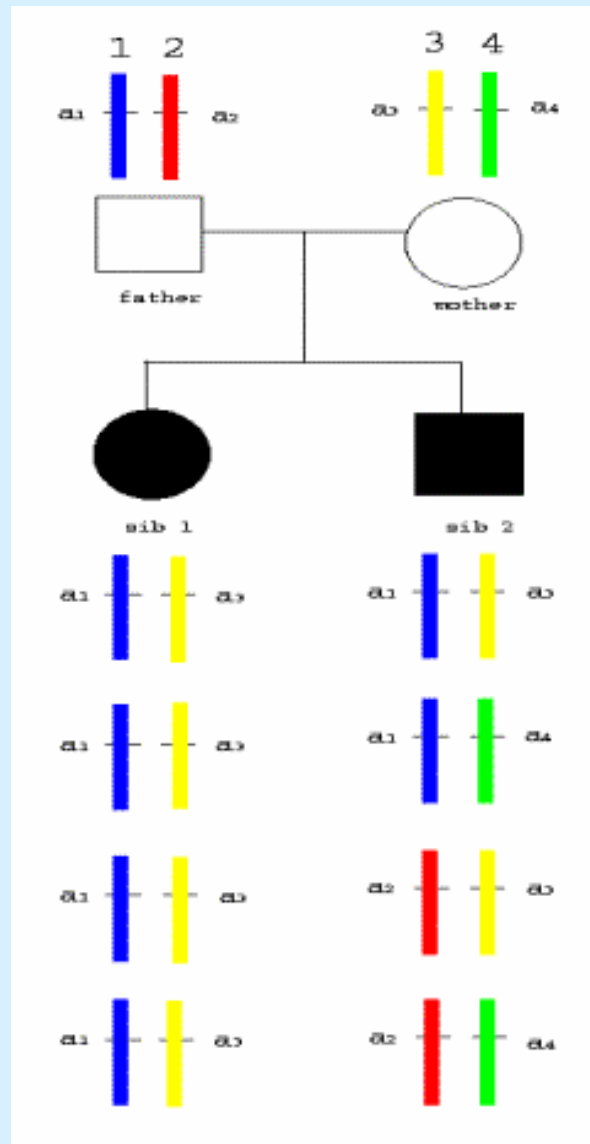
Linkage Analysen

- ◆ Modellbasierte Methoden (z.B. klassische LOD-Score Methode) – bestimmtes Vererbungsmodell wird berücksichtigt
- ◆ Modellfreie Methoden (z.B. affected sib-pair method) – keine Annahmen hinsichtlich Vererbungsmodell – komplexe Erkrankungen

Affected sib-pair Methode

- ◆ Eine nicht-parametrische Methode der Kopplungsanalyse für komplexe Erkrankungen ist die Untersuchung betroffener Geschwisterpaare (Geschwisterpaar-Analyse)
- ◆ Gemeinsame Marker-Allele als Anhaltspunkt für die Assoziation mit der Krankheit
- ◆ Untersuchung betroffener Geschwisterpaare basierend auf der Identität ererbter Anlagen (**identity-by-descent, IBD**)

Affected sib-pair method



Anzahl der
Allele IBD
mit der
Schwester

Wahrschein-
lichkeit des
Auftretens

2

$\frac{1}{4}$

1

1

$\frac{1}{2}$

0

$\frac{1}{4}$

Chi-Quadrat Test

- ◆ Die ASP Methode vergleicht die beobachtete IBD Verteilung in den betroffenen Geschwisterpaaren mit der erwarteten Verteilung (unter der Nullhypothese, dass es kein Linkage gibt)
- ◆ IBD Status des Geschwisterpaars ist nur voll bestimmbar, wenn beide Eltern heterozygot sind und nicht den gleichen Genotyp haben
- ◆ Test auf Linkage – z.B. Chi Quadrat Test

$$T = \sum_i \frac{(o_i - e_i)^2}{e_i}$$

IBD – Linkage Analyse

- ◆ Ist der Marker nicht gekoppelt mit dem Krankheitsgen, dann ist die erwartete Anzahl von IBD Allelen gleich 1
- ◆ Ist der Marker mit dem Krankheitsgen gekoppelt, dann ist die erwartete Anzahl von IBD Allelen >1

IBD: wenn 2 Personen eine ererbte Kopie eines einzelnen Allels besitzen, dann sind diese Allele **identical by descent**

IBS (identity-by-state): wenn 2 Allele dieselbe Form haben (z.B. die gleiche DNA Sequenz haben), dann sind sie **identical by state**

◆ Beispiel:

- Eltern: A1A2, A3A3;
- Kinder: A1A3, A1A3;

A1: IBD; A3: IBS

Likelihood Methoden

- ◆ Likelihood Methoden wurden eingeführt, um mit dem Problem der unvollständigen Daten umgehen zu können
- ◆ Für ein einzelnes Geschwisterpaar (mit Elterndaten) lässt sich die Likelihood-Funktion folgendermaßen anschreiben:

$$L(z) = \sum_{i=0}^2 z_i P(x | IBD = i)$$

wobei z_0, z_1, z_2 die Wahrscheinlichkeit bezeichnet, dass ein Geschwisterpaar 0,1,2 IBD Allele hat.

- ◆ Die Likelihood wird nach z maximiert und durch den Vergleich mit der Likelihood der Daten bei Annahme keiner Kopplung kann der LOD Score berechnet werden

Signifikanzniveau

Lander und Kruglyak haben vorgeschlagen, folgende Kriterien für ASP Studien bei komplexen Erkrankungen zu verwenden:

- ◆ LOD score $> 2,2$ suggestive linkage
- ◆ LOD score $> 3,6$ significant linkage

Weitere Methoden

- ◆ Vergleichen das gemeinsame Vorkommen von Allelen bei anderen Konstellationen von betroffenen Verwandten (Cousins 1. Grades) oder zwischen nicht betroffenen Geschwistern