Введение в эволюционную и медицинскую геномику, часть II

Лекция 3

ФББ МГУ, весна 2008

Журнальный клуб 18 марта:

1. Jimenez-Sanchez G, Childs B, Valle D. Human disease genes. *Nature*. 2001 Feb 15;409(6822):853-5.
2. Steward RE, MacArthur MW, Laskowski RA, Thornton JM. Molecular basis of inherited diseases: a structural perspective. *Trends Genet*. 2003 Sep;19(9):505-13. Review.
3. Di Rienzo A. Population genetics models of common diseases. *Curr Opin Genet Dev*. 2006 Dec;16(6):630-6.
4. Kryukov GV, Pennacchio LA, Sunyaev SR. Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. ***Am J Hum Genet***. 2007 Apr;80(4):727-39.
5. Levy S, et al. The diploid genome sequence of an individual human. ***PLoS Biol***. 2007 Sep 4;5(10):e254.

E-mail: ramensky@imb.ac.ru

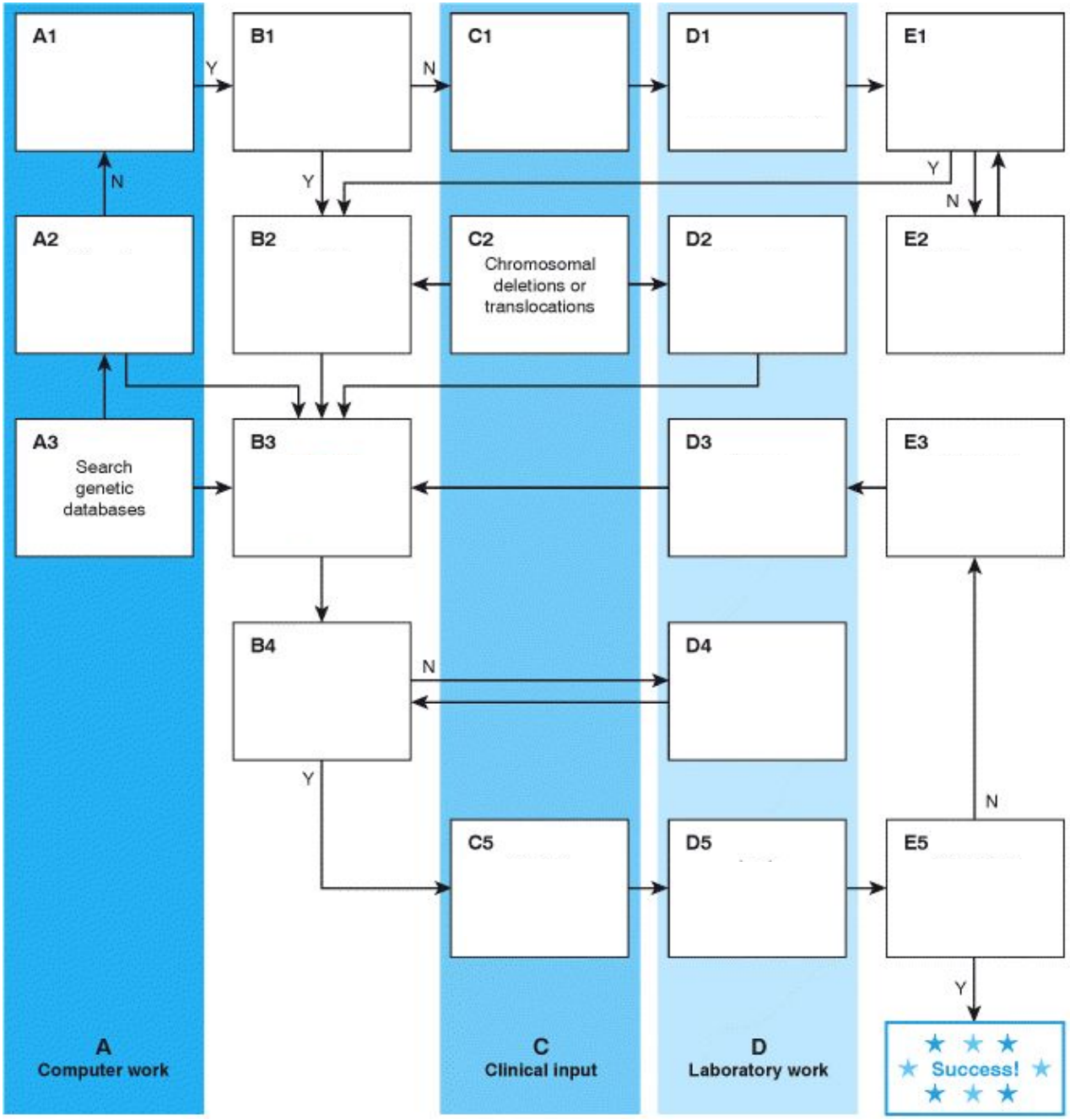


Figure 15.1. How to identify a human disease gene. There is no single pathway to success, but the key step is to arrive at a plausible candidate gene, which can then be tested for mutations in affected people.

"Human Molecular Genetics 2",
Strachan and Read

Стратегии поиска генов заболеваний

J. Pasternak. *An Introduction To Human Molecular Genetics - Mechanisms Of Inherited Diseases* (Wiley, 2005)

Functional/Candidate Gene Cloning: "Either a known protein that is responsible for an inherited disorder or a protein that is considered a likely candidate based on the symptoms and biochemistry of the disease" (Protein -> Gene seq -> Location, PCR, mutations, etc.)

Positional-Candidate Gene Cloning: "The disease gene is mapped to a chromosome location with polymorphic markers. Once the chromosome location is narrowed down, the human genome database is consulted for the genes within this region. From this list of genes, a likely candidate(s) is selected and a mutation detection assay is run with PCR probes based on the gene sequence derived from the database"

Сложные заболевания: примеры

- Аутизм
- Астма
- Диабет
- Ожирение
- Алкоголизм
- Гипертония
- Шизофрения

Особенности сложных заболеваний // Где тут легкий обман?

1. **Неполная пенетрантность** (Не всякая генетическая предрасположенность проявляется как болезнь)
2. **Различный возраст развития заболевания**
3. **Действие факторов окружающей среды** (Образ жизни, диета)
4. **Полигенное наследование, в т.ч. эпистатические взаимодействия**

Легкий обман: **пп. 1-4. характерны также для моногенных заболеваний, так что остается**

***Сухой остаток*: С.з. имеют наследственную компоненту, но наследуются более сложно, чем «по Менделю»**

AN INTRODUCTION TO

Human Molecular Genetics

Mechanisms of Inherited Diseases

Second Edition

Jack J. Pasternak
University of Waterloo
Ontario, Canada

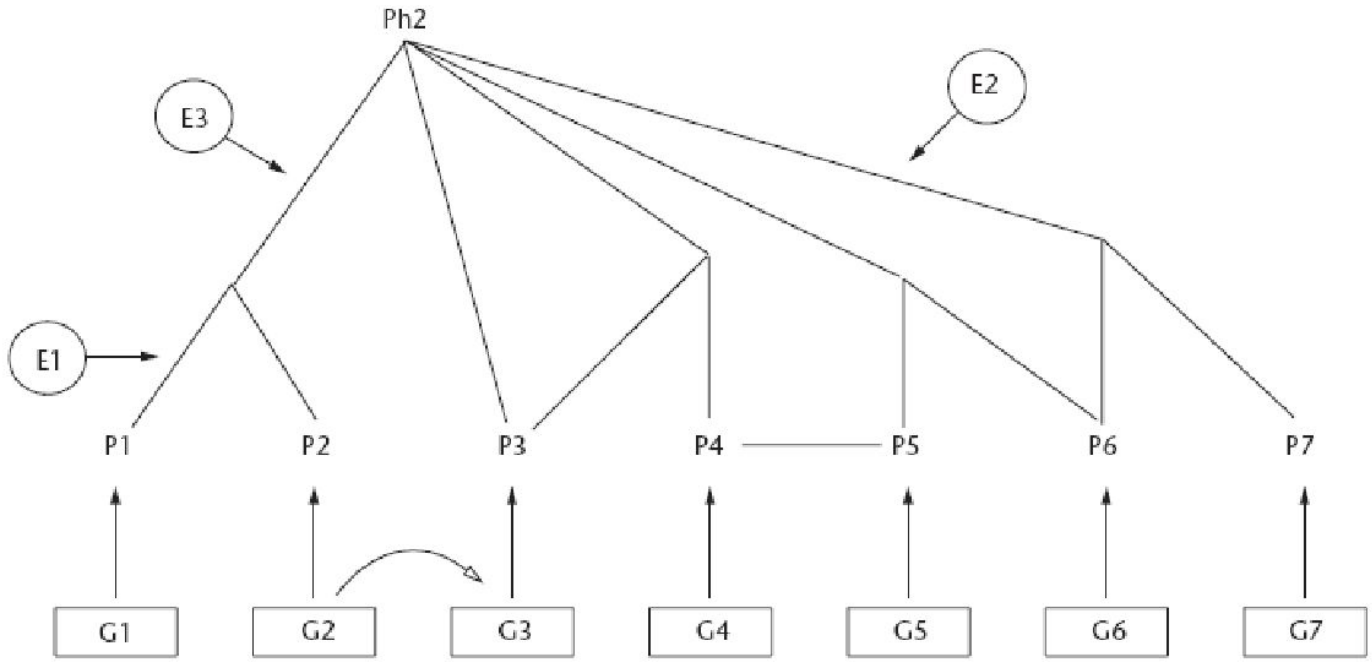


Figure 10.1 Schematic representation showing how phenotypes of a multifactorial trait are determined. (A) The primary products (P1–P7) of seven normal genes (G1–G7) form various protein-protein assemblies and with environmental factors (E1, E2) produce the Ph5 phenotype. The open arrow indicates interaction (epistasis) between genes 2 and 3. (B) The phenotypic outcome (Ph2) is due to the

Evolution revisited: упрощенный взгляд на болезнетворные аллели

Моногенные заболевания

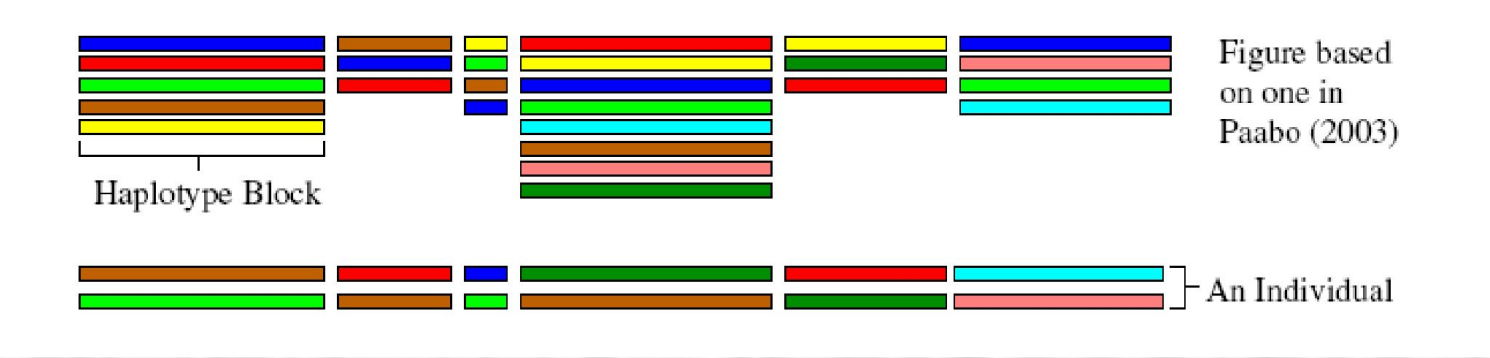
- Сильный эффект аллелей
- Мутационно-отборное равновесие

Сложные заболевания

- Слабый эффект аллелей => слабый отбор
- Давление отбора в некоторых случаях изменялось в ходе эволюции образа жизни и среды обитания

“Common disease / common variant”

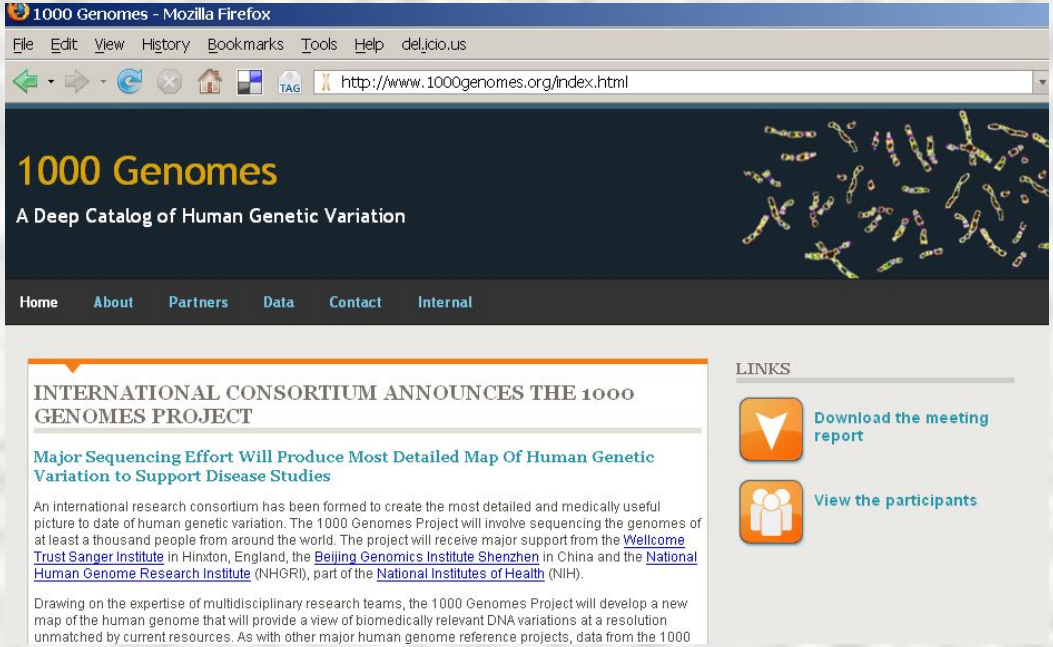
CD/CV Hypothesis: Частые аллели (>1%) составляют значительную долю среди аллелей предрасположенности, и для их поиска применимы ассоциативные исследования, например, с помощью *HapMap**:



** Громкий проект 2002 г.*

“Common disease / rare variant”

CD/RV Hypothesis: Редкие аллели (<1%) составляют основную долю среди аллелей предрасположенности, и нужно *ресеквенирование**



** Громкий проект 2008 г.*

Подтверждения гипотезы CD/CV ?

Table 3 • Some polymorphic, moderate to low risk variants

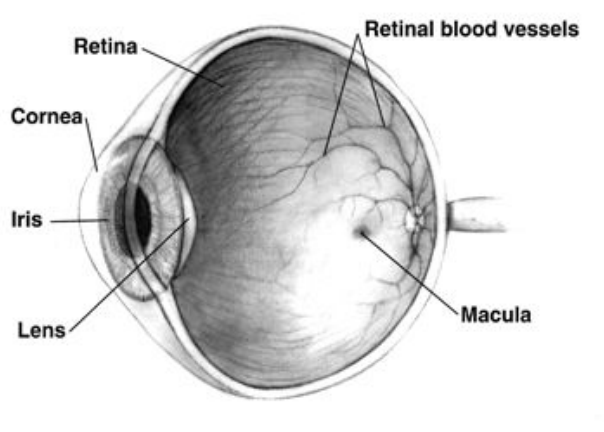
Disease*	Locus	Change	Frequency	Relative risk
Alzheimer's disease	*APOE–APOE4*	C112R	0.09–0.22	4.0–15.0
	APOE–APOE2	C158R	0.04–0.08	0.5
Thrombosis	factor V Leiden	R506Q	0.00–0.08 (Eur)	5.0–10.0
Hemochromatosis	*Hfe*	H63D	0.02–0.22	4.0
NIDDM	*PPARγ*	P12A	0.85 (Eur)	1.25
IDDM	*INS*	*VNTR* (promoter)	0.85 (Eur)	1.5–2.5
HIV	*CCR5*	Δ32	0.01–0.14 (Eur)	high (resistance), moderate (nonprogression)
Crohn's disease	*NOD2/CARD15*	1007fs	0.02 (Eur)	6.0
		G908R	0.01 (Eur)	6.0
		R702W	0.04 (Eur)	3.0
Breast cancer	*BRCA2*	N372H	0.25 (Eur)	1.3
Colon cancer	*APC*	I1307K	0.03 (AJ)	2.0
Neural tube defects	*MTHFR*	C677T (A→V)	0.30 (Eur)	2.0
		A1298C (E→A)	0.30 (Eur)	2.0
FMF	*MEFV*	P369S	0.02 (AJ)	7.0
		E148Q	0.06 (AJ)	3.0
Graves disease	*CTLA4*	T17A	0.35 (Eur)	1.5–2.0
Creutzfeld-Jakob	*PRNP*	M129V	0.65 (Eur)	3.0
Autoimmune diseases	*HLA B, DR, DQ*	numerous amino acid substitutions	polymorphic	low to moderate

*AJ, Ashkenazi Jews; Eur, European; FMF, familial Mediterranean fever; IDDM, insulin-dependent diabetes mellitus; NIDDM, non-insulin-dependent diabetes mellitus.

Botstein & Risch *Nat Genet* (2003) **33**:228

Age-related macular degeneration

Характерные образования в центральной части (macula) сетчатки, вызывающие нарушения зрения



Macular degeneration gene: The genes for the complement system The genes for the complement system proteins factor H (CFH) and factor B (CFB) have been determined to be strongly associated with a person's risk for developing macular degeneration. The mutation in CFH(**Tyr402His**) reduces the affinity of CFH for CRP and probably also alters the ability of factor H to recognise specific glycosaminoglycans.

Recurrence ratios for siblings of an affected individual are three- to sixfold higher than in the general population, but

Family-based analysis has resulted in only modestly significant evidence for linkage.

Age-related macular degeneration

J.Maller *et al.*, *Nat Genet* (2006) **38**(9):1055-9:

Генотипирование 1,536 tagSNPs у 1,238 пациентов и 934 контролей

Table 1 Association between *CFH* variants and age-related macular degeneration

SNP	Allele	Control freq.	Affected freq	χ^2	*P* value
rs1061170[a] Риск: 7.6/2.7	C	0.359	0.615	264.5	1.79×10^{-59}
rs1410996	C	0.571	0.808	272.9	2.65×10^{-61}
rs1061170, rs1410996					3.7×10^{-64}
	CC	0.356	0.609 Риск: ~15		
	TC	0.213	0.193		
	TT	0.428	0.194		

Two-letter allele symbols are used to indicate associated haplotypes of multiple SNPs.
[a]rs1061170 is the variant of *CFH* encoding the Y402H protein variant.

Table 2 Association between *LOC387715* variants and age-related macular degeneration

SNP	Allele	Control freq.	Affected freq.	χ^2	Nominal *P* value
rs10490924	T	0.194	0.455	315.075	1.71×10^{-70}

Age-related macular degeneration

J.Maller *et al*., *Nat Genet* (2006) **38**(9):1055-9:

Полиморфизм в локусе, содержащем гены системы комплемента *C2* и *CFB*, также ассоциируется с AMD:

- BF:R32Q + rs547154 (C2:intron10)
- BF:L9H + C2:E318D

Эти два редких протективных варианта принадлежат разным гаплотипам

Table 3 Association between ***CFB*** and ***C2*** variants and age-related macular degeneration

SNP	Allele	Control freq.	Affected freq.	χ^2	Nominal *P* value	*P* value conditional on rs9332739	*P* value con
rs9332739[a]	C	0.054	0.026	23.6	1.10×10^{-6}	X	9
rs547154	A	0.097	0.054	27.8	1.30×10^{-7}	6.09×10^{-8}	
rs4151667	A	0.050	0.029	13.8	0.0002	0.0397	6
rs641153[b]	T	0.102	0.045	51.9	5.5×10^{-13}	2.99×10^{-13}	

[a]rs9332739 is the variant of *C2* encoding the E318D protein variant, highly correlated to rs4151667 (BF:L9H). [b]rs641153 is the variant of *BF* encoding the R32Q protein correlated to rs547154.

Age-related macular degeneration

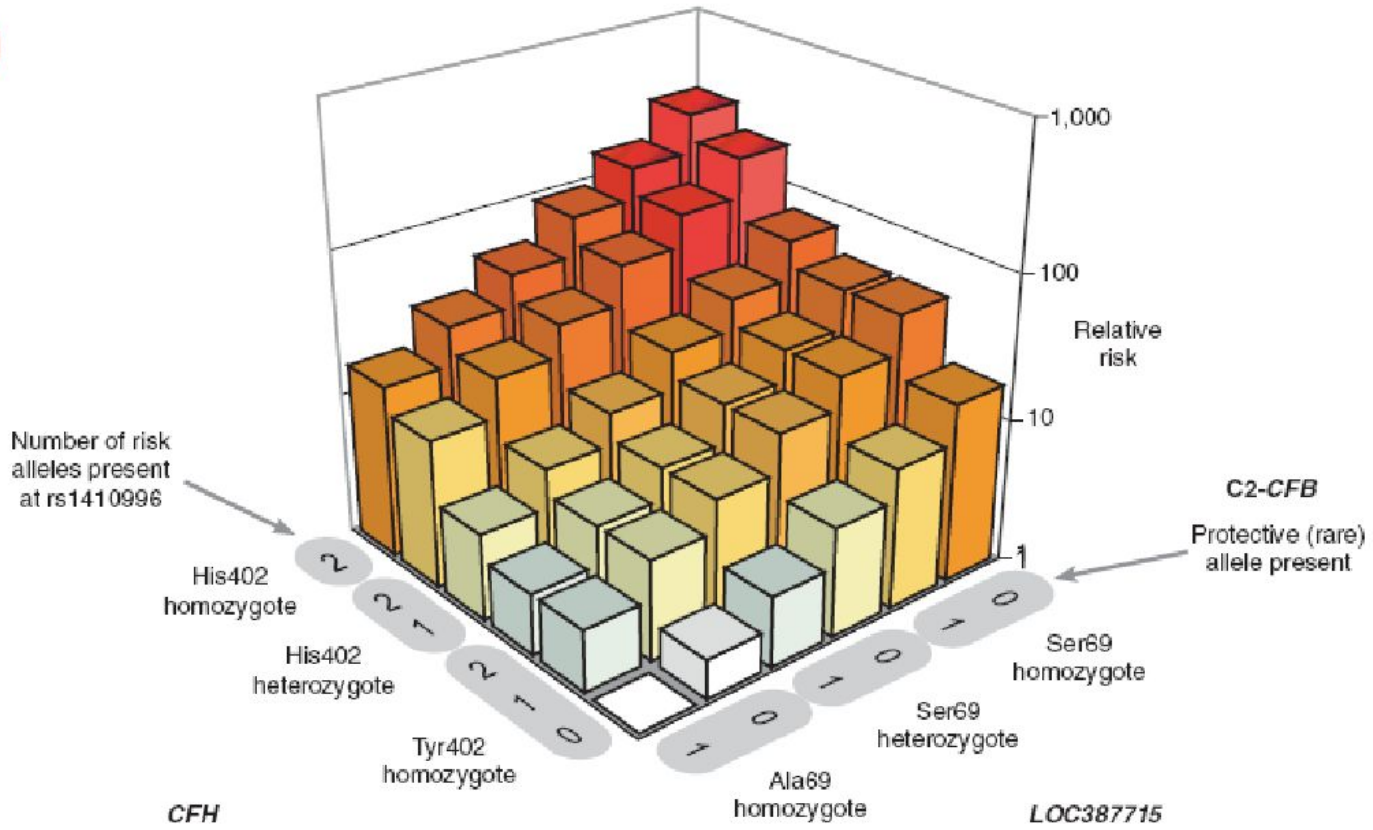


Figure 1 Relative risk plotted as a function of the genetic load of the five variants that influence risk of AMD. Two variants are in the *CFH* gene on chromosome 1: Y402H and rs1410996. Another common variant (A69S) is in hypothetical gene *LOC387715* on chromosome 10. Two relatively rare variants are observed in the *C2* and *BF* genes on chromosome 6. We find no evidence for interaction between any of these variants, suggesting an independent mode of action. (See also **Supplementary Table** 4.)

Age-related macular degeneration

Выводы:

(i) there can **exist common alleles of substantial effect** on common disease (in AMD, these explain at least half of all risk to siblings);
(ii) these can be found outside of 'candidate genes' and **outside of coding regions**;
(iii) for such alleles, **association offers much greater power** and reproducibility than linkage (association results much stronger than similarly sized linkage studies);
(iv) even for late-onset diseases with partial heritability, **common genotypes can strongly influence individual risk**;
(v) there **need not be epistasis** among, or phenotypic sub-stratification by, genes of substantial population effect.

J.Maller *et al*., *Nat Genet* (2006) **38**(9):1055-9

Влияние редких аллелей на уровень холестерина высокой плотности

Низкой уровень холестерина высокой плотности (HDL-C, «хорошего» холестерина): основной фактор риска сердечно-сосудистых заболеваний

Гомозиготы по некоторым мутациям в этих белках не имеют HDL-C вовсе:

1. **Apolipoprotein AI (APOA1)**, the major protein component of HDL;
2. **The adenosine triphosphate binding cassette transporter A1 (ABCA1):** efflux of cholesterol from cells to HDL particles;
3. **Lecithin cholesterol acyltransferase (LCAT):** catalysis of the formation of cholesteryl esters in HDL

Влияние редких аллелей на уровень холестерина высокой плотности

The hypothesis: “rare sequence variations contribute significantly to low plasma levels of high density lipoprotein cholesterol (HDL-C)”

Results: “Of the 128 individuals with low plasma levels of HDL-C, 21 (16%) had sequence variants not present in the high HDL-C group. In contrast, only 3 (2%) of the individuals in the high HDL-C group had sequence variants not found in the low HDL-C group.

Thus, one of six individuals with HDL-C levels below the fifth percentile in the Dallas Heart Study had a rare mutation in *ABCA1* or *APOA1*”

J.Cohen *et al*., *Science* (2004) **305**:869-72

Влияние редких аллелей на уровень холестерина высокой плотности

Table 1. Sequence variations in the coding regions of *ABCA1*, *APOA1*, and *LCAT*. Values represent the numbers of sequence variants identified in 256 individuals from the Dallas Heart Study (DHS) (128 with low HDL-C and 128 with high HDL-C) and 263 Canadians (155 with low HDL-C and 108 with high HDL-C) (*17*). NS, nonsynonymous (nucleotide substitutions resulting in an amino acid change); S, synonymous (coding sequence substitutions that do not result in an amino acid change). GenBank accession numbers for DHS *ABCA1*, *APOA1*, and *LCAT* sequences are NM_005502, NM_000039, and NM_000229, respectively.

	Sequence variants unique to one group				Sequence variants common to both groups	
	Low HDL-C		High HDL-C			
	NS	S	NS	S	NS	S
			DHS			
ABCA1	14	6	2	5	10	19
APOA1	1	0	0	1	0	1
LCAT	0	1	1	0	1	1
			Canadians			
ABCA1	14	2	2	3	7	5
APOA1	0	1	0	0	2	0
LCAT	6	1	0	0	0	0

J.Cohen *et al.*, *Science* (2004) **305**:869-72