Positional cloning:
statistical approaches to gene mapping, i.e. locating genes on the genome

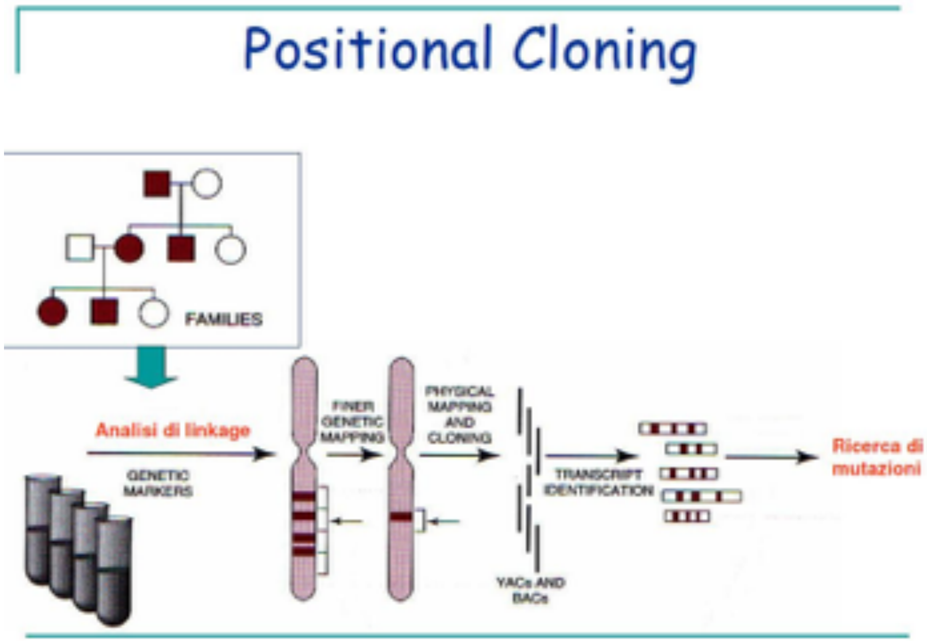
- Linkage analysis
- Association studies (Linkage disequilibrium)

Linkage analysis

- Uses a genetic marker map (a map of polymorphic loci)
- Looks for co-segregation with a marker (polymorphic locus)
- **Simple Idea:**
  - **To determine if marker allele at a known location travels with the disease in a family**

Linkage Analysis

To identify the chromosomal region in which the **disease-gene** is located without knowing its function

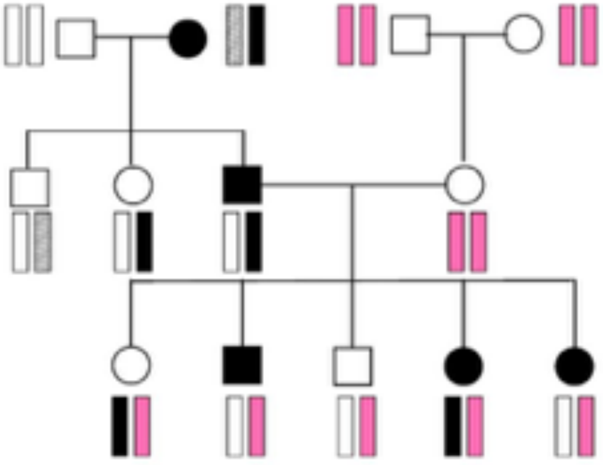


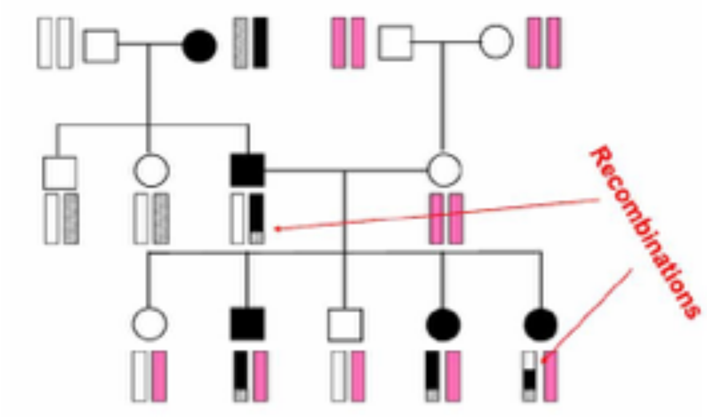
- If in a family a disease D is transmitted associated with specific markers M, then the disease -gene mapped near these markers and D and M segregate together

Main aim of linkage analysis:

To evaluate the distance between D and M

Chromosomal region **non in-linkage** with the disease

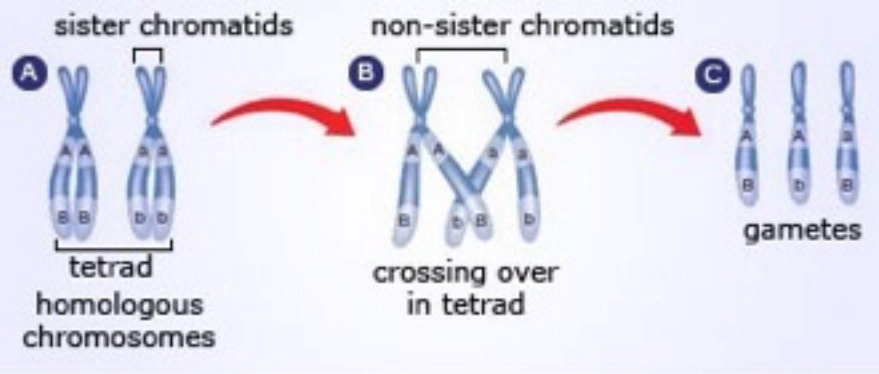


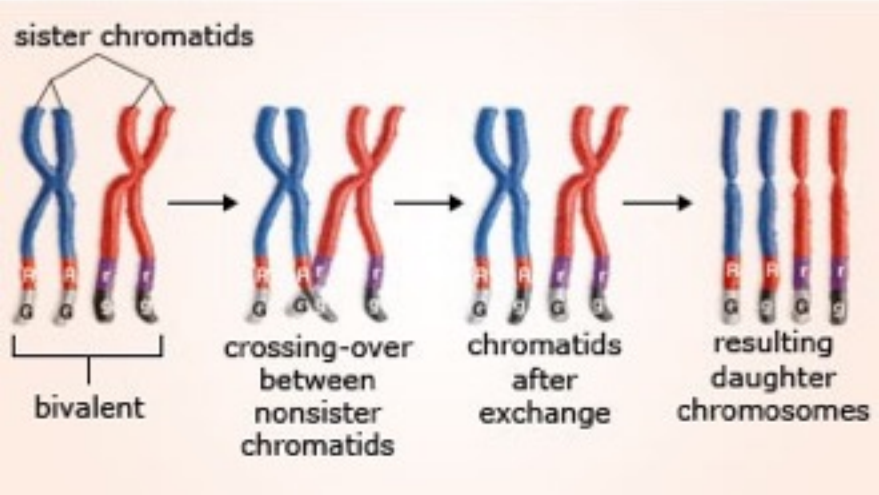


Chromosomal region **in-linkage** with the disease

Meiotic recombination is exploited to define the small region in-linkage with the disease

Linkage Analysis is based on RECOMBINATION




sister chromatids
bivalent
crossing-over between nonsister chromatids
chromatids after exchange
resulting daughter chromosomes

Genetic mapping

the aim is to discover **<u>how often two loci are separated</u>** <u>by meiotic recombination</u>

If two loci are **<u>on different chromosome</u>** they will segregate **<u>independently</u>**

Children will have 50% chance to receive each of these loci.

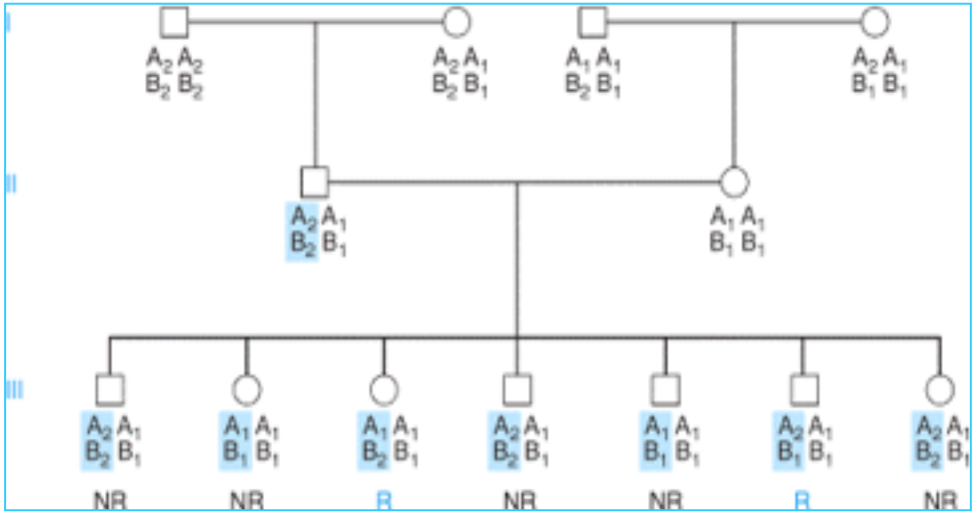
Recombination fraction is θ = 0.5

If loci are on the **<u>same chromosome</u>** they are expected to **<u>segregate together</u>** (θ= 0)
but due to <u>meiotic recombination this **does not always happen**</u> ($0 < \theta < 0.5$)
The further two loci are on the chromosome the more they recombinate
The θ is a **<u>measure of the distance</u>** between two loci

Recombination will rarely separate loci which lie very close (θ=0)

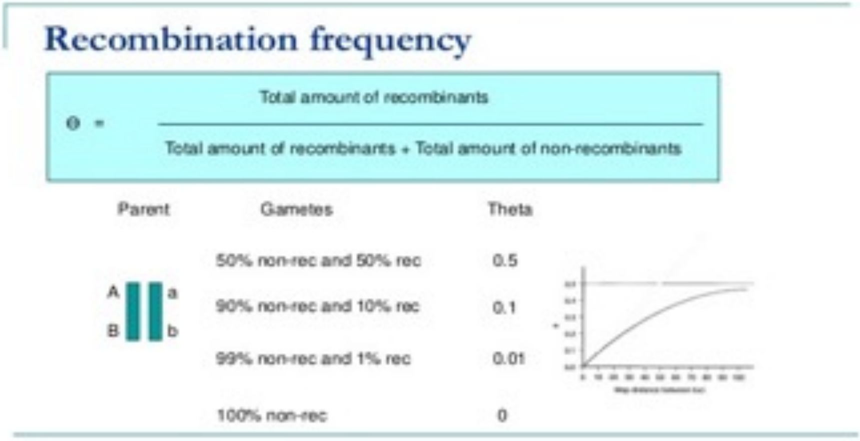
Alleles on the same small chromosome segment tend to be transmitted as a block through a pedigree - called **haplotype** -that can be tracked in families and populations

Recombination fraction defines **genetic distance**



How can we calculate the θ ?

The proportion of children who are recombinant is the <u>recombination fraction</u> between the two loci A and B during meiosis



Recombination frequency

$$\theta = \frac{\text{Total amount of recombinants}}{\text{Total amount of recombinants + Total amount of non-recombinants}}$$

Parent	Gametes	Theta
	50% non-rec and 50% rec	0.5
	90% non-rec and 10% rec	0.1
	99% non-rec and 1% rec	0.01
	100% non-rec	0

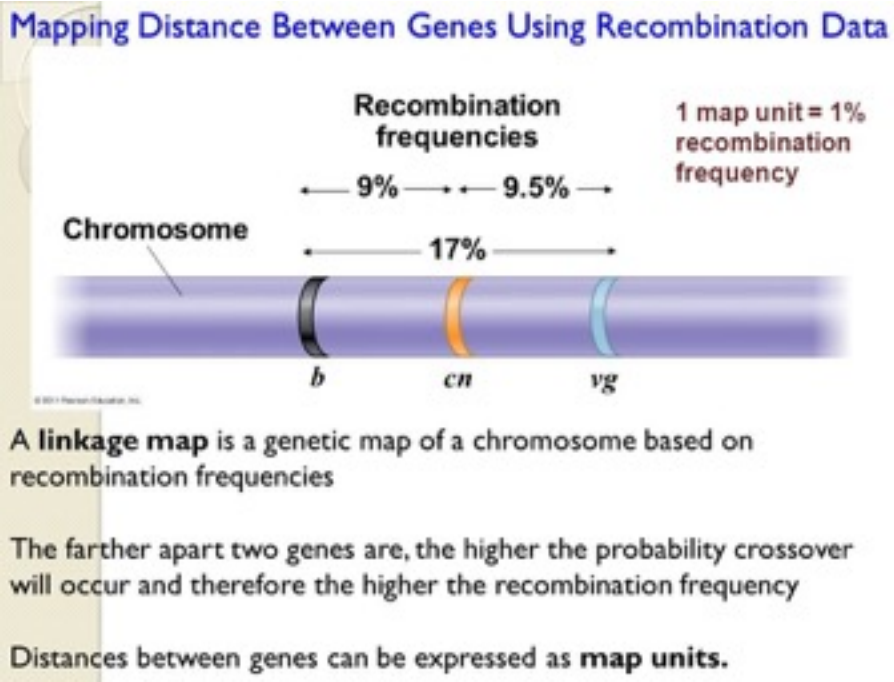
Two loci which show 1% of recombination are defined as 1 centimorgan distance (cM)

The mathematical relationship between
recombination fraction and **genetic map distance** is described by the mapping function
Haldane function $d= 0.5\ln(1- 2\theta)$

However there is interference during crossing-over since one chiasma can inhibit another…
Kosambi function $d=0.25\ln[(1+2\theta)/(1-2\theta)]$

Recombination map

Physical map



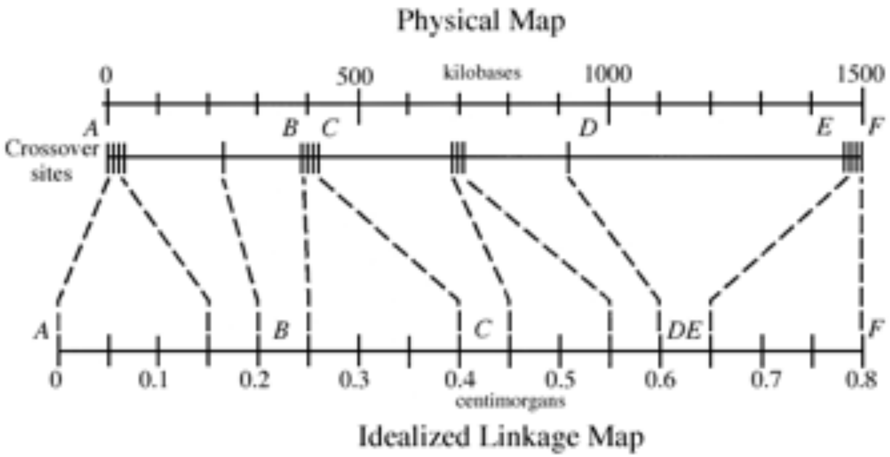
Mapping Distance Between Genes Using Recombination Data

Recombination frequencies
1 map unit = 1% recombination frequency
9%
9.5%
Chromosome
17%
b
cn
vg

A **linkage map** is a genetic map of a chromosome based on recombination frequencies

The farther apart two genes are, the higher the probability crossover will occur and therefore the higher the recombination frequency

Distances between genes can be expressed as **map units.**



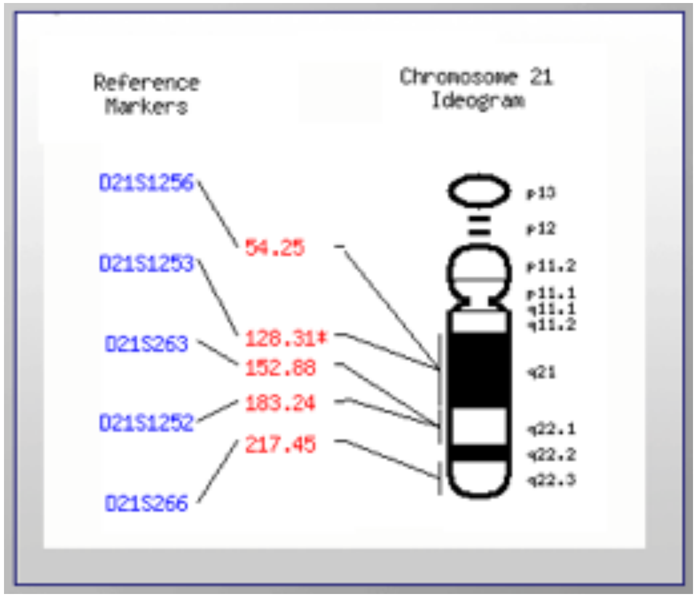
1 male cM = 0.9 Mb
1 female cM = 0.7 Mb

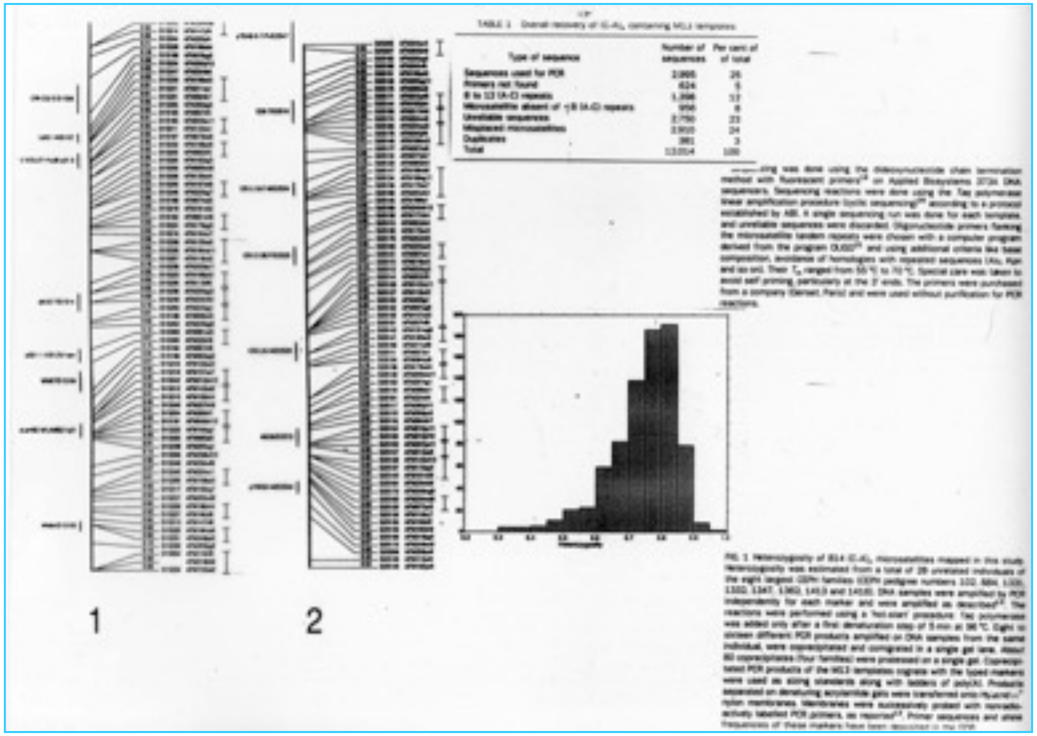
To perform linkage analysis you need **genetic markers**…

Characteristics:

- **highly polymorphic**
- **the rarest allele with a frequency of at least 1%**
- **feasible and stable in the pedigree**
- **well-known position in the genome**
- **genetic map of markers**

Genetic map of markers in 1980…





Human complete genetic map of microsatellite markers in 1992 by Cohen and colleagues

Genetic linkage and disease

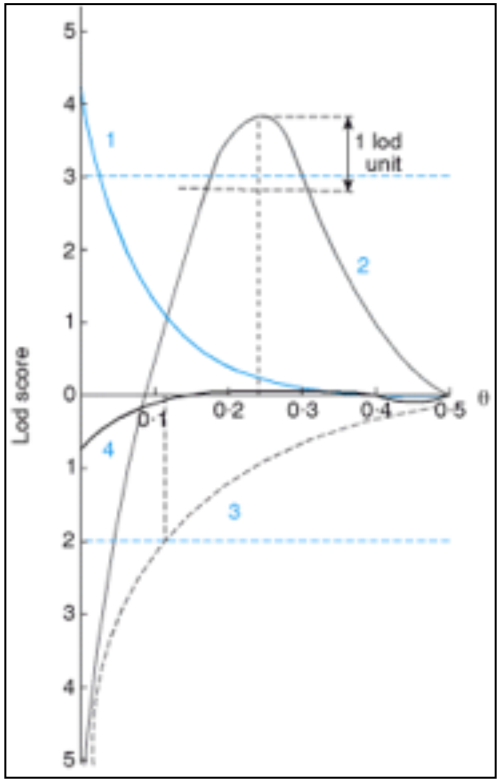
- Suitable large families are collected and <u>segregation of the disease is compared with the segregation of the markers</u>

- By using statistics it is tested the <u>probability</u> that the two loci (markers) <u>are not in linkage</u> (null hypothesis; threshold is p=0.05) in one family (LOD SCORE)

- Data from different families are collected and combined

LOD SCORE = is the logarithm of the odds that the <u>loci are linked rather than unliked</u>

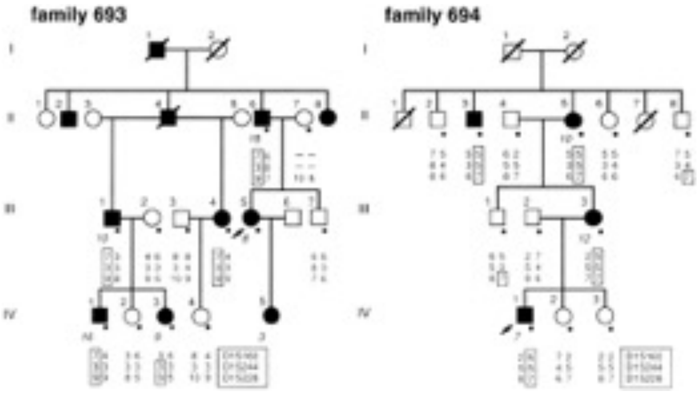
$$LOD = Z = \log 10 \frac{\text{probability of birth sequence with a given linkage value}}{\text{probability of birth sequence with no linkage}}$$
$$= \log 10 \frac{(1-\theta)^{NR} \times \theta^{R}}{0.5^{(NR+R)}}$$



- It is a function of recombination fraction and is the product of the probabilities in each individual family

- When θ is 0.5, lod score is 0

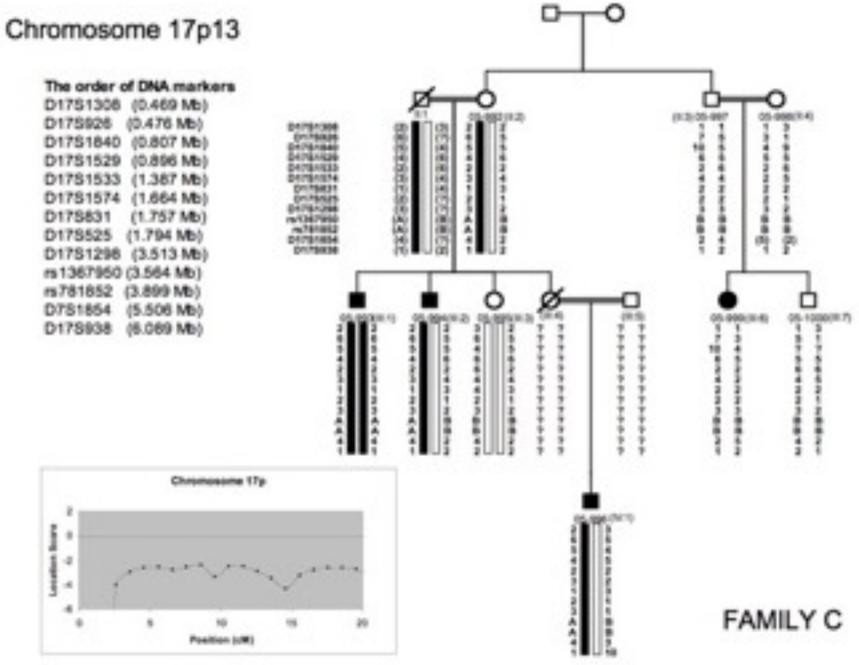
- Z =3 is the threshold to accept linkage



The overall probability of linkage in a set of families is the addition of LOD scores in each family

Linkage analysis can be more efficient if data for more than two loci are analyzed simultaneously

Multipoint mapping

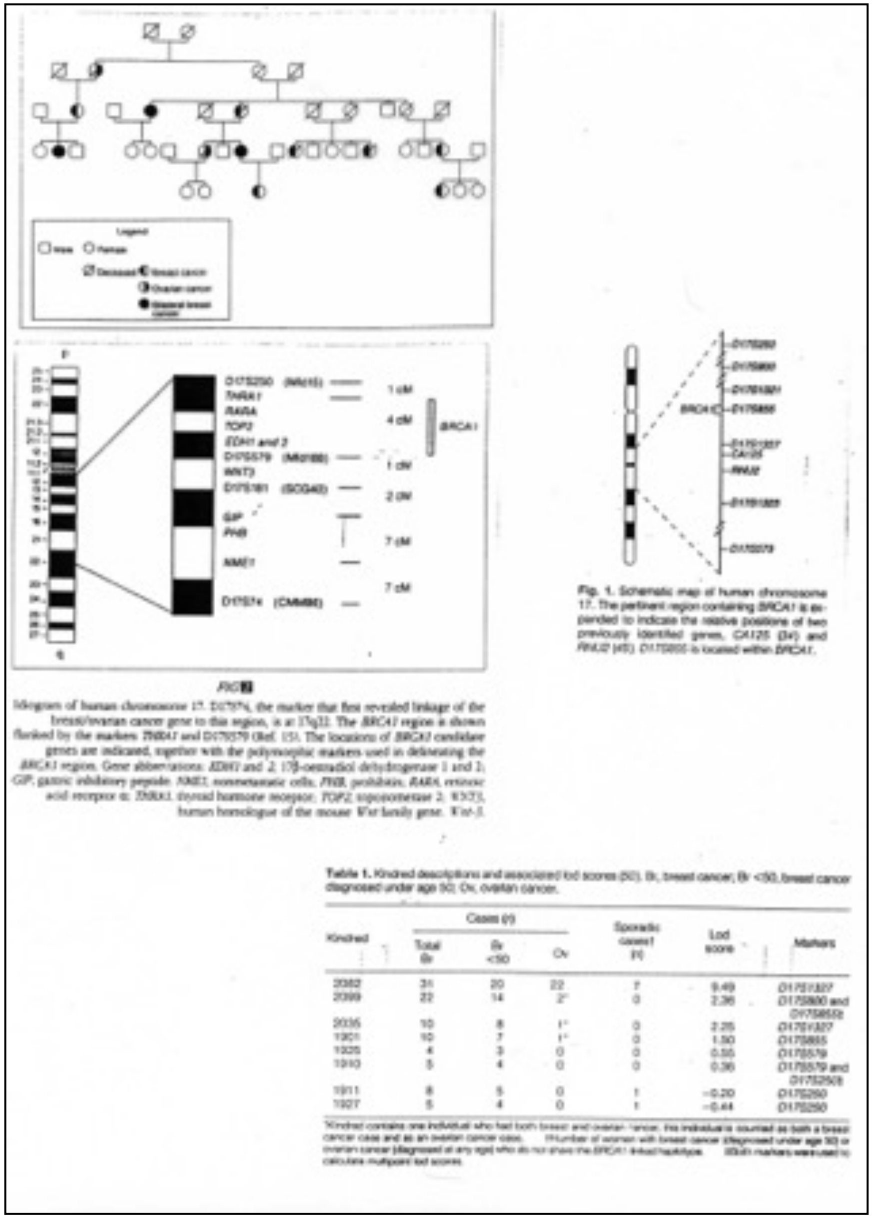


Chromosome 17p13

The order of DNA markers

D17S1308	(0.469 Mb)
D17S926	(0.476 Mb)
D17S1840	(0.807 Mb)
D17S1529	(0.896 Mb)
D17S1533	(1.387 Mb)
D17S1574	(1.664 Mb)
D17S831	(1.757 Mb)
D17S525	(1.794 Mb)
D17S1298	(3.513 Mb)
rs1367950	(3.564 Mb)
rs781852	(3.899 Mb)
D7S1854	(5.506 Mb)
D17S938	(6.069 Mb)

Linkage analysis and positional cloning: BRCA1



FIG

Ideogram of human chromosome 17. D17S74, the marker that first revealed linkage of the breast/ovarian cancer gene to this region, is at 17q21. The *BRCA1* region is shown flanked by the markers *THRA1* and D17S579 (Ref. 15). The locations of *BRCA1* candidate genes are indicated, together with the polymorphic markers used in delineating the *BRCA1* region. Gene abbreviations: *EDH1* and *2*, 17β-oestradiol dehydrogenase 1 and 2; *GIP*, gastric inhibitory peptide; *NME1*, nonmetastatic cells; *PHB*, prohibitin; *RARA*, retinoic acid receptor α; *THRA1*, thyroid hormone receptor; *TOP2*, topoisomerase 2; *WNT3*, human homologue of the mouse *Wnt* family gene, *Wnt-3*.

Fig. 1. Schematic map of human chromosome 17. The pertinent region containing *BRCA1* is expanded to indicate the relative positions of two previously identified genes, *CA125* (34) and *RNU2* (45). *D17S855* is located within *BRCA1*.

Table 1. Kindred descriptions and associated lod scores (50). Br, breast cancer; Br <50, breast cancer diagnosed under age 50; Ov, ovarian cancer.

Kindred	Cases (n)			Sporadic cases† (n)	Lod score	Markers
	Total Br	Br <50	Ov			
2082	31	20	22	7	9.49	*D17S1327*
2099	22	14	2*	0	2.36	*D17S800* and *D17S855*‡
2035	10	8	1*	0	2.25	*D17S1327*
1901	10	7	1*	0	1.50	*D17S855*
1925	4	3	0	0	0.55	*D17S579*
1910	5	4	0	0	0.36	*D17S579* and *D17S250*‡
1911	8	5	0	1	−0.20	*D17S250*
1927	5	4	0	1	−0.44	*D17S250*

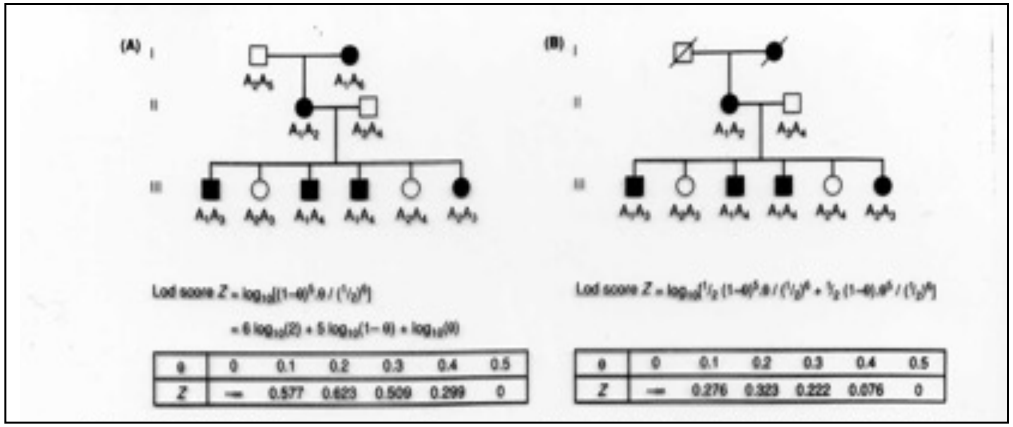
*Kindred contains one individual who had both breast and ovarian cancer; this individual is counted as both a breast cancer case and as an ovarian cancer case. †Number of women with breast cancer (diagnosed under age 50) or ovarian cancer (diagnosed at any age) who do not share the *BRCA1*-linked haplotype. ‡Both markers were used to calculate multipoint lod scores.

"The existence of BRCA1 was proven in 1990 by mapping predisposition to young-onset breast cancer in families to chromosome 17q21. Knowing that such a gene existed and approximately where it lays triggered efforts by public and private groups to clone and sequence it….BRCA1 was positionally cloned in September 1994…"
Mary Claire King

- identification of the more informative families
- identification of the locus for the susceptibility

Linkage analysis and positional cloning: BRCA1

- Families were analyzed with <u>microsatellites</u> spanning the region



Lod score $Z = \log_{10}[(1-\theta)^5 . \theta / (^1/_2)^6]$

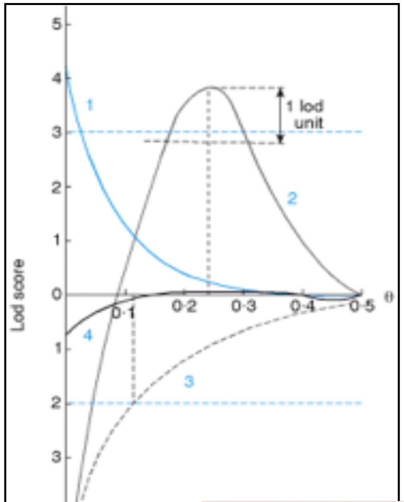
$= 6 \log_{10}(2) + 5 \log_{10}(1-\theta) + \log_{10}(\theta)$

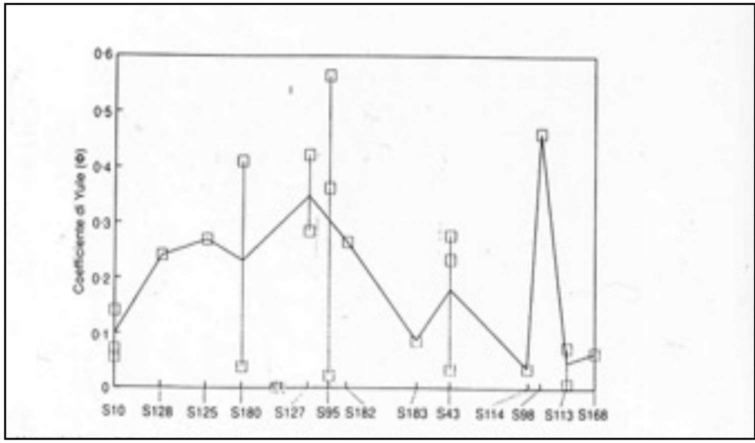
θ	0	0.1	0.2	0.3	0.4	0.5
Z	−∞	0.577	0.623	0.509	0.299	0

Lod score $Z = \log_{10}[^1/_2 (1-\theta)^5 . \theta / (^1/_2)^6 + ^1/_2 (1-\theta) . \theta^5 / (^1/_2)^6]$

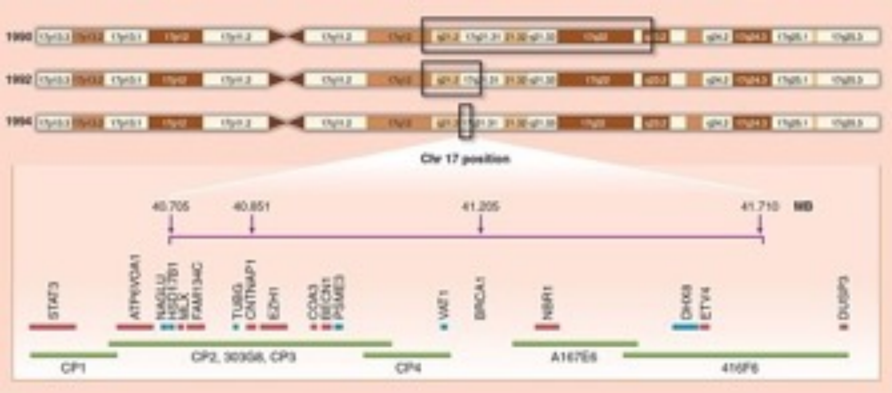
θ	0	0.1	0.2	0.3	0.4	0.5
Z	−∞	0.276	0.323	0.222	0.076	0

- <u>lod scores were calculated</u> for each family and add up for all the families





- characterization of <u>an open reading frame</u>



Parametric Linkage Analysis

Parametric linkage analysis can be applied when there is a probability that a gene important for a disease is linked to a genetic marker

It is studied using the LOD score, which assesses the probability that the disease and the marker are cosegregating.

It can be used when we have a pedigree with a clear type of inheritance and genotype-phenotype correlation

Non- Parametric Linkage Analysis

Non-parametric linkage analysis studies the probability of an allele being identical by descent with itself

It is used when the type of inheritance is not known

Less powerful, but you can apply it to a lot of families

Linkage analysis

- Great success in identifying genes for simple Mendelian diseases
- Few success in identifying genes contributing to complex disease
- Unsuccessful in identifying genes contributing to common complex disease

Linkage disequilibrium (LD)

- The nonrandom association of alleles in the population
- Alleles at neighboring loci tend to cosegregate
- Linkage disequilibrium implies population allelic association

Linkage Disequilibrium Mapping

- Population based
- Look for variant allele in LD with disease
- If most affected individuals in a population share the same mutant allele, then LD can be used to locate the chromosomal region harboring the disease

Association studies: which allele of which gene is associated with the disease?

Case-Control Studies

- Common method in epidemiology
- Cases
- Controls from the same population
  - this implies that cases and controls should have similar genetic backgrounds

Linkage and association are different phenomena

- **Association is a statistical statement about the co-occurrence of alleles or phenotypes** (e.i. allele **A** is associated with disease **D** if people who have **D** also have more **A**). The association can have many possible causes (not all genetics).

- **Linkage is a relationship between loci** and does not of itself produce any association in the general population.

Linkage creates association within families, but not among unrelated people.

However, if two supposedly unrelated people with disease **D** have inherited it from a distant common ancestor, they may well also tend to share particular ancestral alleles at loci closely linked to **D** **(Linkage disequilibrium)**

Statistical association can develop for different reason:

- direct cause-effect
- natural selection
- Stratification of the population
- linkage disequilibrium

Association studies are based on the use of haplotypes

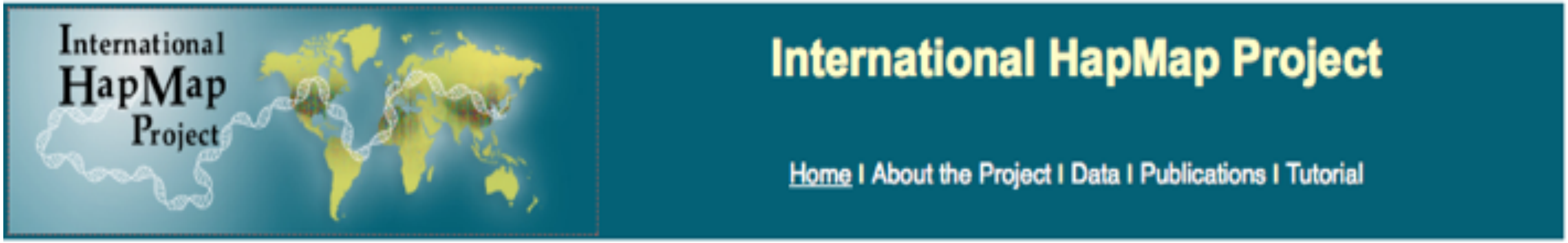
Case-control study: OR

	Cases (disease)	Controls (no-disease)
a allele	A	B
non-a allele	C	D

Odds ratio = odds allele *a* in cases / odds allele *a* in controls

Odds ratio= $\frac{A/B}{B/C}$ **= A/B x D/C= AD/BC**

OR= 1 (no association); OR>1 the allele contributes to the disease



International HapMap Project

http://www.hapmap.org/index.html

HapMap
and location of genes involved in medically important traits

About 10 million SNPs exist in human populations, where the rarer SNP allele has a frequency of at least 1%.

Researchers trying to discover the genes that affect a disease, such as diabetes, will compare a group of people with the disease to a group of people without the disease.

Chromosome regions where the two groups differ in their haplotype frequencies might contain genes affecting the disease.

Theoretically, researchers could look for these regions by genotyping 10 million SNPs. However, the methods to do this are currently too expensive.

The HapMap identifies which **200,000 to 1** million tag SNPs provide almost as much mapping information as the **10 million** SNPs.

This substantial cost reduction makes such studies feasible to do.

“Sporadic” CRC

- is a multifactorial (complex) condition
  - environmental factors
  - genetic factors

•study of twins: 35% of all CRC cases have a genetic component

•first-degree relatives of CRC patients are well- recognized to have a 2- to 4-fold increased risk of developing the disease

–recessive genes?

–pathogenic mutations of low penetrance

–complex gene-gene and gene-environment interactions

Tomlinson et al. ***A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3***
Nature Genetics 40, 623 - 630 (2008)

Tenesa et al. ***Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21***
Nature Genetics 40, 631 - 637 (2008)

In a **genome-wide association** study to identify **loci associated with colorectal cancer (CRC) risk**, we genotyped **555,510 SNPs in 1,012 early-onset Scottish CRC cases and 1,012 controls (phase 1)**. In phase 2, we genotyped the 15,008 highest-ranked SNPs in 2,057 Scottish cases and 2,111 controls. We then genotyped the five highest-ranked SNPs from the joint phase 1 and 2 analysis in **14,500 cases and 13,294** controls from seven populations, and identified a previously unreported association, rs3802842 on **11q23** (OR = 1.1; P = 5.8 times 10- 10), showing population differences in risk. We also replicated and fine-mapped associations at **8q24** (rs7014346; OR = 1.19; P = 8.6 times 10- 26) and **18q21** (rs4939827; OR = 1.2; P = 7.8 times 10- 28).

Carrying all six possible risk alleles yielded OR = 2.6 (95% CI = 1.75-3.89) for CRC. These findings extend our understanding of the role of common genetic variation in CRC etiology.

Tenesa et al. Nat. Genet., 2008

Genetic factors play a role in cancer

- Epidemiology studies show that relatives of cancer patients have a higher risk of developing the disease

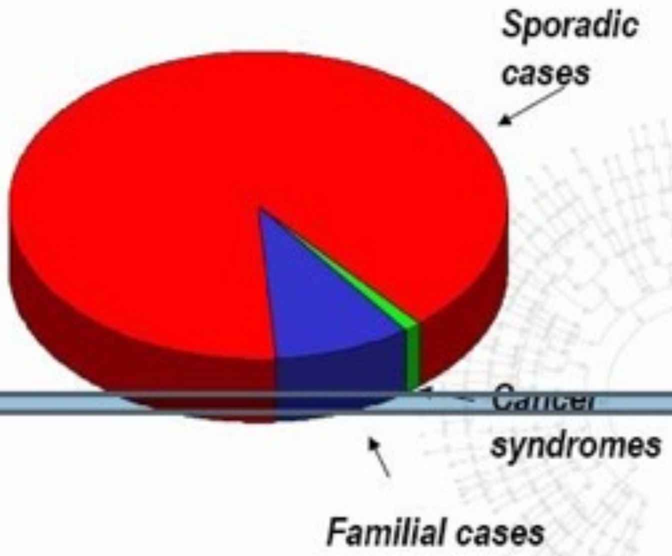
	Relative risk (RR)		
Cancer site	1° relatives	2° relatives	3° relatives
Thyroid	3.02	1.64	1.13
Kidney	2.3	1.31	1.32
Breast	2.02	1.36	1.21
Lung	2	1.39	1.1
Prostate	1.89	1.36	1.19

Amundadottir et al PLOS Med 2004

deCODE genetics

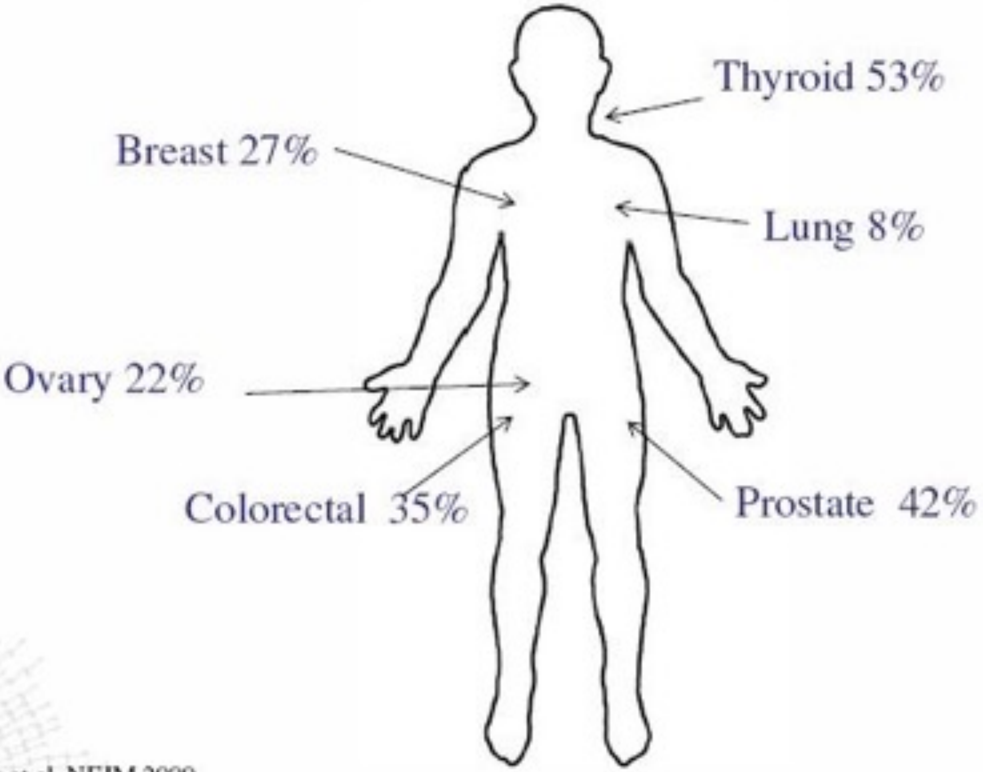
Only a fraction of cancer cases have family history of the disease

- Cancer syndromes (1-2%)
  - Rare, highly penentrant mutations e.g. p53 in Li-Fraumeni syndrome
- Familial cases (10-15%)
  - Mutations of intermediate penetrance, e.g. HNPCC, BRCA1/2
- Sporadic cases (>80%)
  - Family history not notable
  - Genetic factors still play a role



deCODE genetics

How much of cancer risk is explained by genetics ?

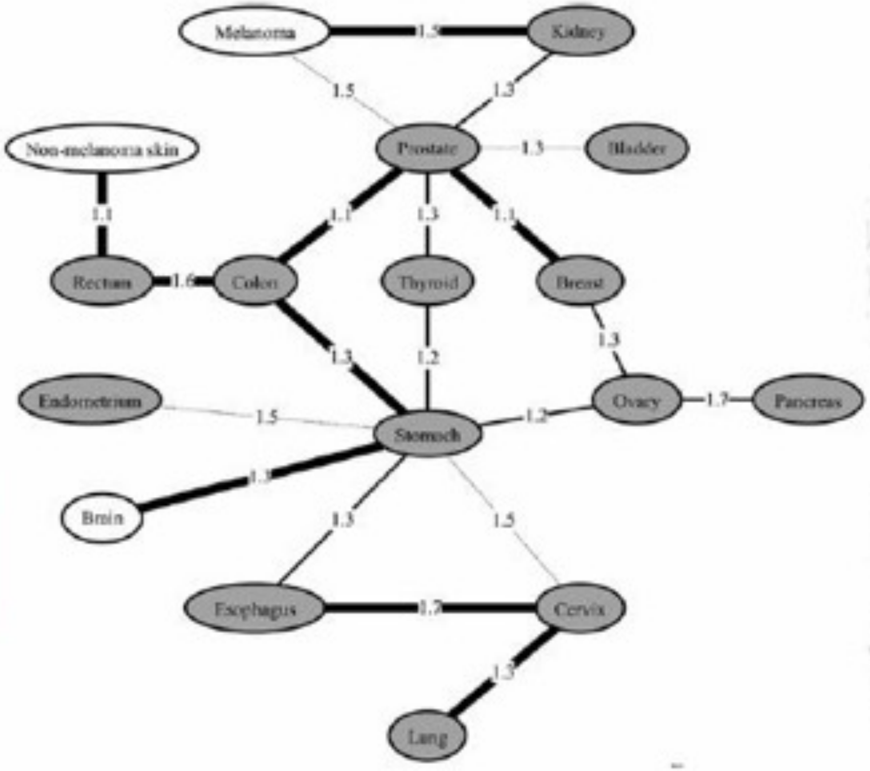


Lichtenstein et al. NEJM 2000
Czen et al. Int J Cancer 2002

deCODE genetics

Cross-risk of cancer in relatives

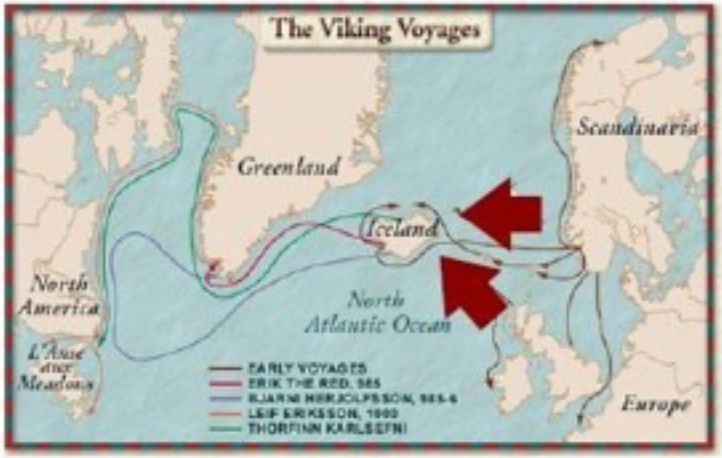
Icelandic Cancer Registry
X
Genealogy of all Icelanders



Can we find genes that explain this picture ? "Multi-cancer" genes

Amundadottir et al PLOS Med 2004

Icelandic history



Iceland founded in 9th century by settlers of mixed Northern European descent

Current Pop. 300,000

- **N-European descent**
- **Isolation for 11 centuries**

- **Same genetic background**
- **Smaller number of mutations**

deCODE genetics

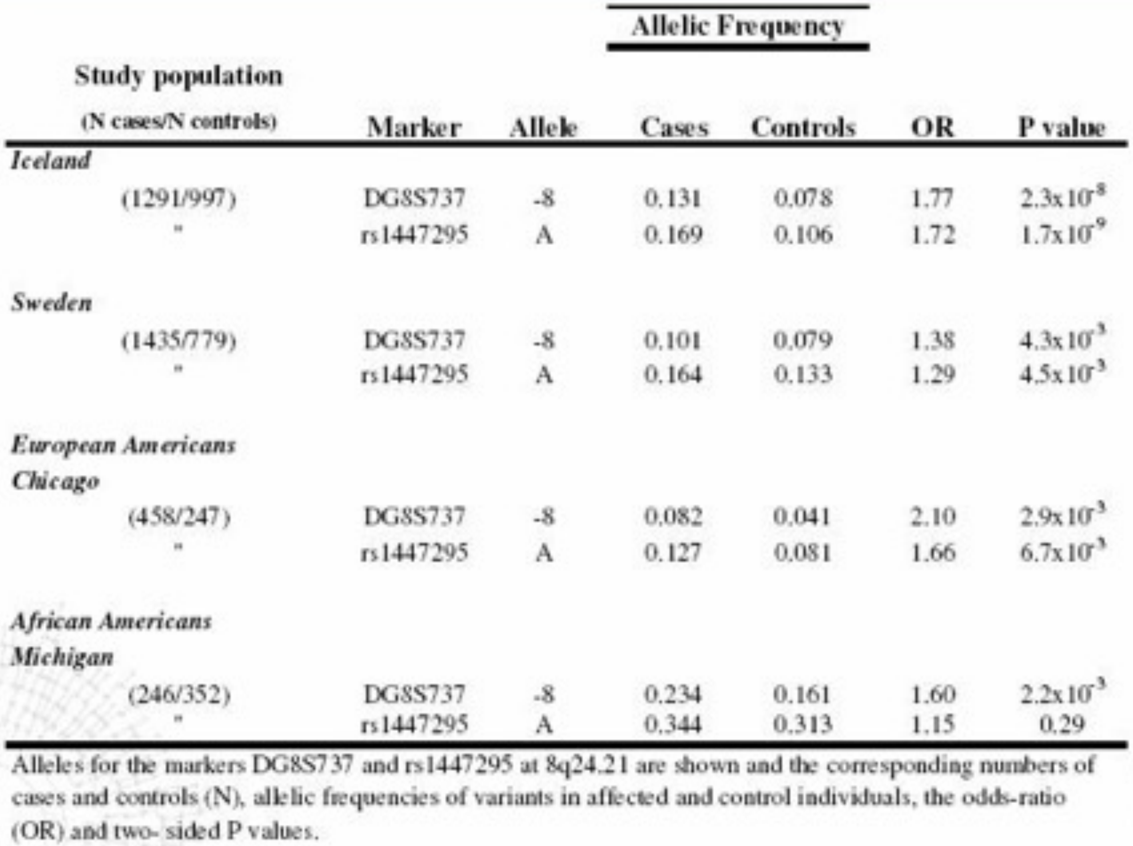
First major cancer project - prostate cancer

- A major public health problem
  - The most common cancer in males in the US
  - Lifetime risk 10% in EU - 16% USA
  - The second leading cause of cancer related deaths in men

- A genetic enigma
  - Genetic component one of the largest of all cancers
  - No highly-penetrant cancer genes isolated that can explain the familiality
  - Common polymorphisms in numerous genes reported to be associated with risk – hard to replicate

deCODE genetics

Results on 8q24 replicated

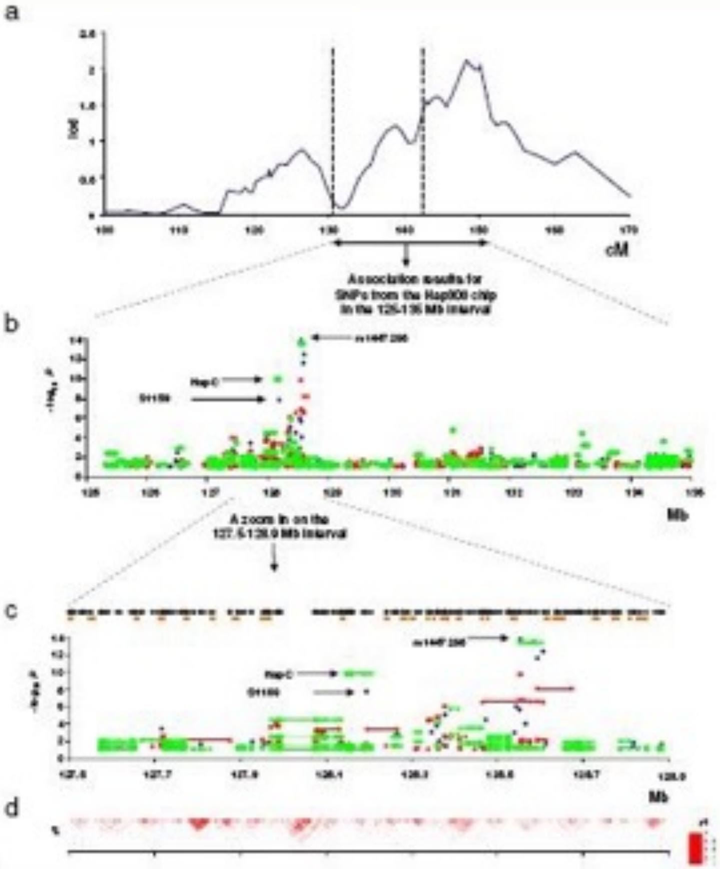


Study population			Allelic Frequency			
(N cases/N controls)	Marker	Allele	Cases	Controls	OR	P value
Iceland						
(1291/997)	DG8S737	-8	0.131	0.078	1.77	$2.3x10^{-8}$
"	rs1447295	A	0.169	0.106	1.72	$1.7x10^{-9}$
Sweden						
(1435/779)	DG8S737	-8	0.101	0.079	1.38	$4.3x10^{-3}$
"	rs1447295	A	0.164	0.133	1.29	$4.5x10^{-3}$
European Americans *Chicago*						
(458/247)	DG8S737	-8	0.082	0.041	2.10	$2.9x10^{-3}$
"	rs1447295	A	0.127	0.081	1.66	$6.7x10^{-3}$
African Americans *Michigan*						
(246/352)	DG8S737	-8	0.234	0.161	1.60	$2.2x10^{-3}$
"	rs1447295	A	0.344	0.313	1.15	0.29

Alleles for the markers DG8S737 and rs1447295 at 8q24.21 are shown and the corresponding numbers of cases and controls (N), allelic frequencies of variants in affected and control individuals, the odds-ratio (OR) and two- sided P values.

deCODE genetics

GWA identifies a second signal on 8q24



Gudmundsson et al., Nat.Genetics, 2007

deCODE genetics

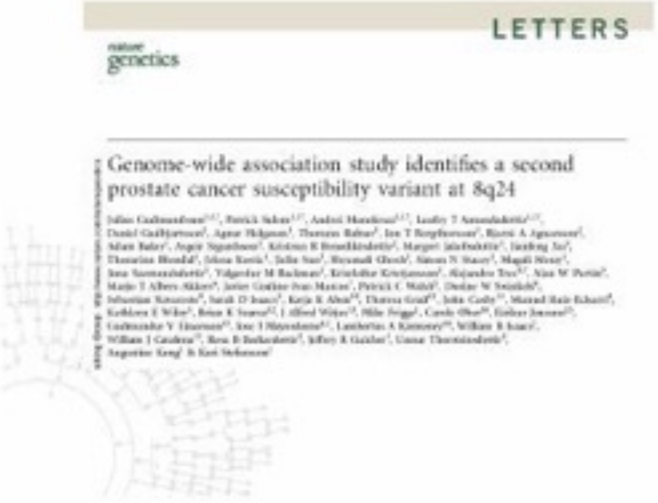
Several independent prostate cancer risk loci on chr8q24

LETTERS

nature genetics

A common variant associated with prostate cancer in European and African populations

Laufey T Amundadottir[1,10], Patrick Sulem[1,10], Julius Gudmundsson[1,10], Agnar Helgason[1], Adam Baker[1], Bjarni A Agnarsson[2], Asgeir Sigurdsson[1], Kristrun R Benediktsdottir[2], Jean-Baptiste Cazier[1], Jesus Sainz[1], Margret Jakobsdottir[1], Jelena Kostic[1], Droplaug N Magnusdottir[1], Shyamali Ghosh[1], Kari Agnarsson[1], Birgitta Birgisdottir[1], Louise Le Roux[1], Adalbjorg Olafsdottir[1], Thorarinn Blondal[1], Margret Andresdottir[1], Olafia Stanala Gretarsdottir[1], Jon T Bergthorsson[1], Daniel Gudbjartsson[1], Arnaldur Gylfason[1], Gudmar Thorleifsson[1], Andrei Manolescu[1], Kristleifur Kristjansson[1], Gudmundur Geirsson[3], Helgi Isaksson[3], Julie Douglas[4], Jan-Erik Johansson[5], Katarina Bälter[6], Fredrik Wiklund[6], James E Montie[7], Xiaoying Yu[8], Brian K Suarez[8], Carole Ober[9], Kathleen A Cooney[7,4], Henrik Gronberg[6], William J Catalona[8], Gudmundur V Einarsson[3], Rosa B Barkardottir[2], Jeffrey R Gulcher[1], Augustine Kong[1], Unnur Thorsteinsdottir[1] & Kari Stefansson[1]

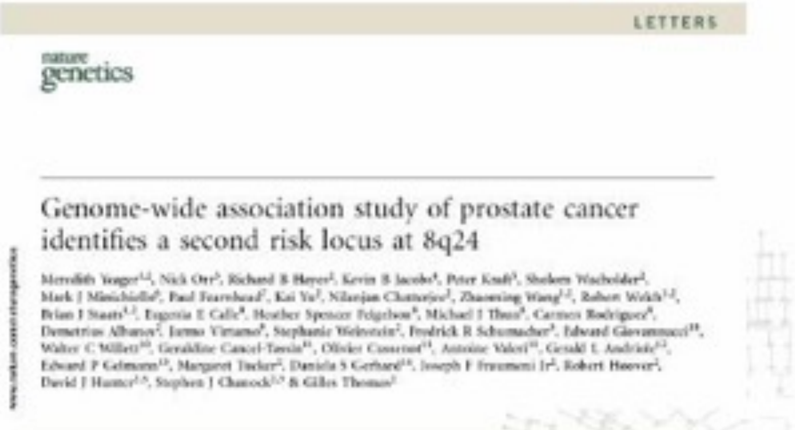


LETTERS

nature genetics

Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24

Julius Gudmundsson[1,11], Patrick Sulem[1,11], Andrei Manolescu[1,11], Laufey T Amundadottir[1,11], Daniel Gudbjartsson[1], Agnar Helgason[1], Thorunn Rafnar[1], Jon T Bergthorsson[1], Bjarni A Agnarsson[2], Adam Baker[1], Asgeir Sigurdsson[1], Kristrun R Benediktsdottir[2], Margret Jakobsdottir[1], Jianfeng Xu[3], Thorarinn Blondal[1], Jelena Kostic[1], Jielin Sun[3], Shyamali Ghosh[1], Simon N Stacey[1], Magali Mouy[1], Jona Saemundsdottir[1], Valgerdur M Backman[1], Kristleifur Kristjansson[1], Alejandro Tres[4,5], Alan W Partin[6], Maarjo T Albers-Akkers[7], Javier Godino-Ivan Marcos[5], Patrick C Walsh[6], Dorine W Swinkels[8], Sebastian Navarrete[9], Sarah D Isaacs[6], Katja K Aben[10], Theresa Graif[3], John Cashy[3], Manuel Ruiz-Echarri[9], Kathleen E Wiley[6], Brian K Suarez[12], J Alfred Witjes[11], Mike Frigge[1], Carole Ober[13], Eirikur Jonsson[14], Gudmundur V Einarsson[14], Jose I Mayordomo[4,5], Lambertus A Kiemeney[10], William B Isaacs[6], William J Catalona[3], Rosa B Barkardottir[2], Jeffrey R Gulcher[1], Unnur Thorsteinsdottir[1], Augustine Kong[1] & Kari Stefansson[1]



LETTERS

nature genetics

Genome-wide association study of prostate cancer identifies a second risk locus at 8q24

Meredith Yeager[1,2], Nick Orr[3], Richard B Hayes[2], Kevin B Jacobs[4], Peter Kraft[5], Sholom Wacholder[2], Mark J Minichiello[6], Paul Fearnhead[7], Kai Yu[2], Nilanjan Chatterjee[2], Zhaoming Wang[1,2], Robert Welch[1,2], Brian J Staats[1,2], Eugenia E Calle[8], Heather Spencer Feigelson[8], Michael J Thun[8], Carmen Rodriguez[8], Demetrius Albanes[2], Jarmo Virtamo[9], Stephanie Weinstein[2], Fredrick R Schumacher[5], Edward Giovannucci[10], Walter C Willett[10], Geraldine Cancel-Tassin[11], Olivier Cussenot[11], Antoine Valeri[11], Gerald L Andriole[12], Edward P Gelmann[13], Margaret Tucker[2], Daniela S Gerhard[14], Joseph F Fraumeni Jr[2], Robert Hoover[2], David J Hunter[2,5], Stephen J Chanock[1,2] & Gilles Thomas[2]

www.nature.com/naturegenetics

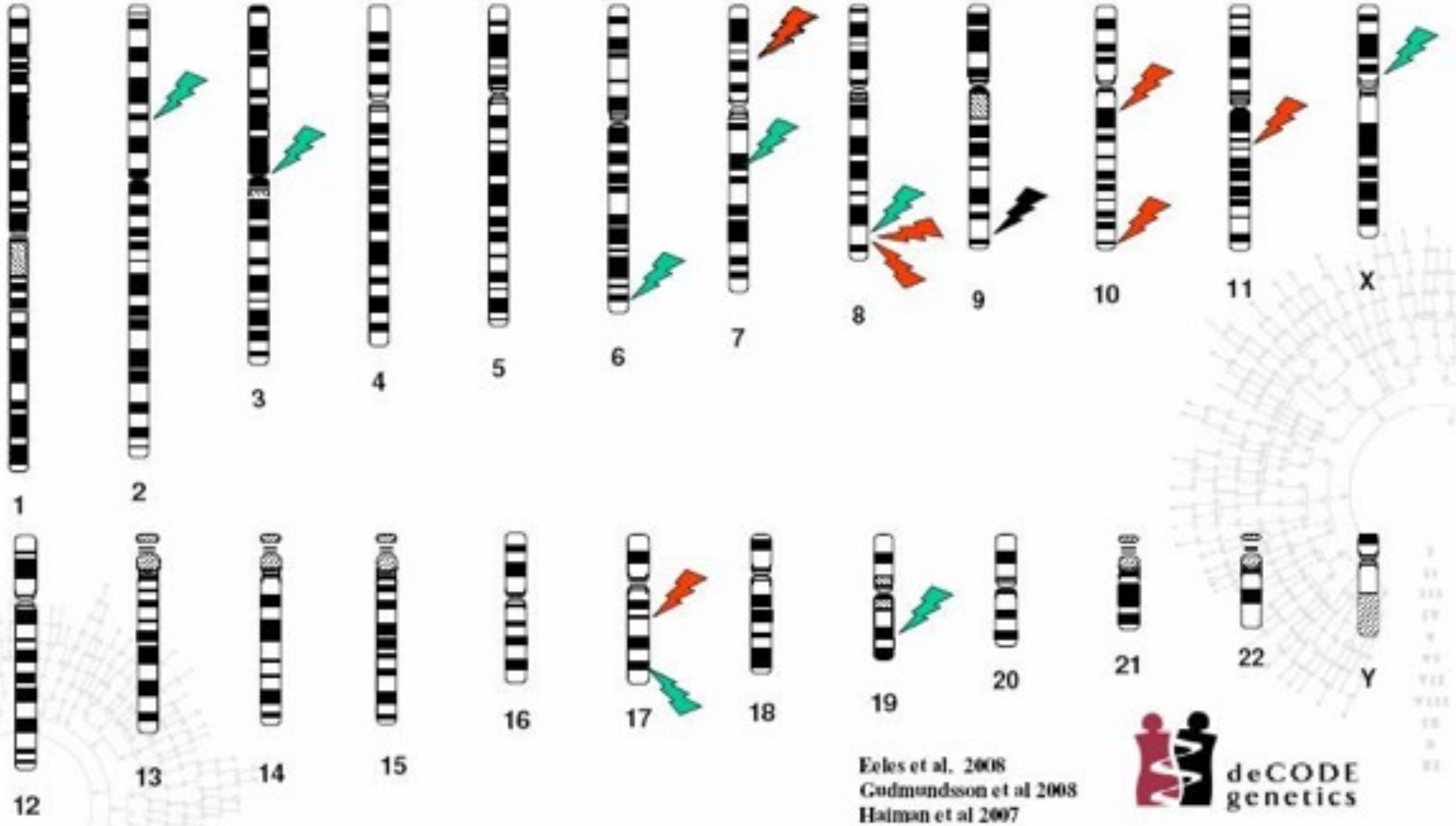
LETTERS

nature genetics

Multiple newly identified loci associated with prostate cancer susceptibility

Rosalind A Eeles[1,2], Zsofia Kote-Jarai[1,14], Graham G Giles[3,4,14], Ali Amin Al Olama[5,14], Michelle Guy[1,14], Sarah K Jugurnauth[1], Shani Mulholland[1], Daniel A Leongamornlert[1], Stephen M Edwards[1], Jonathan Morrison[5], Helen I Field[6], Melissa C Southey[7], Gianluca Severi[3,4], Jenny L Donovan[8], Freddie C Hamdy[9], David P Dearnaley[1,2], Kenneth R Muir[10], Charmaine Smith[5], Melisa Bagnato[1], Audrey T Ardern-Jones[2], Amanda L Hall[1,2], Lynne T O'Brien[1], Beatrice N Gehr-Swain[1,2], Rosemary A Wilkinson[1], Angie Cox[9], Sarah Lewis[8], Paul M Brown[11], Sameer G Jhavar[1], Malgorzata Tymrakiewicz[1], Artitaya Lophatananon[10], Sarah L Bryant[1], The UK Genetic Prostate Cancer Study Collaborators[11], British Association of Urological Surgeons' Section of Oncology[11], The UK ProtecT Study Collaborators[11], Alan Horwich[12], Robert A Huddart[1,2], Vincent S Khoo[2,1], Christopher C Parker[1,2], Christopher J Woodhouse[2], Alan Thompson[2], Tim Christmas[2], Chris Ogden[2], Cyril Fisher[2], Charles Jamieson[2], Colin S Cooper[1], Dallas R English[4], John L Hopper[4], David E Neal[11,13,14] & Douglas F Easton[5]

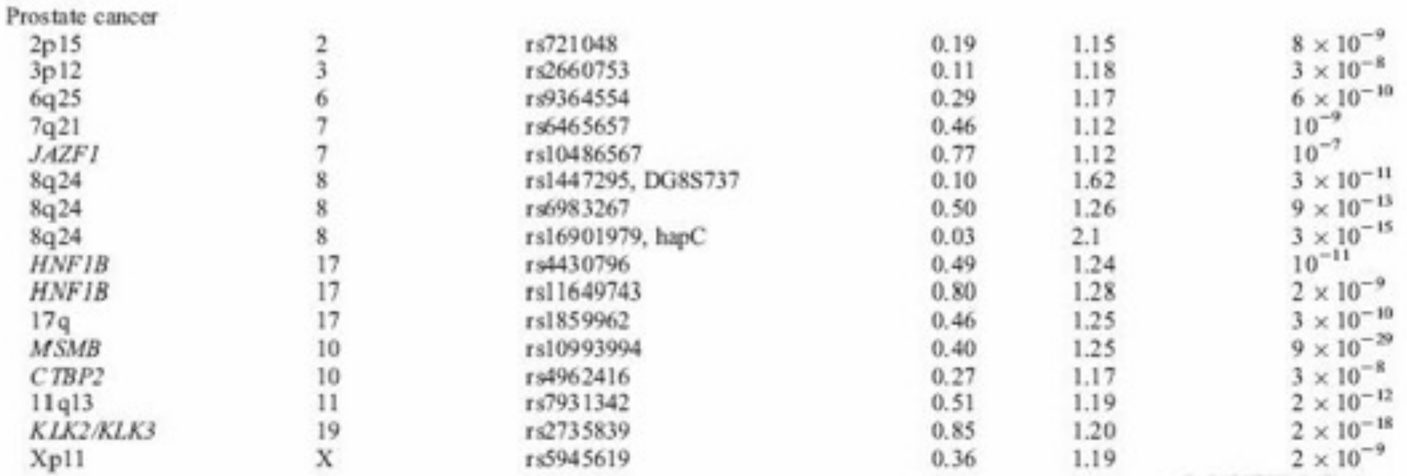
GWA studies have identified >16 loci involved in prostate cancer



Eeles et al. 2008
Gudmundsson et al 2008
Haiman et al 2007

deCODE genetics

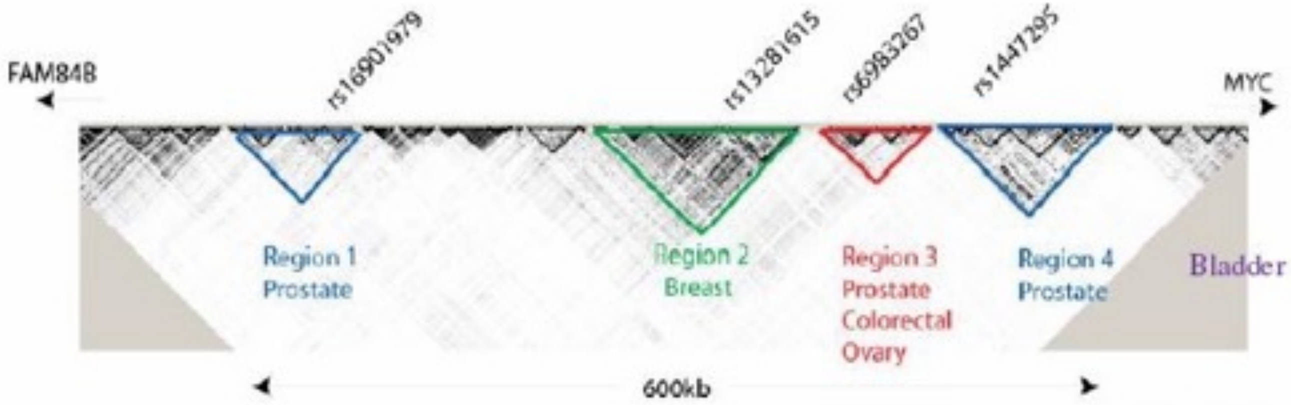
16 prostate cancer variants....



Prostate cancer					
2p15	2	rs721048	0.19	1.15	8×10^{-9}
3p12	3	rs2660753	0.11	1.18	3×10^{-8}
6q25	6	rs9364554	0.29	1.17	6×10^{-10}
7q21	7	rs6465657	0.46	1.12	10^{-9}
JAZF1	7	rs10486567	0.77	1.12	10^{-7}
8q24	8	rs1447295, DG8S737	0.10	1.62	3×10^{-11}
8q24	8	rs6983267	0.50	1.26	9×10^{-13}
8q24	8	rs16901979, hapC	0.03	2.1	3×10^{-15}
HNF1B	17	rs4430796	0.49	1.24	10^{-11}
HNF1B	17	rs11649743	0.80	1.28	2×10^{-9}
17q	17	rs1859962	0.46	1.25	3×10^{-10}
MSMB	10	rs10993994	0.40	1.25	9×10^{-29}
CTBP2	10	rs4962416	0.27	1.17	3×10^{-8}
11q13	11	rs7931342	0.51	1.19	2×10^{-12}
KLK2/KLK3	19	rs2735839	0.85	1.20	2×10^{-18}
Xp11	X	rs5945619	0.36	1.19	2×10^{-9}

deCODE genetics

Cancer risk variants on chr8q24



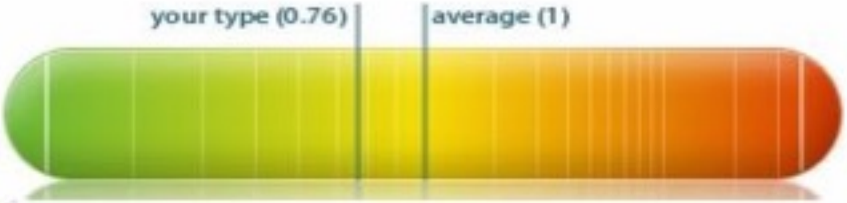
Most variants specific for a particular cancer

Does not explain cross-risk of cancers

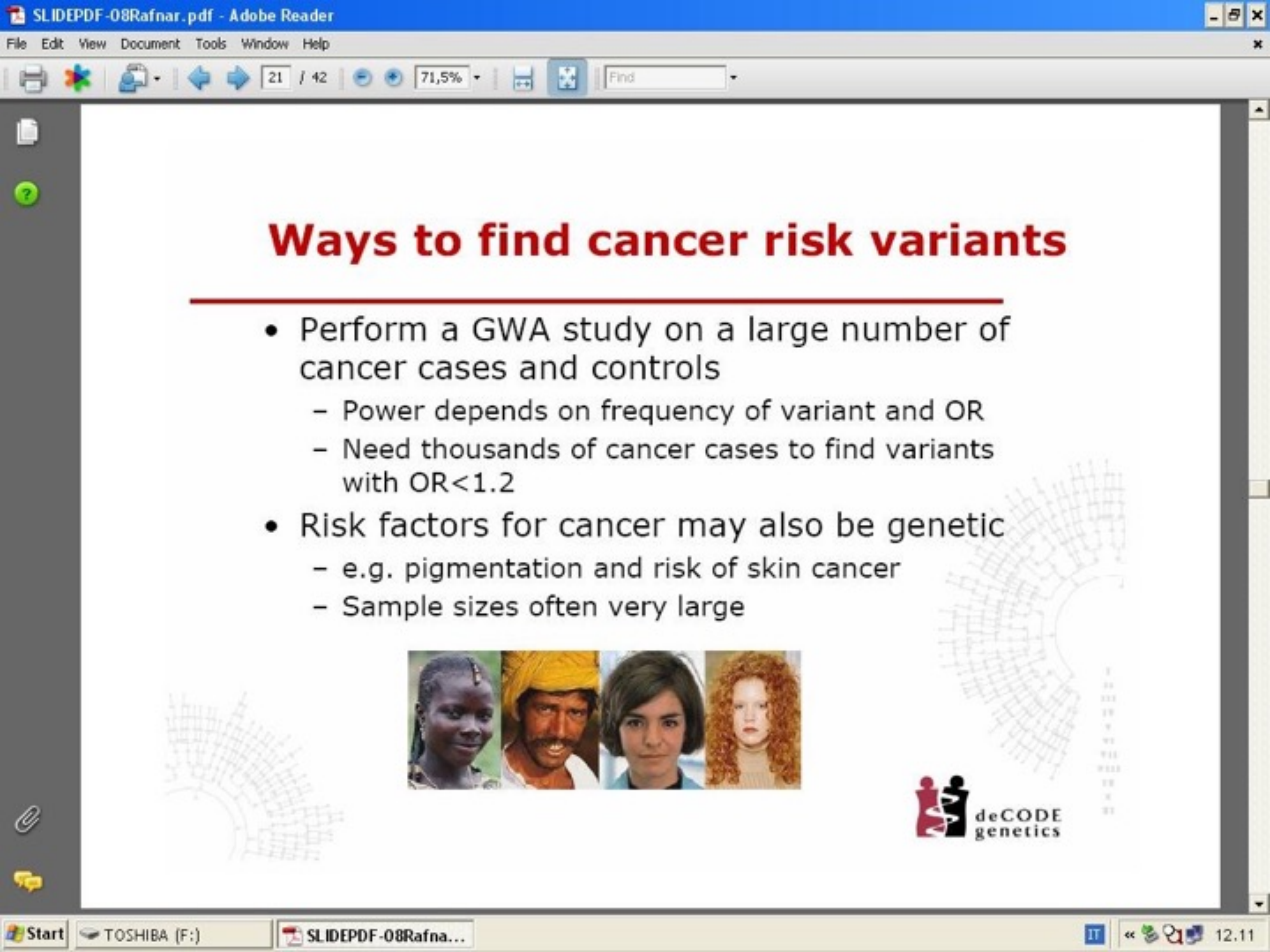
deCODE genetics

Genetic risk assessment model for prostate cancer

- Genotype 13 variants
  - Multiplicative model
  - Results presented as
    - Relative risk of developing the disease
    - Lifetime risk (average 10% in the EU)
- Does not include family history



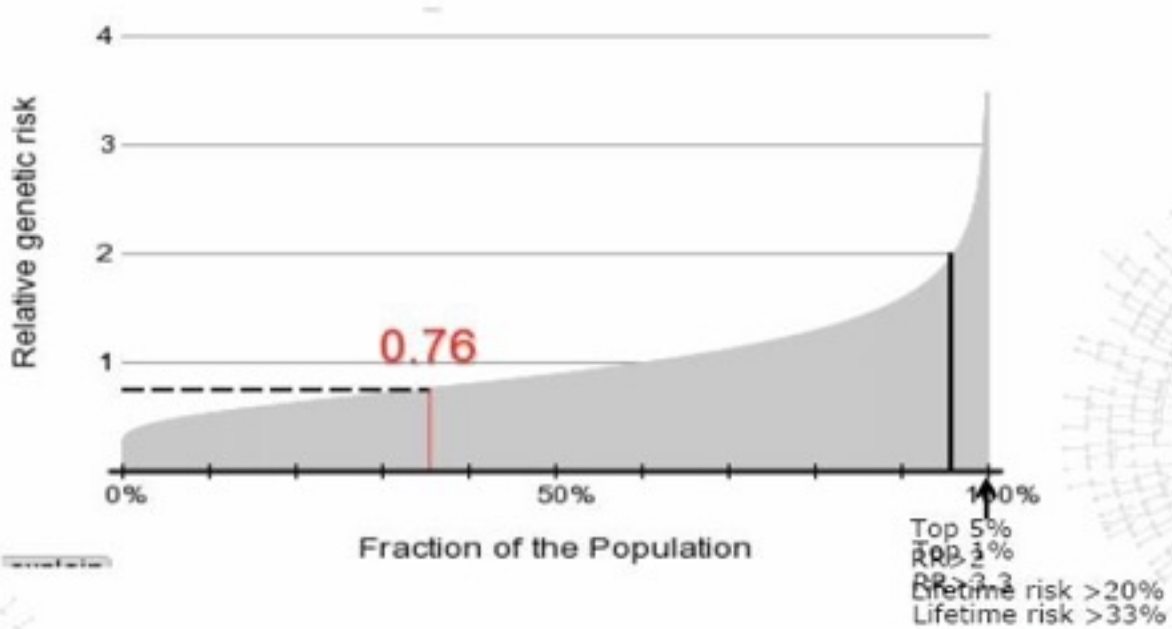
deCODE genetics


SLIDEPDF-08Rafnar.pdf - Adobe Reader
File Edit View Document Tools Window Help
21 / 42
71,5%
Find
Start
TOSHIBA (F:)
SLIDEPDF-08Rafna...
IT
12.11
deCODE genetics

Ways to find cancer risk variants

- Perform a GWA study on a large number of cancer cases and controls
    - Power depends on frequency of variant and OR
    - Need thousands of cancer cases to find variants with OR<1.2
- Risk factors for cancer may also be genetic
    - e.g. pigmentation and risk of skin cancer
    - Sample sizes often very large

Results from prostate cancer risk test



deCODE genetics

Smoking behavior and lung cancer

- Smoking is the major risk factor for lung cancer
  - Over 90% of cases in males and 80% of cases in females attributed to smoking
- Evidence for genetic influence on smoking behaviour and nicotine addiction
- Genetic studies on smoking....

deCODE genetics

Studies on smoking behaviour

- 11,000 smokers; divided into 4 groups based on the number of cigarettes per day (cpd)

  1-10

  11-20

  21-30

  31 or more

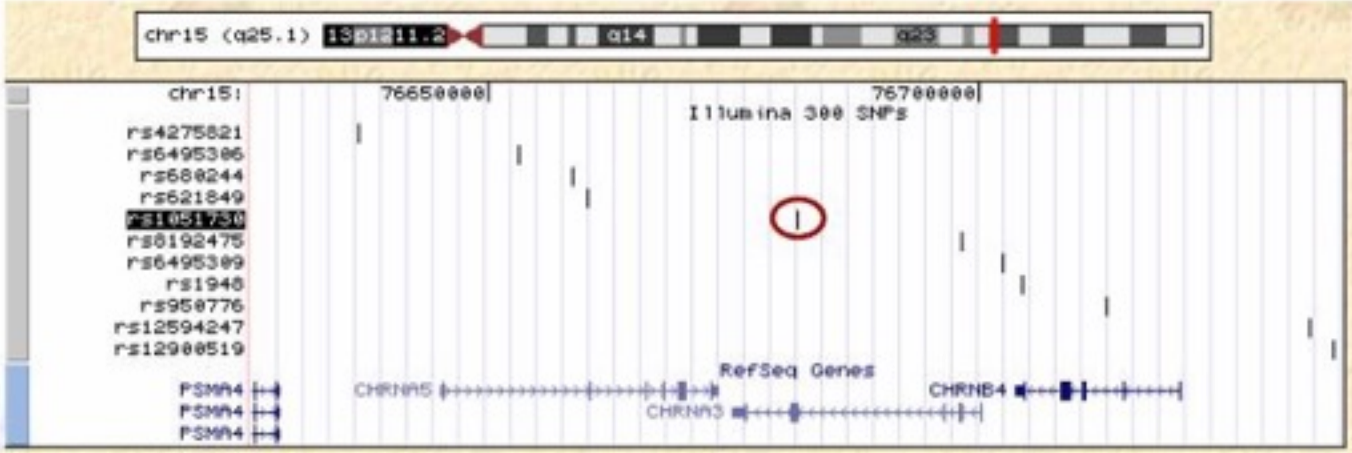
- Genotyped for 370.000 SNPs
  - Search for variants that are more common in heavier smokers

Thorgeirsson et al Nature 2008

deCODE genetics

Variants in nicotine receptor cluster associate with more smoking

6 SNPs in the nicotinic acetylcholine receptor gene cluster on chromosome 15q
Associate with more smoking and nicotine addiction



rs1051730 (T)

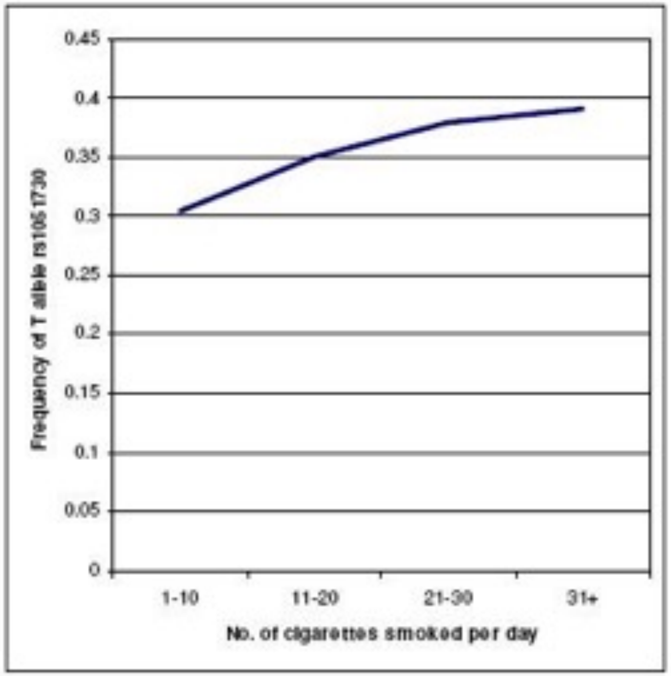
Genetics of smoking

- Results confirmed in an independent group of smokers from
  - Iceland (2950)
  - The Netherlands (1375)
  - Spain (523)
- For all groups combined, regression analysis adjusted for gender and age, $P=6\times10^{-20}$.
  - Each copy of the "risk" variant increases smoking by 1 cigarette per day
- Also associated with nicotine addiction

Thorgeirsson et al Nature 2008

deCODE genetics

Association between rs1051730(T)and number of cigarettes per day



Thorgeirsson et al Nature 2008

deCODE genetics

rs1051730(T) associates with risk of lung cancer

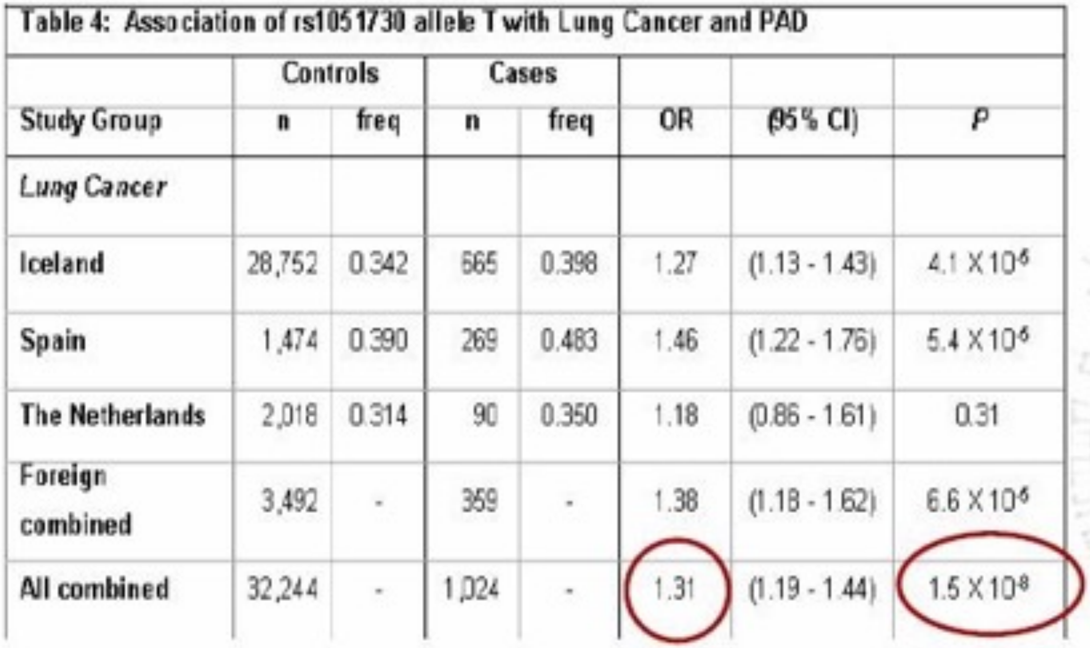
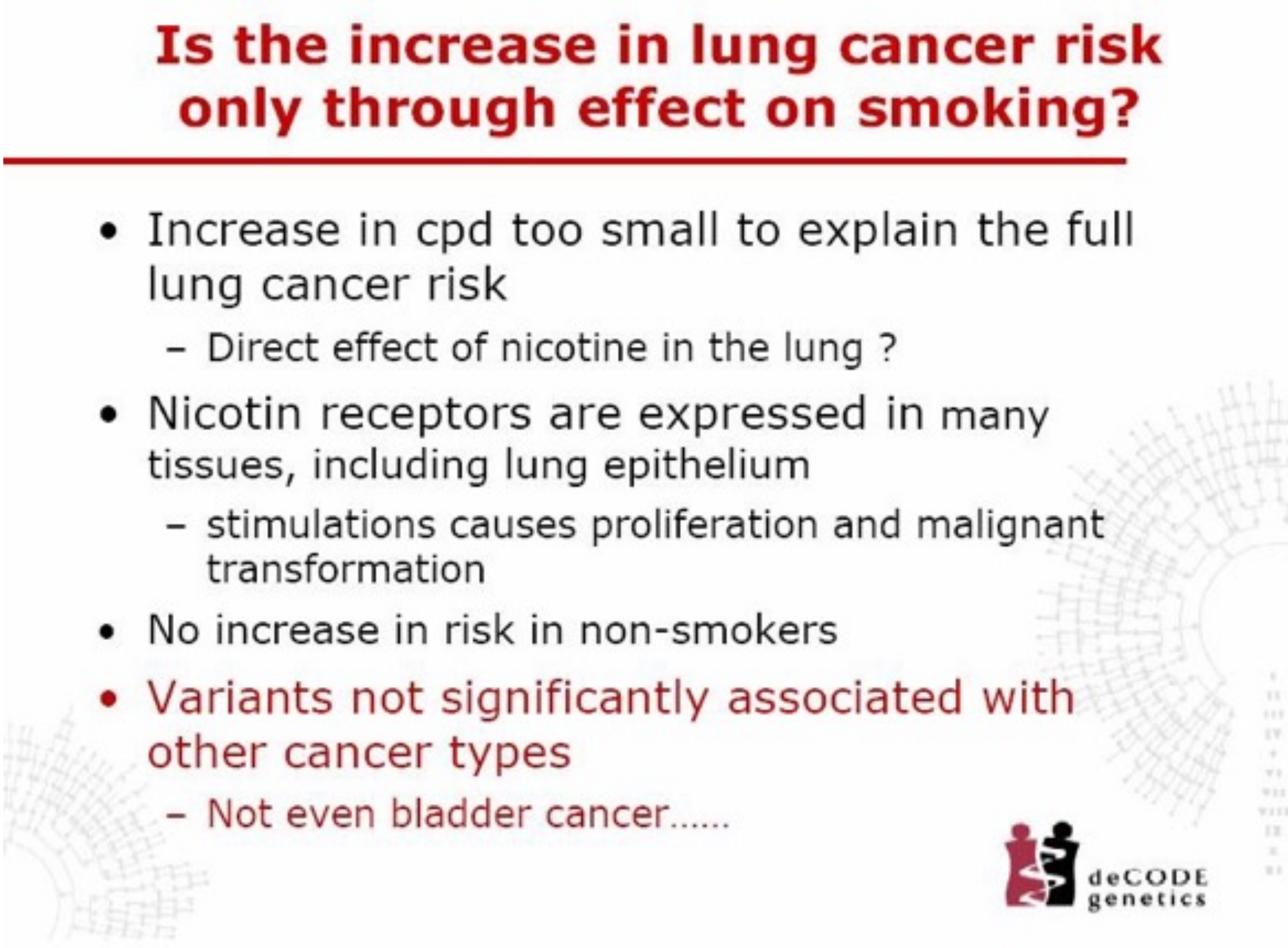


Table 4: Association of rs1051730 allele T with Lung Cancer and PAD

	Controls		Cases				
Study Group	n	freq	n	freq	OR	(95% CI)	*P*
Lung Cancer							
Iceland	28,752	0.342	665	0.398	1.27	(1.13 - 1.43)	4.1×10^{-6}
Spain	1,474	0.390	269	0.483	1.46	(1.22 - 1.76)	5.4×10^{-6}
The Netherlands	2,018	0.314	90	0.350	1.18	(0.86 - 1.61)	0.31
Foreign combined	3,492	-	359	-	1.38	(1.18 - 1.62)	6.6×10^{-6}
All combined	32,244	-	1,024	-	1.31	(1.19 - 1.44)	1.5×10^{-8}

Thorgeirsson et al Nature 2008



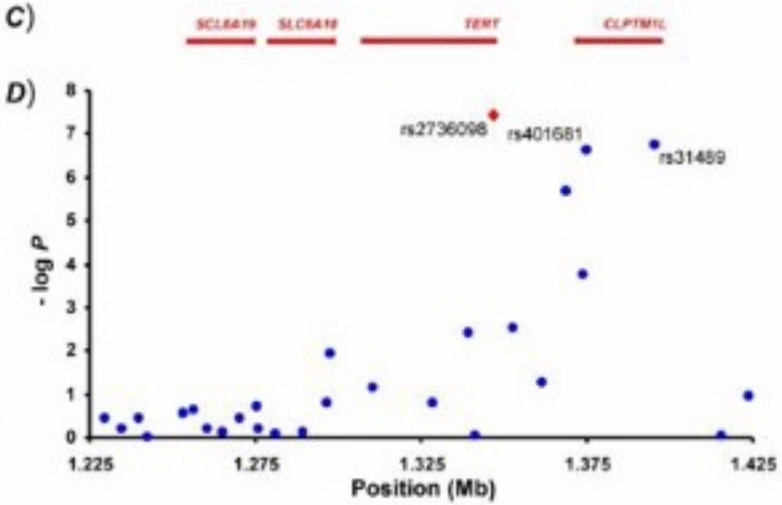
Is the increase in lung cancer risk only through effect on smoking?

- Increase in cpd too small to explain the full lung cancer risk
  - Direct effect of nicotine in the lung ?
- Nicotin receptors are expressed in many tissues, including lung epithelium
  - stimulations causes proliferation and malignant transformation
- No increase in risk in non-smokers
- Variants not significantly associated with other cancer types
  - Not even bladder cancer……

deCODE genetics

Finally, a variant that affects risk of many types of cancer !

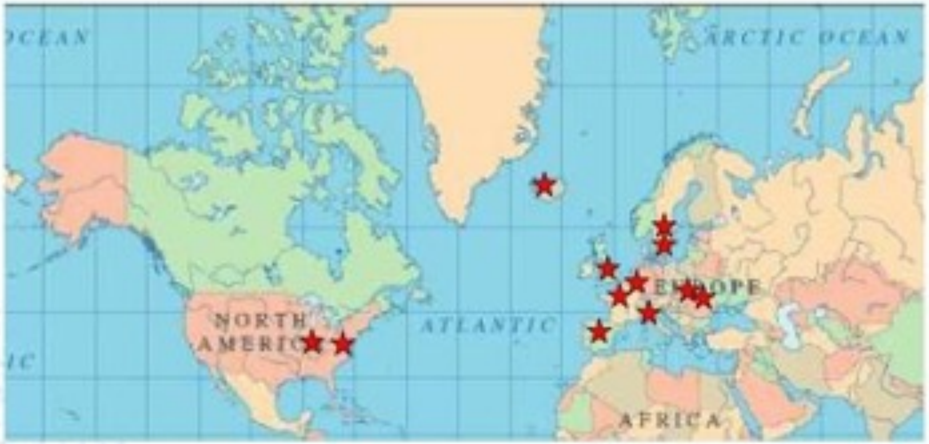
- GWA study on basal cell carcinoma (BCC) identified several regions that associate with increased risk of skin cancer (Stacey et al 2008)
- One on chr5p near two known "cancer genes"
  - *CLPTM1L* (cisplatin resistance related protein) gene
  - **h*TERT*** (human telomerase reverse transcriptase) gene



deCODE genetics

TERT plays a role in the progression of most forms of cancer

- Examine if variation in this region is associated with risk of other cancer types
- Test cancer at **17 cancer sites**, using **30,000** cancer cases and **45,000** controls from Iceland, Europe and USA



deCODE genetics

rs401681 (C) associates with risk of cancer at 5 sites

Study population	Number		Frequency		OR	95% CI	*P* value
	Cases	Controls	Cases	Controls			
Basal cell carcinoma							
Iceland all	2,040	28,890	0.604	0.545	1.27	1.19-1.36	9.5×10^{-12}
Eastern Europe	525	515	0.616	0.575	1.16	0.97-1.39	0.098
All combined	**2,565**	**515**	**0.610**	**0.560**	**1.25**	**1.18-1.34**	$\mathbf{3.7\times10^{-12}}$
Lung cancer							
Iceland all	1,449	28,890	0.575	0.545	1.13	1.04-1.23	3.6×10^{-3}
The Netherlands	529	1,832	0.610	0.570	1.18	1.02-1.35	0.021
Spain	367	1,427	0.582	0.538	1.19	1.01-1.41	0.034
IARC	1,920	2,517	0.617	0.586	1.16	1.06-1.27	8×10-4
All combined	**4,265**	**34,666**	**0.596**	**0.560**	**1.15**	**1.10-1.22**	$\mathbf{7.2\times10^{-8}}$
Bladder cancer							
Iceland all	780	28,890	0.583	0.545	1.16	1.05-1.29	4.5×10^{-3}
The Netherlands	1,277	1,832	0.584	0.570	1.06	0.96-1.17	0.27
UK	707	506	0.564	0.514	1.23	1.04-1.44	0.014
Italy-Torino	329	379	0.550	0.545	1.02	0.84-1.24	0.84
Italy-Brescia	122	156	0.574	0.564	1.04	0.74-1.46	0.82
Belgium	199	378	0.603	0.554	1.22	0.95-1.56	0.11
Eastern Europe	214	515	0.619	0.575	1.20	0.96-1.51	0.12
Sweden	346	905	0.545	0.521	1.10	0.92-1.31	0.30
Spain	173	1,427	0.546	0.538	1.03	0.83-1.29	0.78
All combined	**4,147**	**34,988**	**0.578**	**0.535**	**1.12**	**1.06-1.18**	$\mathbf{5.7\times10^{-5}}$
Prostate cancer							
Iceland all	2,276	28,890	0.569	0.545	1.10	1.03-1.17	3.75×10^{-3}
The Netherlands	994	1,832	0.576	0.570	1.02	0.92-1.14	0.67
Chicago, US	635	693	0.581	0.568	1.06	0.90-1.23	0.49
Spain	459	1,427	0.559	0.538	1.09	0.94-1.26	0.27
CGEMS	5,109	5,059	0.558	0.543	1.06	1.00-1.11	0.036
All combined	**9,473**	**37,901**	**0.569**	**0.553**	**1.07**	**1.03-1.11**	$\mathbf{3.6\times10^{-4}}$
Cervical cancer							
Iceland all	369	28,890	0.611	0.545	1.31	1.13-1.51	2.6×10^{-4}

deCODE genetics