

Polygene Krankheiten

Peter N. Robinson

Institut für medizinische Genetik
Charité Universitätsmedizin Berlin

23. Juni 2008

Kartierung mendelnder Erkrankungen (Rückblick)

- Die Suche nach Krankheitsgenen mit der Iod-Wert-Analyse setzt ein definiertes Vererbungsmodell voraus
 - ▶ Autosomal dominant
 - ▶ Autosomal rezessiv
 - ▶ X-chromosomal rezessiv
- Für komplexe Krankheiten haben wir jedoch keine genauen Vorstellungen über den Vererbungsmodus, Allelhäufigkeiten, Penetranz ...

dichotome und kontinuierliche Merkmale

- dichotom: gr. *dicha* = entzwei, *tom* ≈ Schnitt
- Ein "Entweder-Oder"-Merkmal
 - ▶ z.B. Hämophilie oder normale Gerinnung
 - ▶ Mendelsche Erkrankungen lassen sich als dichotome Merkmale verstehen
- Dichotome Merkmale lassen sich als Folge von Variation in einem Genlocus erklären

Kontinuierliche/Quantitative Merkmale

- Kontinuierliche Merkmale werden durch das Zusammenwirken von Allelen an mehreren Genorten (und durch Umweltfaktoren) bedingt
- **polygen**: durch eine große Zahl unabhängiger mendelscher Merkmale bestimmt

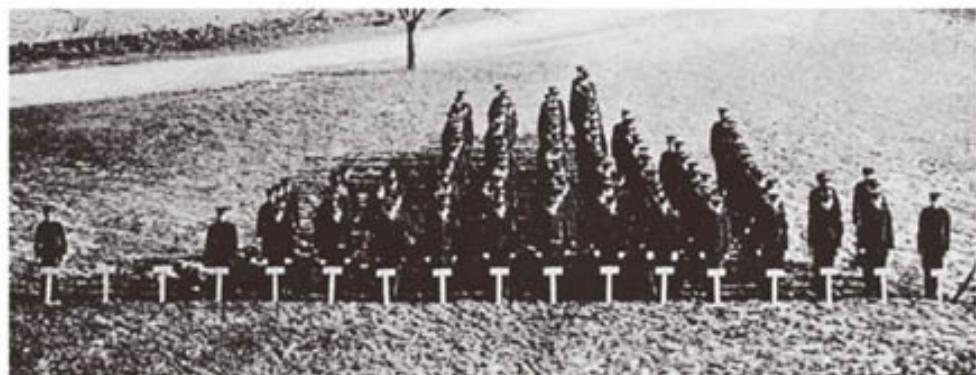
Gauß-Verteilung polygener Merkmale

Kleinbuchstabe: -5
Großbuchstabe: +5

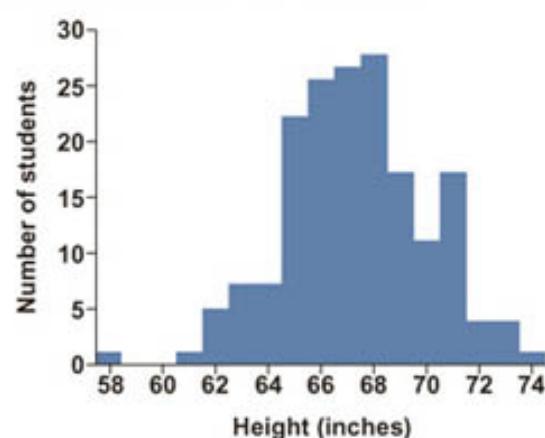
Urheberrechtlich geschütztes Bild entfernt

Körpergröße

A “living histogram”—distribution of height in a college class



Normal distribution = bell curve

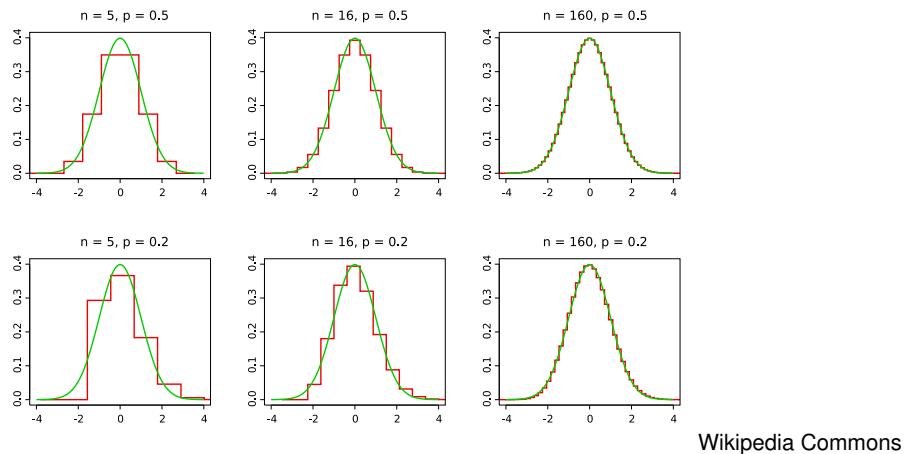


Zentraler Grenzwertsatz

- Unter der Annahme, dass die Ausprägung eines Merkmals das Ergebnis additiver genetischer Effekt ist

$$X = \sum X_i$$

- dann gilt der zentrale Grenzwertsatz (central limit theorem)
die Summe einer großen Zahl von unabhängigen, identisch verteilten Zufallsvariablen ist annähernd normalverteilt



Wikipedia Commons

Schwellenmodell

- Medizinisch interessanter sind jedoch die Krankheiten, die tendenziell familiär auftreten, ohne ein mendelndes Vererbungsmuster zu zeigen
- Viele solche Krankheiten sind dichotom, z.B. Gaumenspalte
- Anfälligkeit, eine Krankheit zu entwickeln, könnte jedoch einer Gauß'schen Verteilung entsprechen.
- Übersteigt die Anfälligkeit einen bestimmten Schwellenwert, kommt es zur Entwicklung der Krankheit

Urheberrechtlich geschütztes Bild entfernt

Der λ -Wert ist ein Maß für die Häufung in Familien

- Die erste Entscheidung ist, ob ein nicht mendelndes Merkmal genetisch (mit)bedingt ist
- Dazu muss erst gezeigt werden, dass eine Krankheit in Familien gehäuft auftritt
- Sei K die Prävalenz einer Erkrankung in der Bevölkerung
- Sei K_R die Prävalenz einer Erkrankung bei Verwandten (Relatives) eines erkrankten Proband
- Der λ -Wert von Risch[†] definiert sich als

$$\lambda_R = \frac{K_R}{K}$$

[†]Risch N (1987) Assessing the Role of HLA-linked and Unlinked Determinants of Disease
Am J Human Genet 40:1-14, 1987

Der λ -Wert ist ein Maß für die Häufung in Familien

- Man kann für jeden Verwandtschaftsgrad einen eigenen Wert bestimmen
- z.B. beträgt für Geschwister (**Siblings**) eines Probanden mit einer Schizophrenie $\lambda_S = 12,6$
- Unter bestimmten Voraussetzungen nimmt λ_R mit jedem Verwandtschaftsgrad um $\frac{1}{2}$ ab.

Erkrankte Geschwister Paare

Urheberrechtlich geschütztes Bild entfernt

- Affected sib pairs (ASP)-Methode
- Ziel: Auffinden von chromosomalen Segmenten, die betroffene Personen gemeinsam haben (Hinweis auf Anfälligkeitss locus)
 - ▶ Hierbei keine Annahme eines bestimmten Modells wie autosomal dominant
 - ▶ daher "nichtparametrisch"

Identical by state (IBS)

Urheberrechtlich geschütztes Bild entfernt

- identity by state = eigenschaftsidentisch
- Dieses Geschwisterpaar besitzt zwei unabhängige Kopien von A₁

Identical by descent (IBD)

Urheberrechtlich geschütztes Bild entfernt

- identity by descent = Identität durch Abstammung
- Dieses Geschwisterpaar besitzt Kopien desselben elterlichen A₁-Allels

Urheberrechtlich geschütztes Bild entfernt

- Für ein zufällig gewähltes Locus ist eine zufällige Segregation zu erwarten.
 - ▶ kein elterliches Allel gemeinsam: 1/4 (AC – BD)
 - ▶ ein elterliches Allel gemeinsam: 1/2 (AC – AD und AC – BC)
 - ▶ zwei elterliche Allele gemeinsam 1/4 (AC – AC)

Urheberrechtlich geschütztes Bild entfernt

- Handelt es sich beim Locus um den Krankheitslocus für eine dominante Krankheit, besitzen die Geschwister mindestens ein übereinstimmendes Allel.

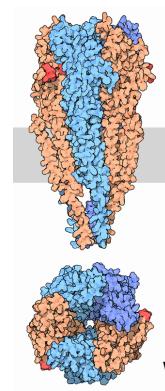
Urheberrechtlich geschütztes Bild entfernt

- Handelt es sich beim Locus um den Krankheitslocus für eine rezessive Krankheit, besitzen die Geschwister zwei übereinstimmende Allele.

- Im Allgemeinen (ungeachtet des genetischen Modells) ist eine signifikante Abweichung von den Häufigkeiten $1/4 - 1/2 - 1/4$ ein Hinweis für eine Kopplung an einen prädisponierenden Locus für eine Krankheit
- Daher lässt sich eine ASP-Analyse ohne vorherige Annahmen über die Genetik der Krankheit durchführen

Einzelnes Anfälligkeitssallel

- Was ist ein Anfälligkeitss locus (susceptibility locus)?
- Beispiel: Nikotinische Acetylcholinrezeptoren werden durch Acetylcholin aber auch durch Nikotin geöffnet
- Nikotinrezeptoren können an Karzinogene binden
- Eine Variante D398N im Nikotinrezeptorgen *CHRNA5* erhöht das Risiko für Lungenkrebs[†]



Wikipedia commons

[†] Hung RJ et al (2008) A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. Nature 452:633–637

Einzellocusmodell

- Betrachten wir nun ein Modell[†], wobei eine Variante in einem einzigen Locus für eine erhöhte Anfälligkeit für eine bestimmte (nicht mendelnde) Erkrankung verantwortlich ist
- Betrachten wir Verwandte vom Typ R (z.B. R=S, sibs/Geschwister)
 - Φ_R : Wahrscheinlichkeit, dass zwei Verwandte vom Typ R keine IBD-Allele teilen
- z.B. $\Phi_R = \frac{1}{4}$ für Geschwister, $\Phi_R = \frac{1}{2}$ für Halbgeschwister bzw. für Großeltern oder Enkel, ...
- K : Prävalenz der Erkrankung in der Allgemeinbevölkerung
- K_R : Prävalenz der Erkrankung unter Verwandten vom Typ R eines Probanden mit der Erkrankung

[†] Risch N (1987) Assessing the Role of HLA-linked and Unlinked Determinants of Disease Am J Human Genet 40:1-14, 1987

Einzellocusmodell (1)

- Falls die Verwandten kein gemeinsames IBD-Allel teilen, dann sind sie diesbezüglich genetisch unabhängig
- Die Wahrscheinlichkeit, dass eine Person in der Bevölkerung die Erkrankung entwickelt, ist K
- Die Wahrscheinlichkeit, dass zwei unabhängige Personen die Erkrankung entwickeln, ist K^2
- Die Wahrscheinlichkeit, dass zwei Verwandte erkranken, gegeben dass sie keine IBD-Allele teilen, ist daher

$$P(\text{2 R erkrankt} | \text{2 R mit 0 IBD}) = K^2$$

Einzellocusmodell (2)

- Ohne *a priori* Information über IBD-Allele, kann die Wahrscheinlichkeit, dass zwei Verwandte beide erkrankt sind, errechnet werden aus der Wahrscheinlichkeit, dass eine Person erkrankt (K) mal die Wahrscheinlichkeit, dass ein Verwandter dieser Person erkrankt (K_R)

$$P(\text{2 R erkrankt}) = K \cdot K_R$$

Einzellocusmodell (3)

- Die Wahrscheinlichkeit, dass zwei Verwandte erkranken und keine IBD-Allele teilen, lässt sich anhand der Definition der bedingten Wahrscheinlichkeit berechnen:

$$p(X, Y) = p(X|Y)p(Y)$$

$$\begin{aligned} & P(\text{2 R erkrankt und 2 R mit 0 IBD}) \\ &= P(\text{2 R erkrankt} | \text{2 R mit 0 IBD})P(\text{2 R mit 0 IBD}) \\ &= K^2\Phi_R \end{aligned}$$

Einzellocusmodell (4)

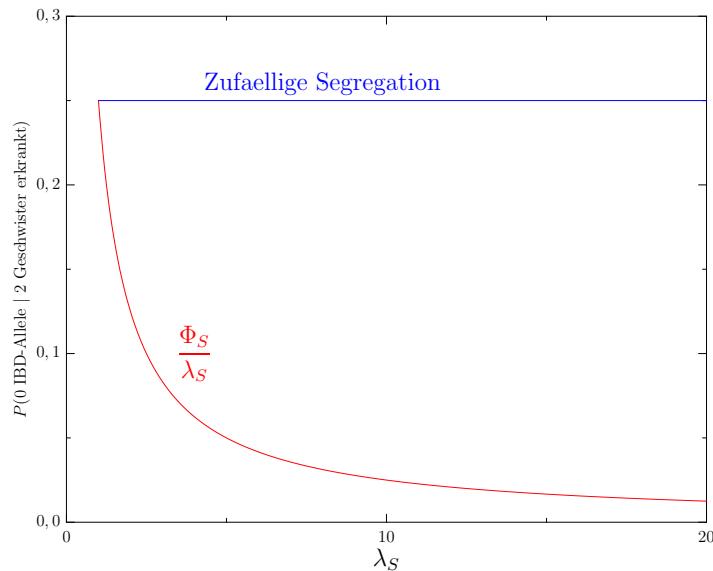
- Bayes' Gesetz

$$p(Y|X) = \frac{p(X|Y)p(Y)}{p(X)}$$

daher

$$\begin{aligned} & P(\text{2 R mit 0 IBD} | \text{2 R erkrankt}) \\ &= \frac{P(\text{2 R erkrankt} | \text{2 R mit 0 IBD})P(\text{2 R mit 0 IBD})}{P(\text{2 R erkrankt})} \\ &= \frac{K^2 \cdot \Phi_R}{K \cdot K_R} \\ &= \Phi_R \cdot \frac{K}{K_R} = \boxed{\frac{\Phi_R}{\lambda_R}} \end{aligned}$$

Einzellocusmodell (4)



- Je größer das Risiko für Geschwister eines erkrankten Probanden, selber zu erkranken (λ_S), desto unwahrscheinlich ist es, dass erkrankte Geschwisterpaare keine IBD-Allele teilen
- Umgekehrt steigt mit λ_S die Wahrscheinlichkeit für erkrankte Geschwisterpaare, gemeinsame IBD-Allele zu teilen.

Einzellocusmodell (5)

Urheberrechtlich geschütztes Bild entfernt

- Im Allgemeinen nimmt man die Existenz mehrfacher (jeweils unterschiedlicher) Anfälligkeitsgene für häufige Krankheiten wie Schizophrenie, Herzinfarkt, Diabetes mellitus...
- Eine Vielzahl mathematischer Modelle für multiple Anfälligkeitssoci sind entwickelt worden[†]

[†] z.B. Risch N (1990) Linkage Strategies for Genetically Complex Traits. 1. Multilocus Models Am. J. Hum. Genet. 46:222–228

ASP-Analyse: Ein Locus

- Verschiedene statistische Methoden. Am einfachsten ist der χ^2 -Test:

geteilte IBD-Allele	0	1	2
beobachtet:	O_0	O_1	O_2
erwartet:	E_0	E_1	E_2

$$\chi_{df=2}^2 = \sum_{i=0}^2 \frac{(O_i - E_i)^2}{E_i}$$

ASP-Analyse: Mehrere Loci

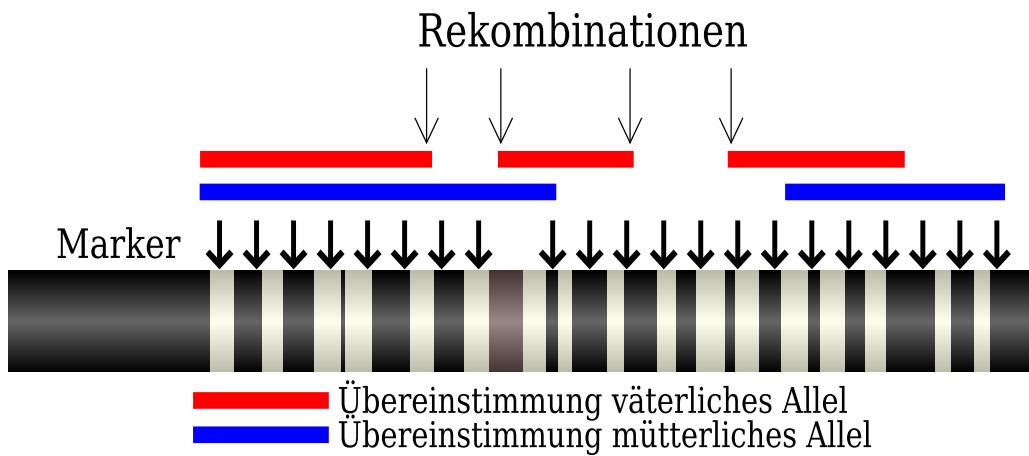
- Familien charakterisieren¹

Urheberrechtlich geschütztes Bild entfernt

¹ Worthington J. et al. (1994) Br. J. Rheumatol., 33, 970–976

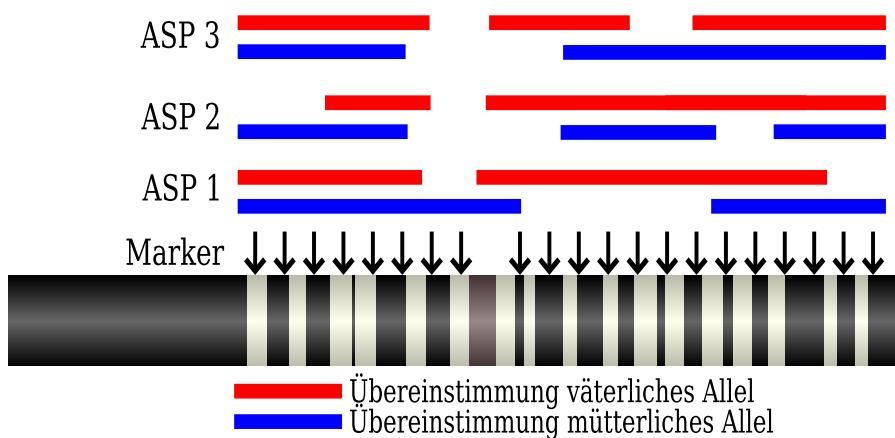
ASP-Analyse: Mehrere Loci

- Loci testen



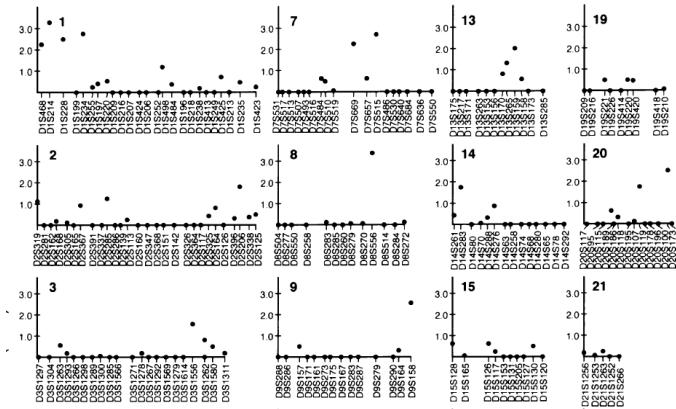
ASP-Analyse: Mehrere Loci

- Ergebnisse für alle Geschwisterpaare auswerten (mehrere statistische Verfahren)



ASP-Analyse: Mehrere Loci

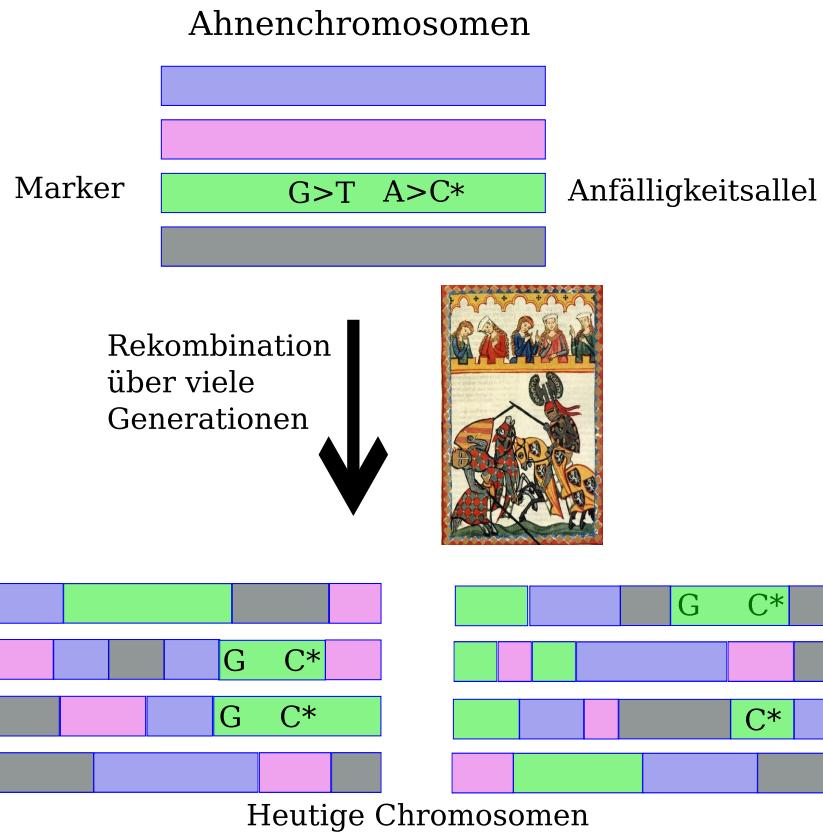
- Ergebnisse für alle Geschwisterpaare auswerten (mehrere statistische Verfahren)



Assoziation vs. Kopplung

Kopplung \leftrightarrow Loci	Wie weit ist ein Marker vom Krankheitslokus entfernt?
Assoziation \leftrightarrow Allele	Wird ein bestimmtes Markerallel mit dem prädisponierenden Allel mitvererbt?
Kopplung	familienspezisch
Assoziation	aufgrund eines Kopplungsungleichgewichts innerhalb einer Population
Assoziation	besteht über viel geringerer Abstände im Genom als Kopplung

Assoziation



Assoziation

Ewans WJ, Spielman RS (1995)

Urheberrechtlich geschütztes Bild entfernt

Am. J. Hum. Genet. 57:455–464

- Anfangs isolierte Populationen
- Nach der Zusammenkunft der ehemals getrennten Populationen, nimmt die Kopplung zwischen Markern entsprechend θ ab

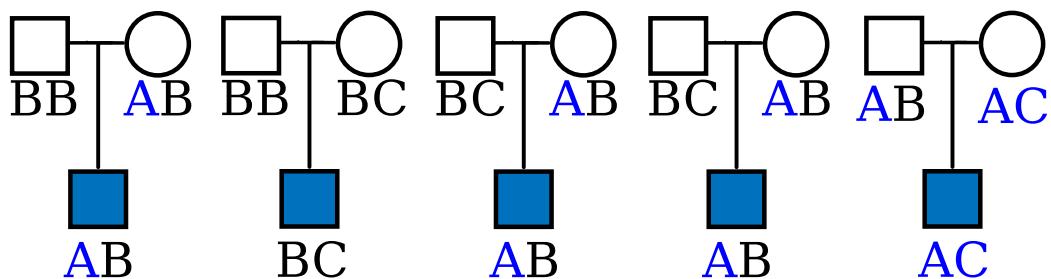
Linkage Disequilibrium

2 LD-Blöcke im CYP2C-CLuster

Urheberrechtlich geschütztes Bild entfernt

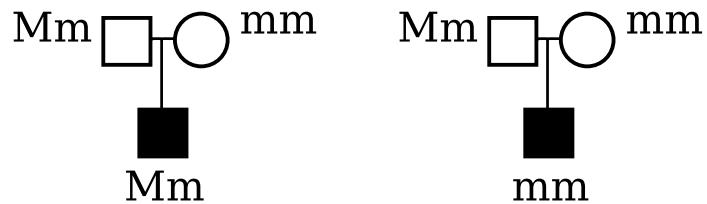
Walton et al (2005) Nature Genetics

Transmission-Disequilibrium-Test (TDT)



- In 4/5 Fällen wurde das Allel A an ein erkranktes Kind vererbt
- Annahme: Allel A ist in LD mit einem "angestammten" Anfälligkeitssallel/Krankheitsgen

Transmission-Disequilibrium-Test (TDT)



- $a = \#$ Fälle, bei denen M vererbt wurde
- $b = \#$ Fälle, bei denen M nicht vererbt wurde
- Statistik:

$$\chi_{td}^2 = \frac{(a-b)^2}{(a+b)}$$

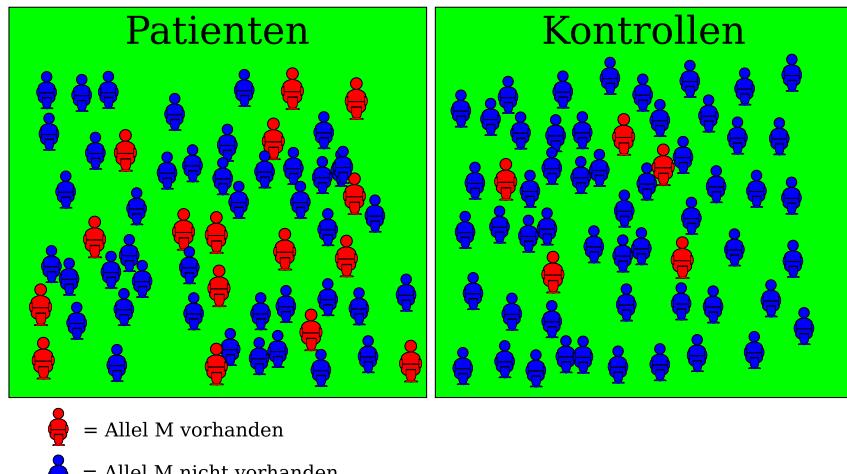
- Ungefähr χ^2 -verteilt mit einem Freiheitsgrad

Fall-Kontroll-Studien

- Die Häufigkeit von SNP-Allelen in zwei Gruppen vergleichen
 - ▶ Personen mit der Erkrankung
 - ▶ Unbetroffene Kontrollen oder zufällig aus der Bevölkerung ausgewählte Personen
- Eine erhöhte Häufigkeit eines SNP-Allels in Patienten verglichen mit Kontrollen deutet darauf hin, dass das Allel das Risiko für die Erkrankung erhöht

Fall-Kontroll-Studien

Fall-Kontroll-Studie



Genotyp	Patienten	Kontrollen
M vorhanden	a	b
M nicht vorhanden	c	d

- Statistische Auswertung mit χ^2 -Test

Fall-Kontroll-Studien

	AA	AB	BB
Patienten	a	b	c
Kontrollen	d	e	f

$$\chi^2 = \sum_{i=0}^2 \frac{(O_i - E_i)^2}{E_i}$$

- Mehrere Varianten für spezifische Fragestellungen[†]

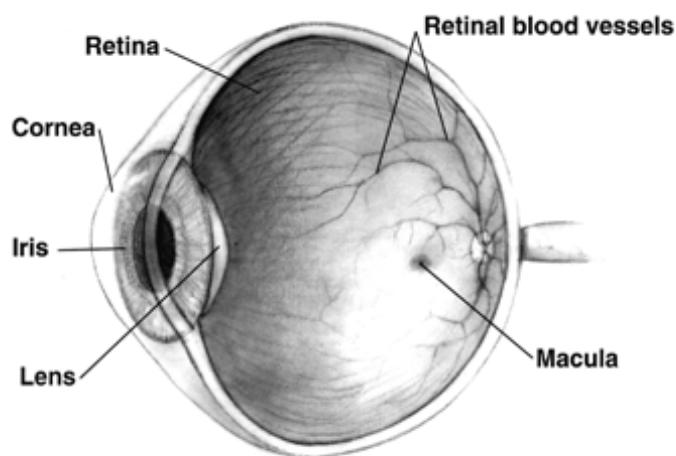
[†] vgl. Lewis CM (2002) Genetic association studies: design, analysis and interpretation. Brief Bioinform. 3:146-53.

Urheberrechtlich geschütztes Bild entfernt

- SNP-Chips mit bis zu 500.000 SNPs (bald mehr) von Affymetrix und anderen Anbietern

Bildquelle: www.imagenes-bio.de

Beispiel: Makuladegeneration



Bildquelle: Wikipedia commons

- Degeneration der Makula lutea ("Punkt des schärfsten Sehens")
- Hauptursache der Erblindung bei Menschen über 55 Jahre alt

Beispiel: Makuladegeneration

- Bis 2005 hatten Assoziationsstudien mit Kandidatengenen kein positives Ergebnis erbracht.
- Studie mit 100K SNP-Chip[†] mit 96 Fällen und 50 Kontrollen
- 116.204 SNPs ausgewertet, davon waren 105.980 informativ. Diejenigen 103.611 SNPs wurden in die Endanalyse einbezogen, die auf einem der 22 Autosomen gelegen sind

[†] Klein RJ, et al. (2005) Complement factor H polymorphism in age-related macular degeneration. Science. 308:385–9

Beispiel: Makuladegeneration

- Für jeden SNP wurde ein Test auf allelische Assoziation mit Makuladegeneration durchgeführt.
- Als Korrektur für multiples Testen wurde eine Bonferroni-Korrektur durchgeführt, d.h., nur diejenigen SNPs wurden für signifikant gehalten, welche einen p -Wert $< 0,05/103.611 = 4.8 \times 10^{-7}$ aufwiesen.

Beispiel: Makuladegeneration

- Zwei SNPs waren signifikant

Urheberrechtlich geschütztes Bild entfernt

Bildquelle: Klein et al. (2005)

Beispiel: Makuladegeneration

- Beide SNPs lagen in einem Intron des Gens für Complement H auf Chromosom 1q31
- Da die SNPs in nicht konservierten Sequenzen gelegen sind und keine funktionelle Bedeutsamkeit ersichtlich war, fragte man sich, ob die SNPs in Kopplungsungleichgewicht mit einem funktionellen Polymorphismus
- Eine LD-Analyse zeigte, dass die SNPs innerhalb einer Region mit hohem LD gelegen sind

Urheberrechtlich geschütztes Bild entfernt

Bildquelle: Klein et al. (2005)

Beispiel: Makuladegeneration

- Die SNP-Studie hat gekoppelte Marker innerhalb einer Region mit starkem Kopplungsungleichgewicht identifiziert
- Jetzt galt es, den eigentlichen funktionellen Polymorphismus zu finden, welcher die Entstehung der Makuladegeneration bei den Patienten begünstigt hatte
- Alle Exons des Gens für Complement H wurden sequenziert → Y402H
- Ablagerung von Complement wurden in der Netzhaut betroffener Patienten nachgewiesen

Urheberrechtlich geschütztes Bild entfernt

Bildquelle: Klein et al. (2005)

Assoziationsstudien

- Die Studie über Makuladegeneration wurde inzwischen mehrfach bestätigt
- Im Allgemeinen können jedoch viele Assoziationen nicht in unabhängigen Studien reproduziert werden
 - ▶ Methodische/statistische Unzulänglichkeiten
 - ▶ populationsspezifische Mutationen
 - ▶ Unterschiedliche klinische Kriterien
 - ▶ Unterschiedliche Umwelteffekte in unterschiedlichen Populationen

Gene

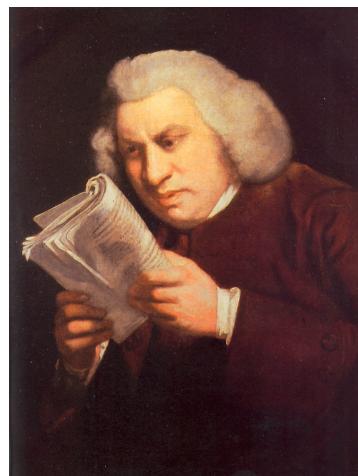
Umwelt

Marfan Syndrome Darmkrebs
Neurofibromatosis Brustkrebs
Hämophilie
Mukoviszidose

M. Alzheimer Diabetes
Asthma
Schizophrenie

Lungenkrebs Verkehrsunfall
Hautkrebs
Tuberkulose

The End of the Lecture as We Know It



- Diese Vorlesungsdias stehen unter der GNU-Lizenz für freie Dokumentation^b
- Kontakt:
peter.robinson@charite.de
- Strachan & Read, Kapitel 15

^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).