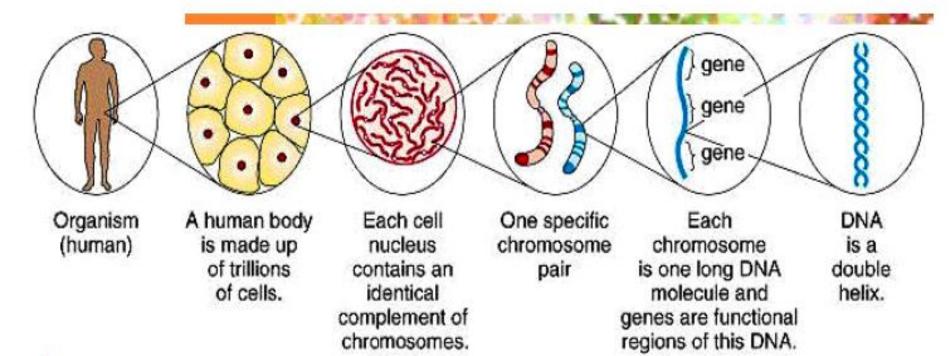
Online Genetic Variation Databases

Department of Biomathematics University of Thessaly, Larissa November 2011

Genetics

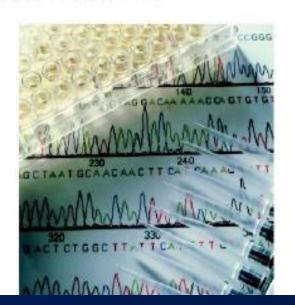
Scientific study of how physical, biochemical, and behavioral traits are transmitted from parents to their offspring. In essence, the science of heredity.

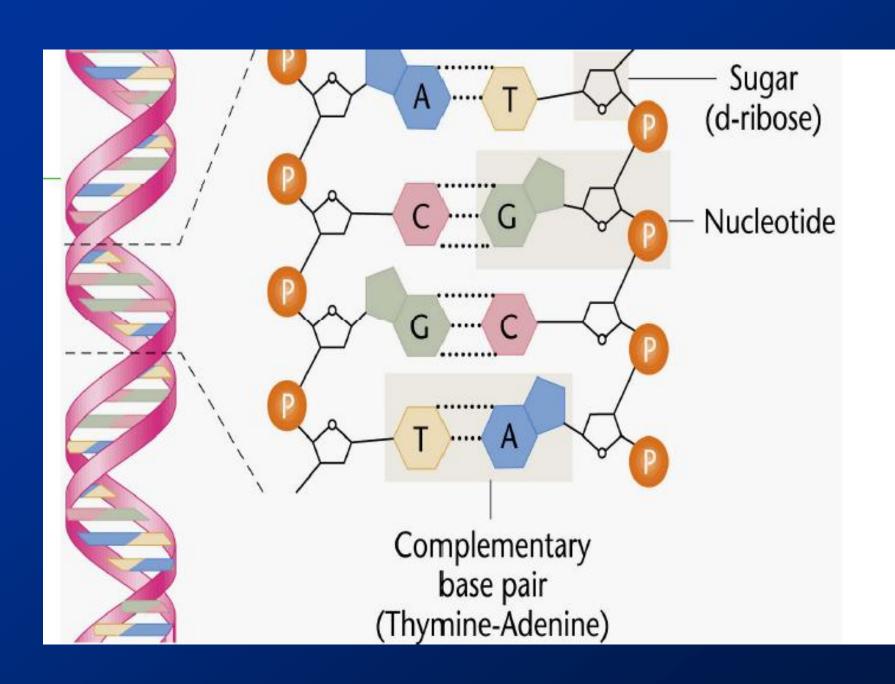


Genetic information contained in DNA

DNA is composed of nucleotides

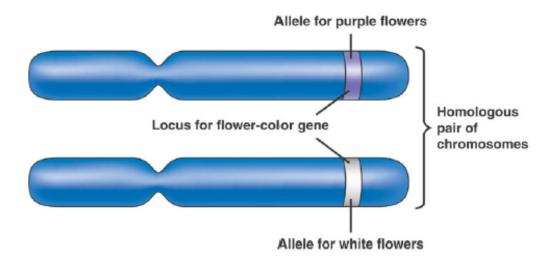
A gene is a unique sequence of nucleotides





Central Dogma Transcription Translation mRNA Protein Gene (DNA)

Some Genetic Terms



- Genes are instructions for producing a trait
- Locus is the spot each genes has on a chromosome
- Diploid cells have two genes (a gene pair) for each trait, each on a homologous chromosome
- Alleles are various molecular forms of a gene encoding for the same trait (i.e. flower color)

Τι είναι γονίδιο

- Βασική μονάδα κληρονομικότητας
- Ακολουθία βάσεων η οποία μεταφέρει την πληροφορία που απαιτείται για να δημιουργηθεί μια συγκεκριμένη πρωτεΐνη
- Ένα γονίδιο κωδικοποιεί μια πρωτεΐνη ή ένα μόριο RNA

Structure of a Gene Exon 1 Exon 2 Exon 3 Exon 4 Intron 1 Promoter Intron 2 Intron 3 Gene (DNA) Transcription Primary transcript (RNA) Splicing Mature transcript (mRNA) Protein synthesis Protein

@ Wellcome Trust

HUMAN INDIVIDUALITY





DNA sequence of all human beings is 99.9% identical

- Our DNAs differ by 0.1%.
- ⊃ Does it make a difference ?

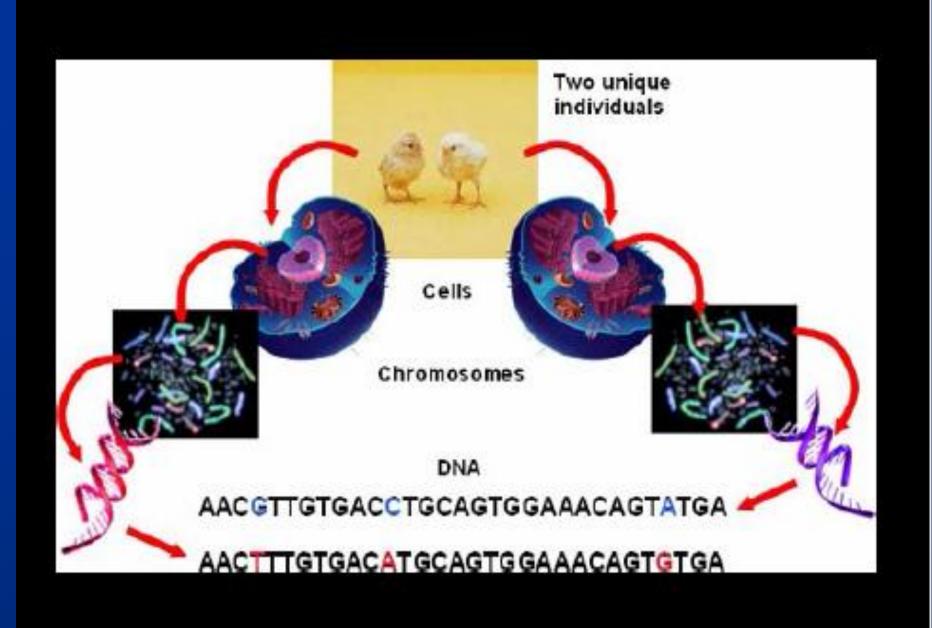


0.1% difference translates into 3 million separate "spelling" differences in a genome of 3 billion bases

Small variations in genotype can make big differences to phenotype (1.24%)

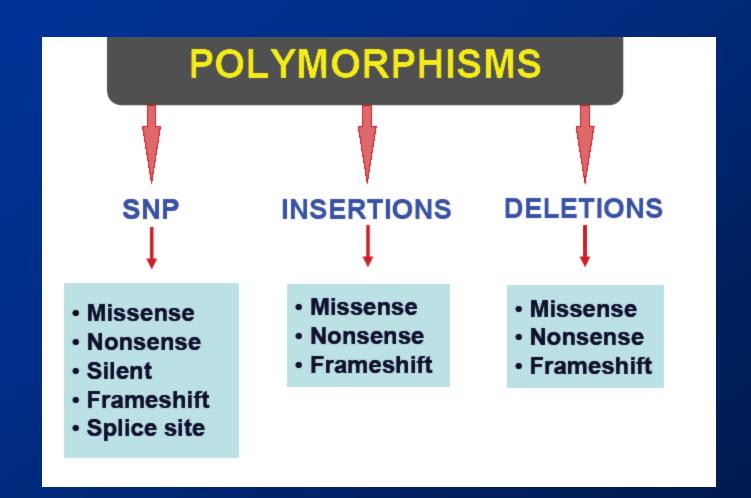






definitions

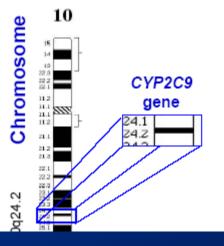
- Polymorphism: a variation in nucleotide sequence at a given position in the genome, seen in the population at a frequency of 1% or higher
- Mutation: a change in nucleotide sequence which can occur in germ line cells or somatic cells. In germ line cells, mutations can be passed on to offspring who inherit the mutation. Mutations generally have a frequency of less than 1% in the population.
- Allele: the specific variant (nucleotide) of a polymorphism at a single locus on a chromosome; for example A, C, G or T



Single nucleotide polymorphisms (SNP)

Most common, Incidence 1 per 300 - 600bp

Example:



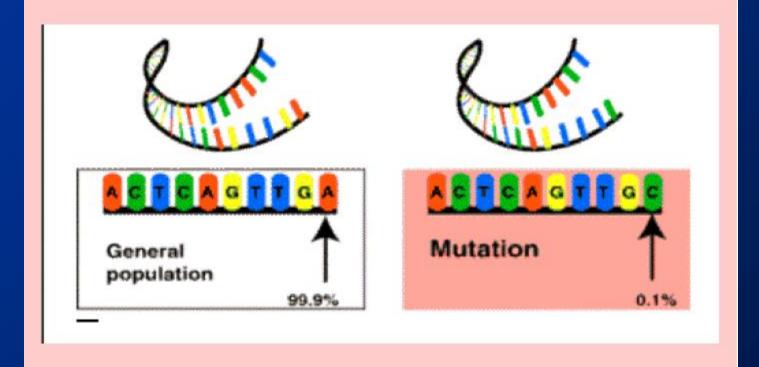
Nucleotide position 430



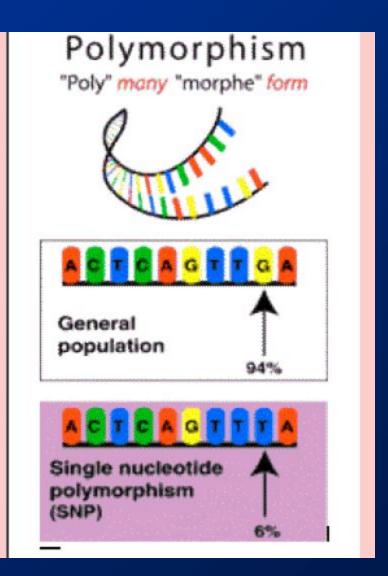


Decreased enzyme activity

Mutation



SNPs (pronounced snips)



Mendelian Traits in Humans



Recessive phenotype: Albinism

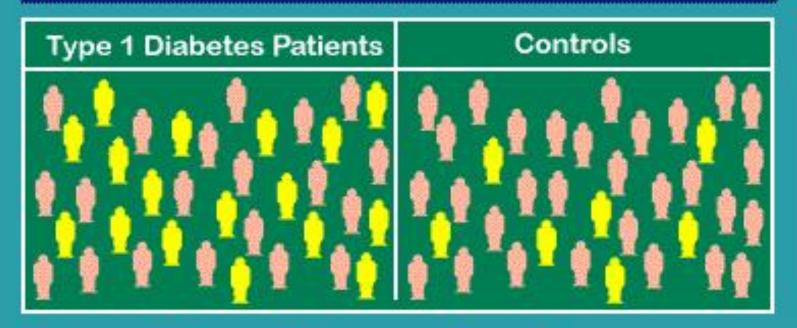


Dominant phenotype: Achondroplasia

 However, common Multifactorial disorders result from polymorphisms in multiple genes, often coupled with environmental causes. The complicated bases of these diseases make them difficult to study and to treat. Heart disorder, diabetes and cancer are examples of this type of disorder.

- SNPs may fall within coding sequences of genes, non-coding regions of genes, or in the intergenic regions between genes. SNPs within a coding sequence will not necessarily change the aminoacid sequence of the protein that is produced, due to degeneracy of the genetic code. A SNP in which both forms lead to the same polypeptide sequence is termed synonymous if a different polypeptide sequence is produced they are non-synonymous.
- Variations in the DNA sequences of humans can affect how humans develop diseases and respond to pathogens, chemicals, drugs, vaccines, and other agents. However, their greatest importance in biomedical research is for comparing regions of the genome between cohorts (such as with matched cohorts with and without a disease).
- Genetic association studies

Association Studies



Genotype	Type 1	Controls	Total
HLA DR4	17	7	24
NON-HLA DR4	20	30	50
	37	37	

 $\chi_{.05}^2 = 5.377$

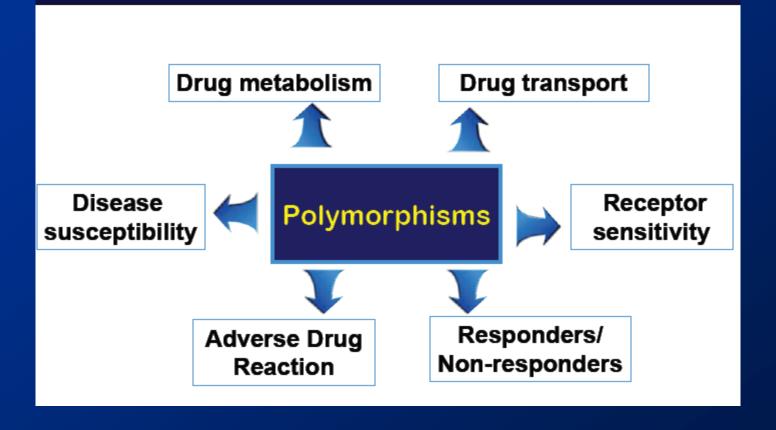
p< 0.025

Odds Ratio: 3.6 95% CI = 1.3 to 10.4

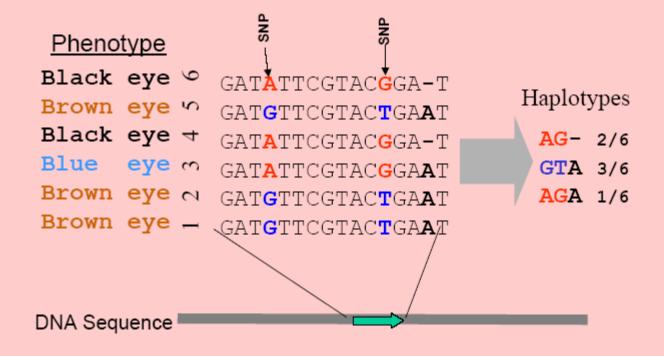
= non-HLA DR4



Consequences of polymorphisms



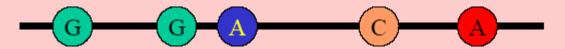
From SNP to Haplotype



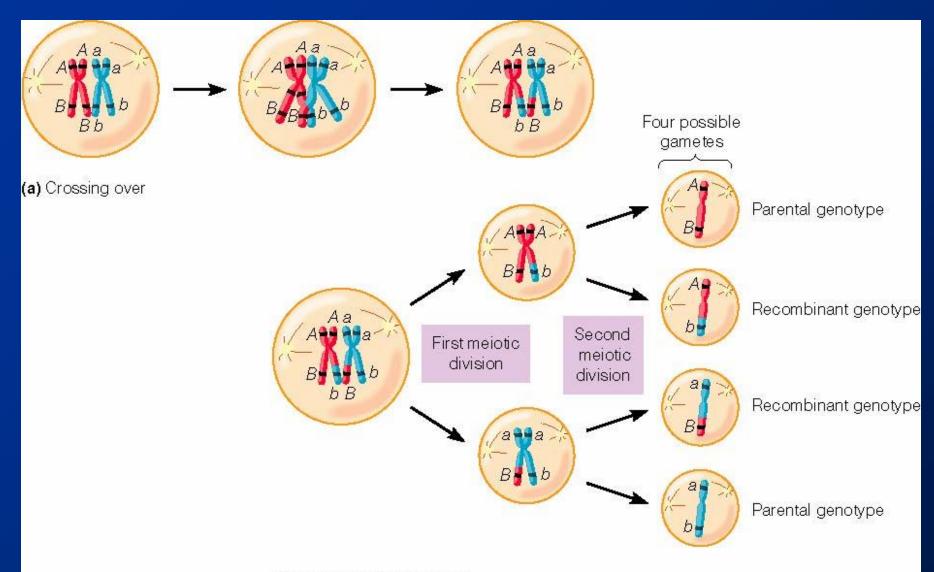
 Haplotype: the combination of two or more <u>alleles</u> at physically neighbouring positions within the <u>genome</u>

Haplotype

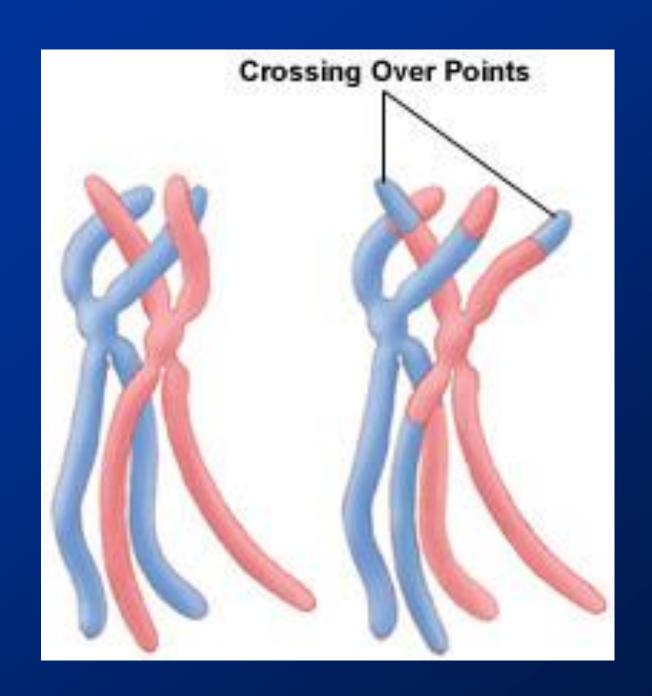
 Haplotype: A set of closely linked genetic markers present on one chromosome which tend to be inherited together (not easily separable by recombination).



Set of SNP polymorphisms: a SNP haplotype



(b) Results of crossing over



Linkage disequilibrium

 The occurrence together on the same chromosome of specific alleles at closely linked loci more frequently than would be expected by chance. Because linkage disequilibrium is a function of the distance between loci, it can be used to help infer the order of genes on chromosomes.

• LD metrics: D, r², LOD

PERSPECTIVES

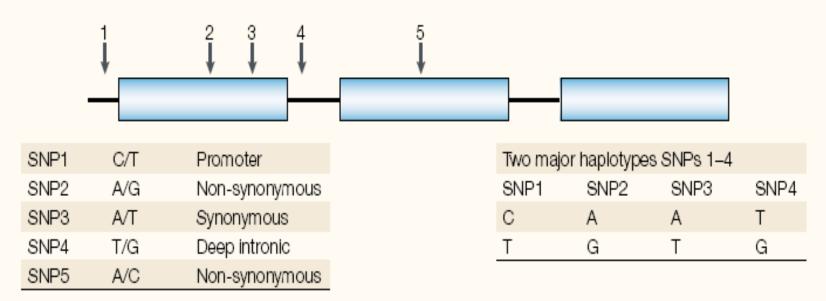
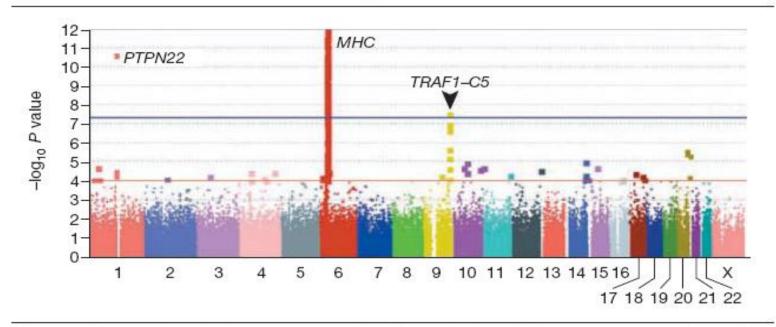


Figure 1 | Haplotypes and linkage disequilibrium in single-nucleotide-polymorphism selection. Single-nucleotide polymorphisms (SNPs) 1–4 are in linkage disequilibrium and form two common haplotypes, which can be characterized by any of the variants. Therefore, it is only necessary to determine the genotype of one of the SNPs to capture the information about all four SNPs. SNP5 is not in linkage disequilibrium with SNPs 1–4 and is not part of the haplotype, but it might still contribute independently to the risk of the phenotype.

Genome wide association studies

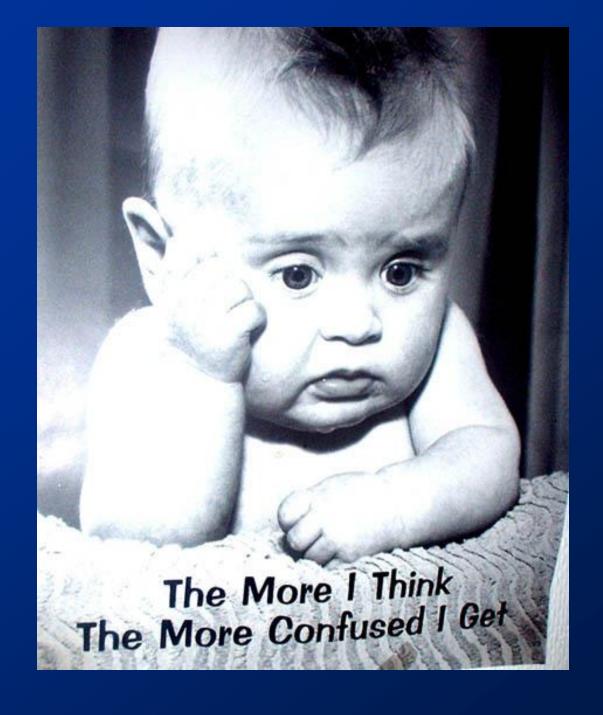
Figure 3. Genome-wide Association Findings in Rheumatoid Arthritis



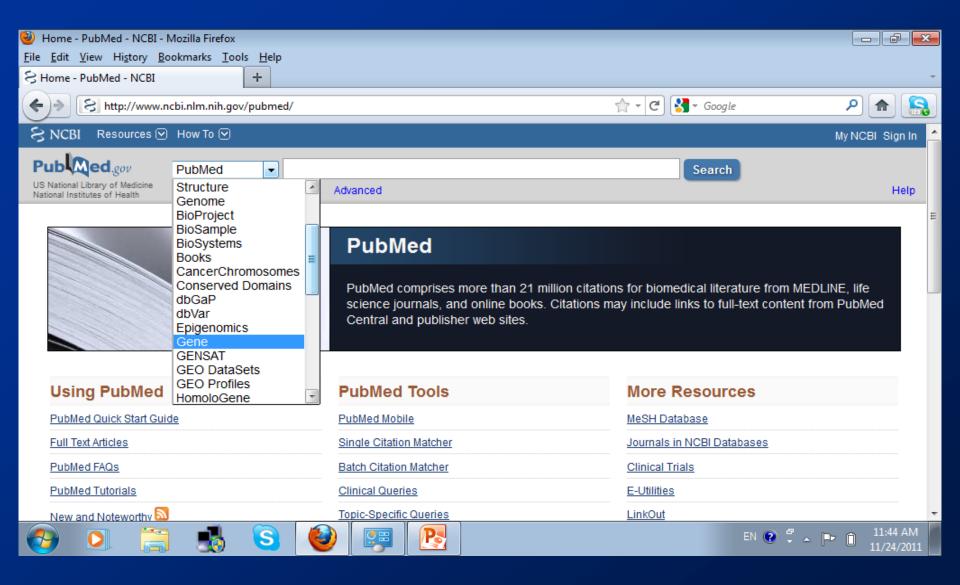
Genome-wide association studies assume a priori hypotheses about candidate genes or regions that might be associated with disease; rather, they test single-nucleotide polymorphisms (SNPs) throughout the genome for possible evidence of genetic susceptibility. Associations plotted as $-\log_{10}P$ values for a genome-wide association study in 1522 cases with rheumatoid arthritis and 1850 controls, showing single data points for SNPs with $P < 10^{-4}$ (lower horizontal red line) for 22 autosomes and the X chromosome. The predefined level of significance, at 5×10^{-8} is shown with a horizontal blue line. SNPs at PTPN22 on chromosome 1, the major histocompatibility comples (MHC) on chromosome 6, and the TRAF1-C5 locus on chromosome 9 exceed this threshold. Reproduced with permission from Plenge et al.⁴⁷

Overview

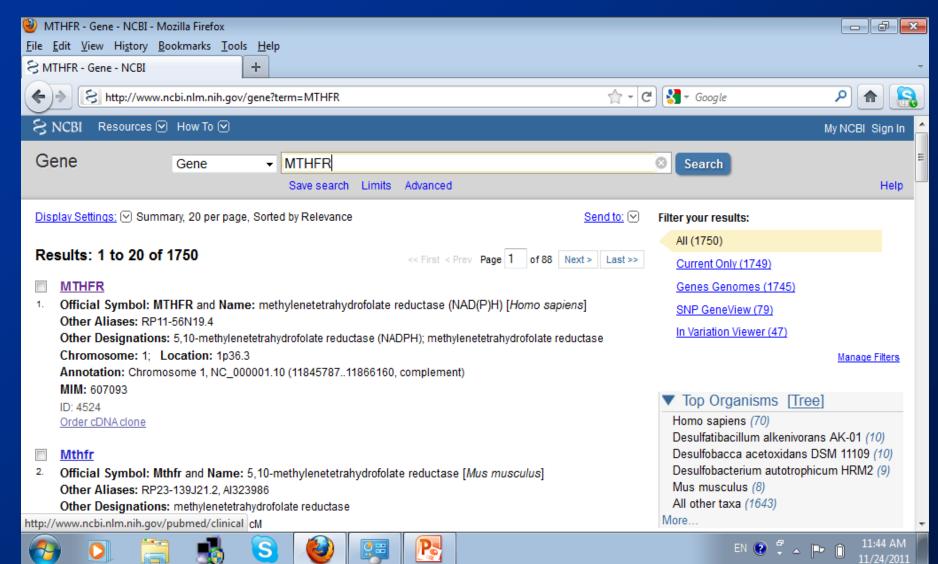
- Gene
- Polymorphism
- Mutation
- Genetic association studies
- Haplotype
- Linkage disequilibrium
- Genome-wide association studies



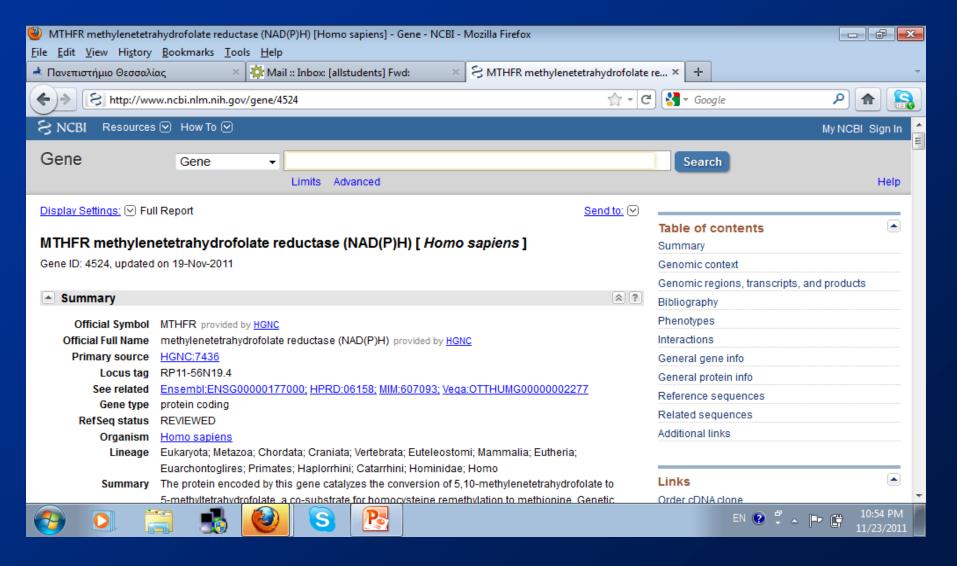
Example: MTHFR gene. Go to PubMed (<u>www.pubmed.com</u>) and type MTHFR and search GENE

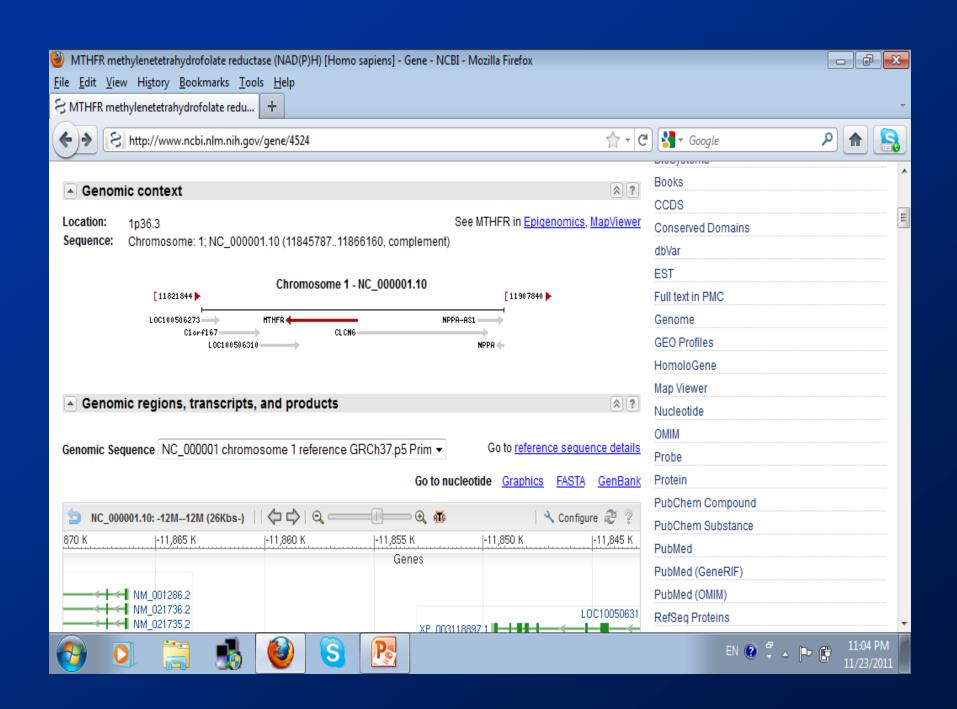


Click MTHFR for Homo Sapiens

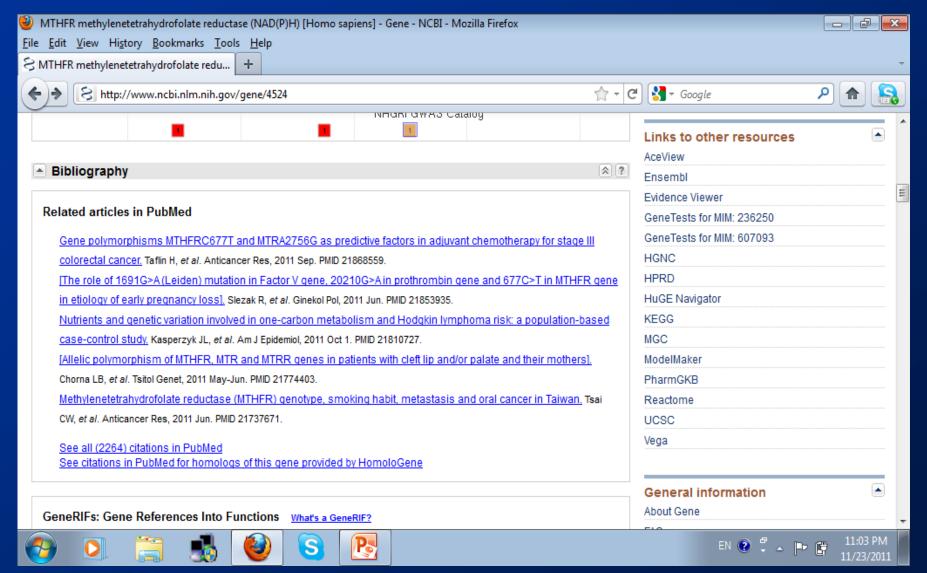


Available information regarding :Summary, Genomic regions, transcripts, Genomic context, Bibliography, Interactions, General gene information General protein information, Reference Sequences, Related Sequences

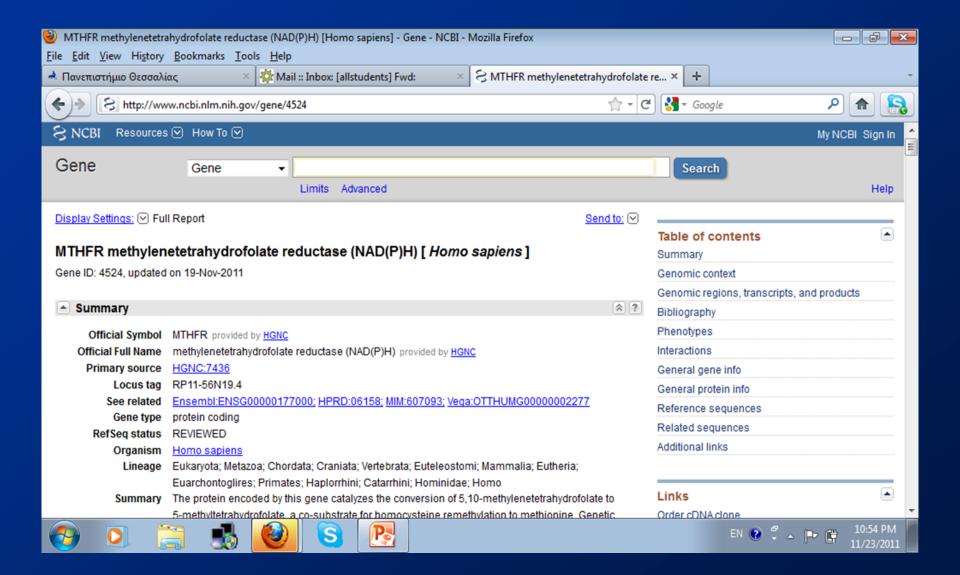




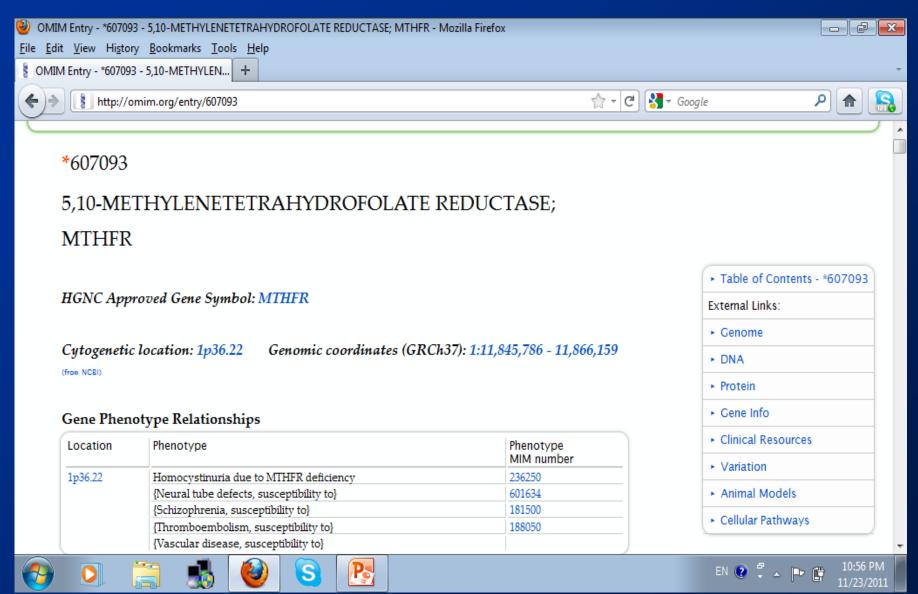
Have a look to the references from PubMed...and how the gene and its SNPs are named in the literature...



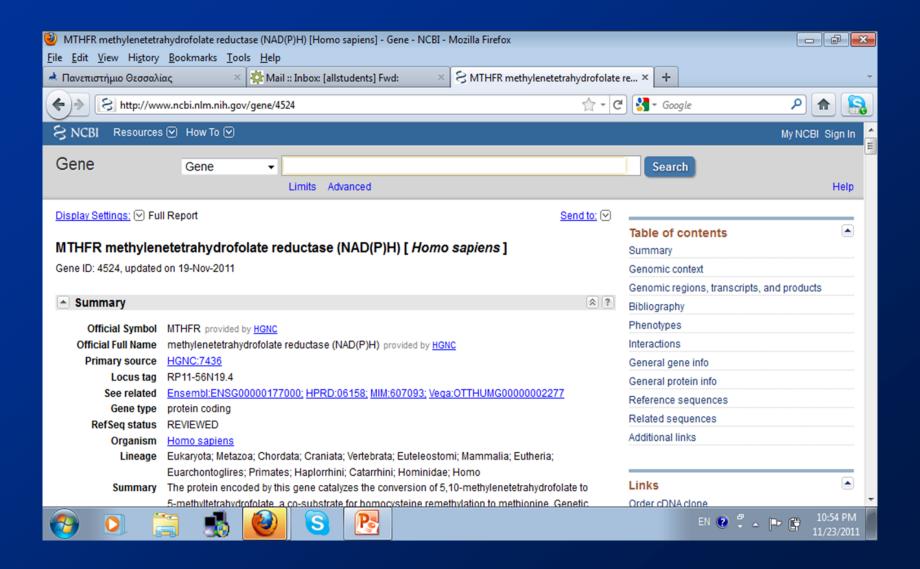
click on the link MIM 607093, hyperlink to OMIM database



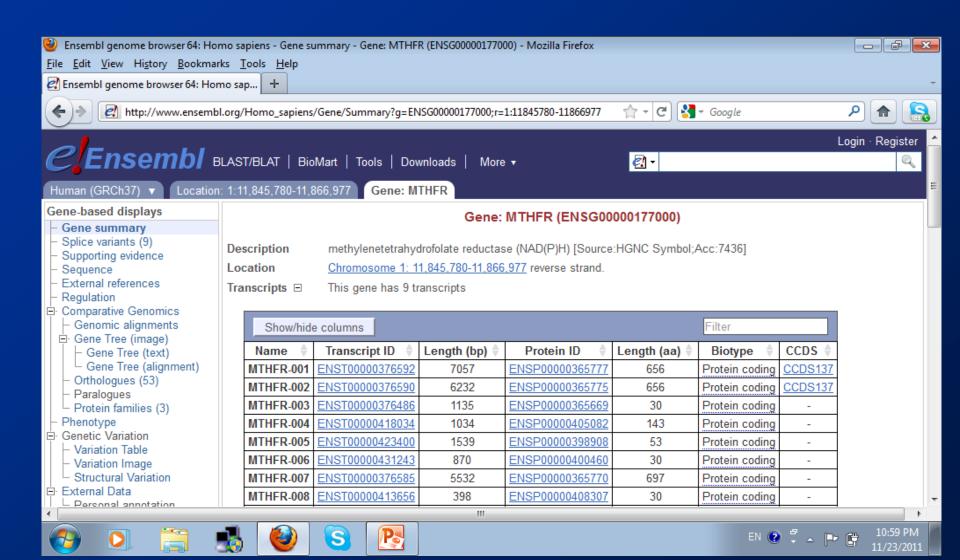
High quality description of the gene in text format

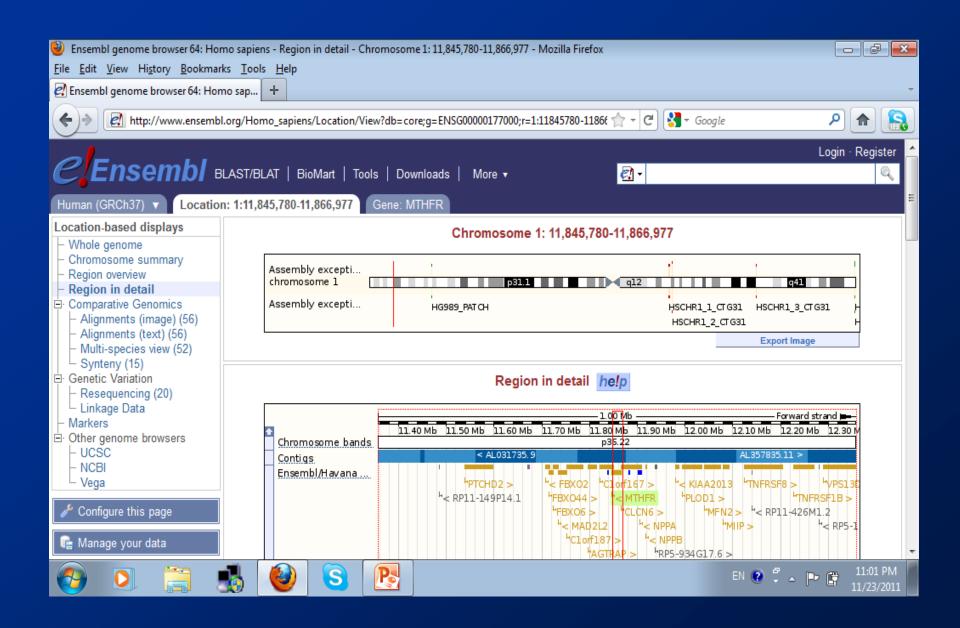


Back to GENE, click on ENSEMBL link

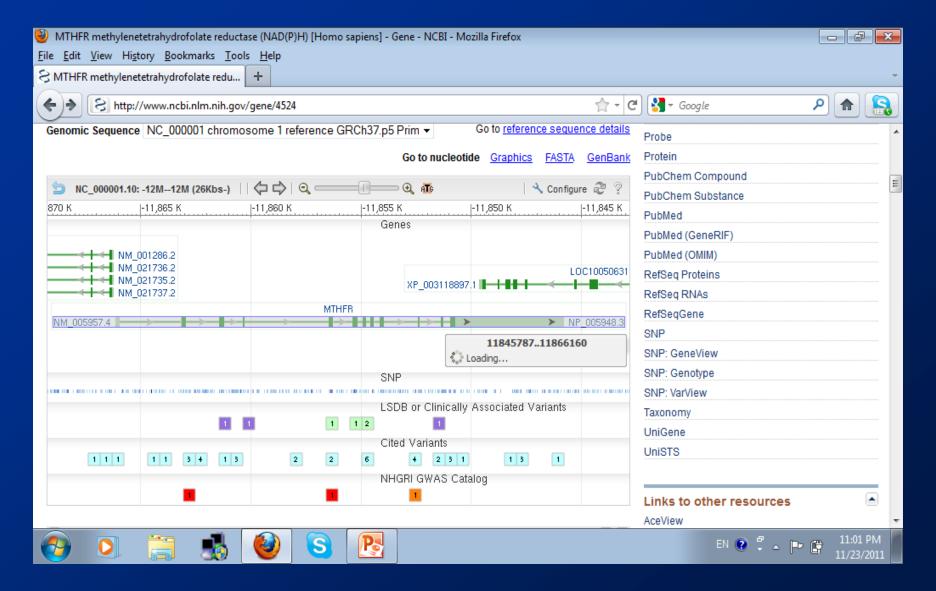


Detailed description of the gene with many bioinformatic tools. Click on the location link to view its location...

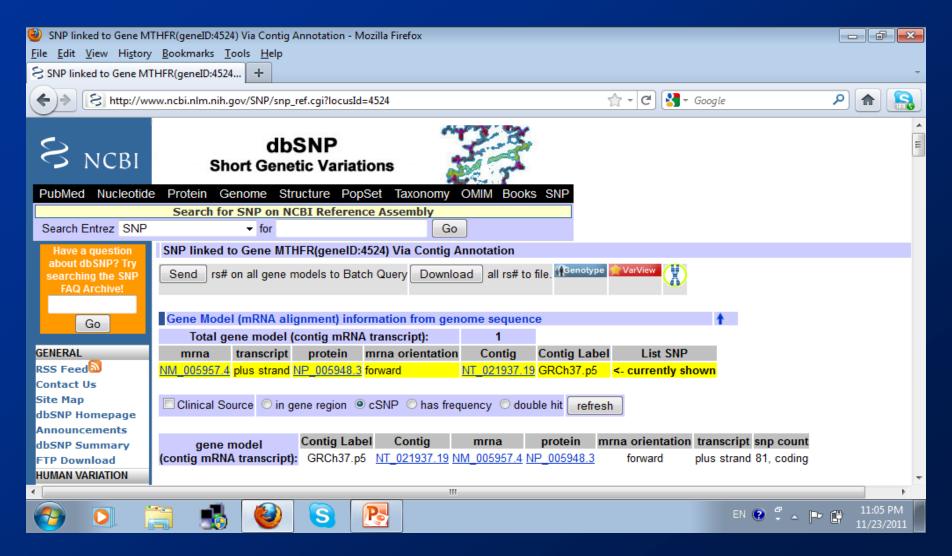




Back to GENE, click on SNP geneview, to search for MTHFR SNPs



dbSNP for MTHFR gene: all coding SNPs are shown (default)

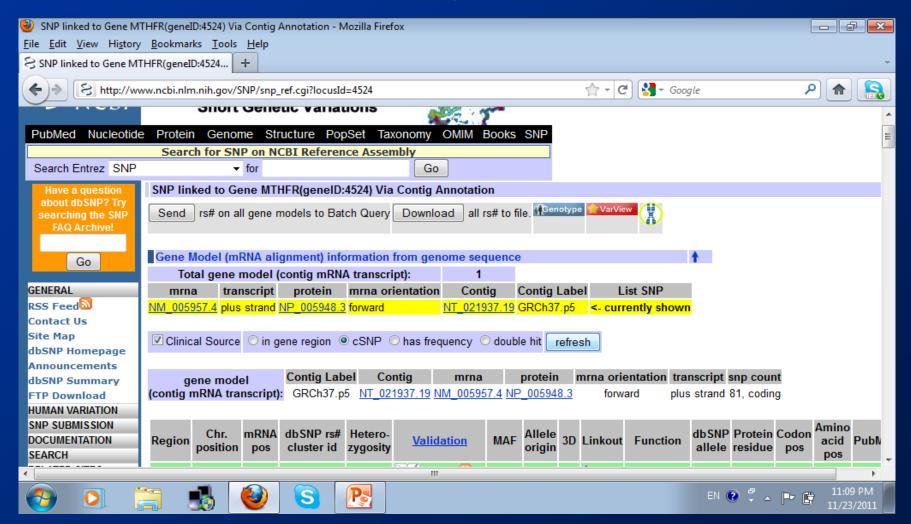


Scroll down...

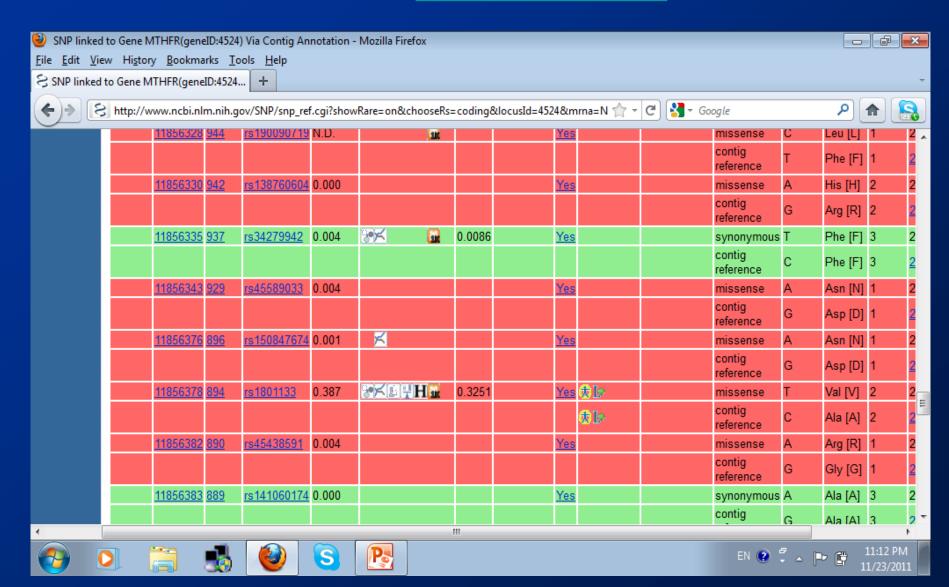


EN 😰 🗸 📭 😭

From bibliography, we are interested for a SNP with a nucleotide change C677T resulting in aminoacid change Ala222Val. This SNP has been implicated with many disorders and we are interested in identifying its unique rs code. Tick include clinically sources and then click refresh



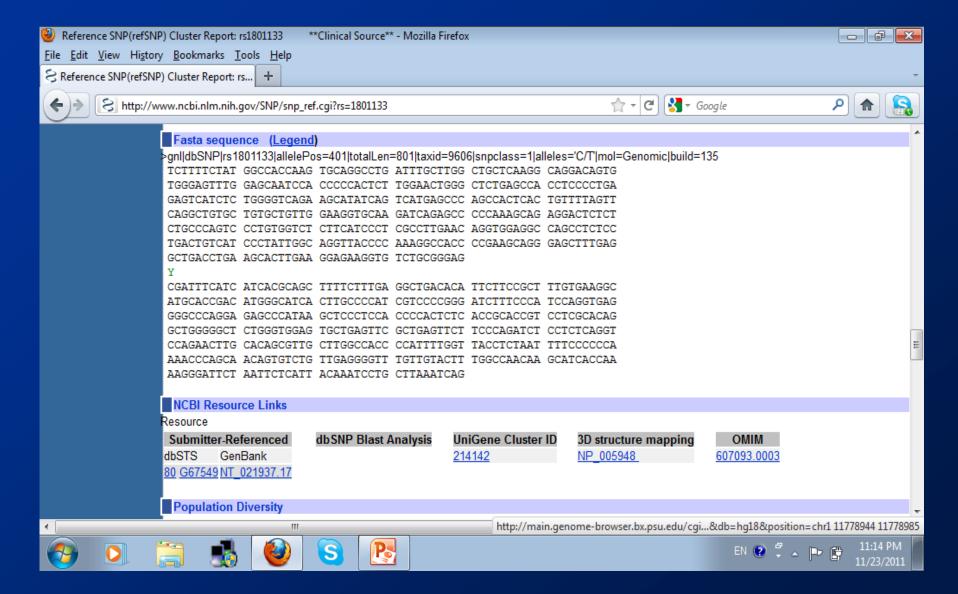
In exon 5, it is <u>rs1801133</u>. Click...



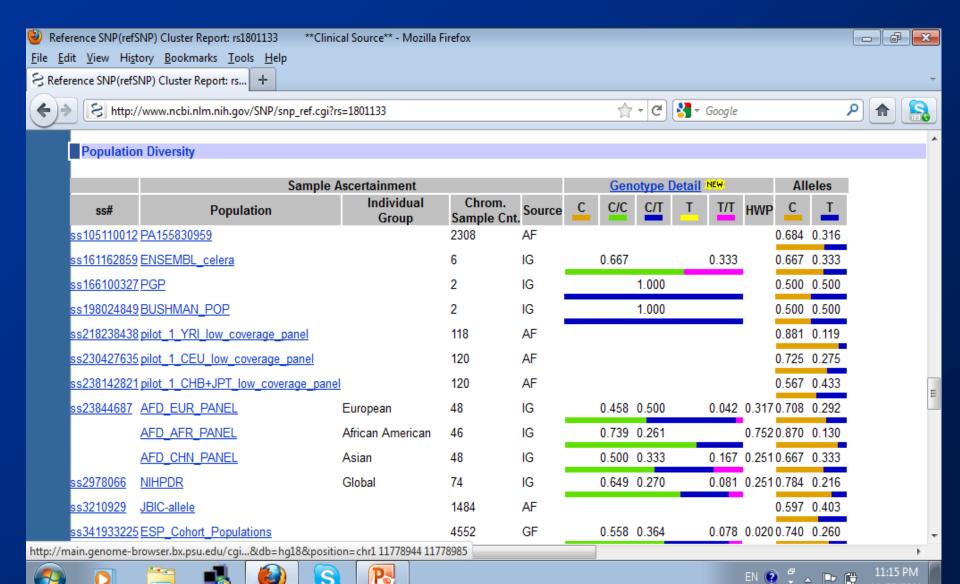
dbSNP for rs1801133



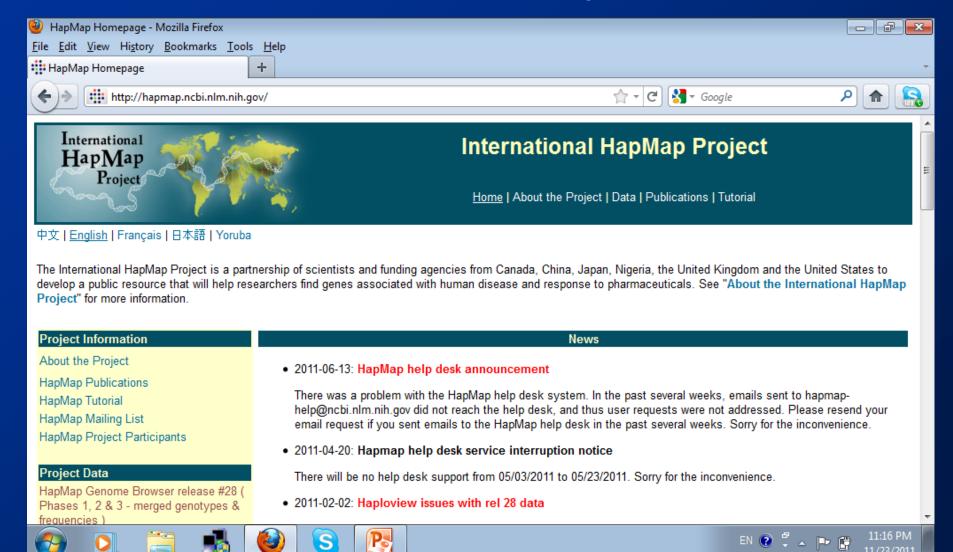
dbSNP for rs1801133



dbSNP for rs1801133

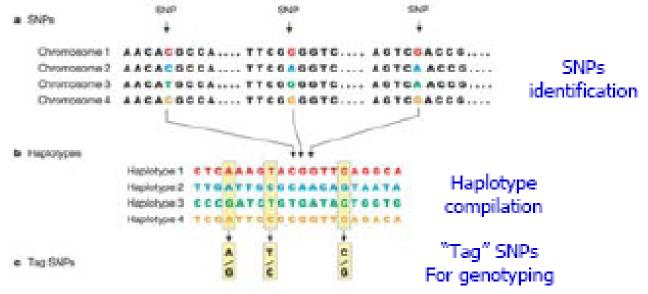


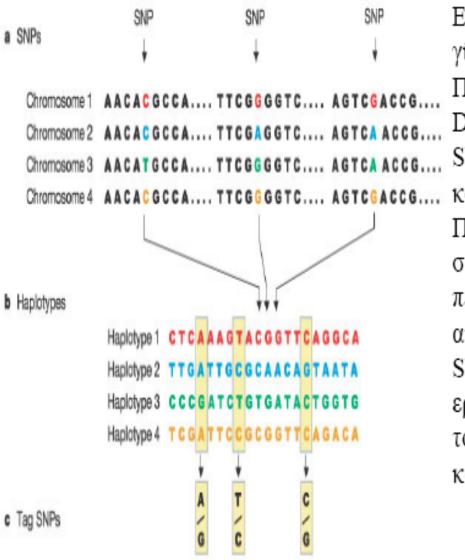
We are now interested to see the haplotype structure of the gene. Is our SNP in LD with other SNPs? Type www.hapmap.org



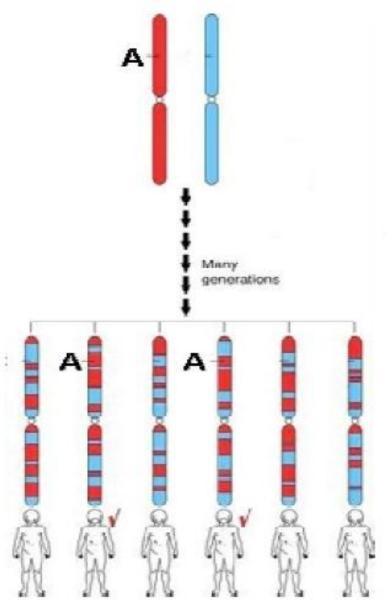
International HapMap Project

- A catalog of common genetic variants that occur in human beings.
 - what these variants are
 - where they occur in our DNA
 - how they are distributed among people within populations and among populations in different parts of the world



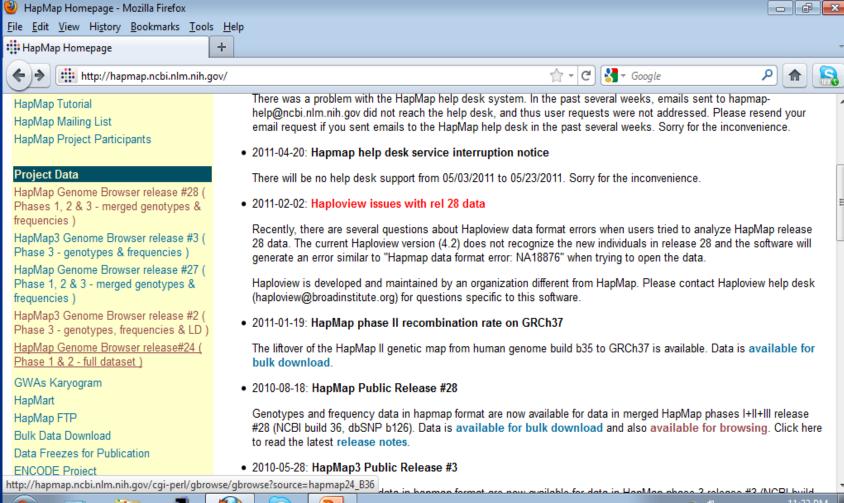


Εικόνα 2: Η κατασκευή του ΗαρΜαρ γίνεται σε 3 στάδια. (a) Προσδιορίζονται SNPs σε δείγματα DNA από πολλά άτομα. (b) Γειτονικοί SNPs που κληρονομούνται μαζί κατατάσσονται σε απλοτύπους (c) Προσδιορίζονται"Tag" SNPs μέσα στους απλοτύπους τα οποία περιγράφουν πλήρως τους απλοτύπους. Με τη γονοτύπωση 3 tag SNPs που φαίνονται στην εικόνα, οι ερευνητές προσδιορίζουν ποιός από τους 4 απλοτύπους είναι παρόν σε κάθε άτομο..



Το διάγραμμα δείχνει 2 προγονικά χρωμοσώματα που αναμιγνύονται μέσω ανασυνδυασμού για πολλές γενιές ώστε να δώσουν διαφορετικά χρωμοσώματα απογόνων. Εάν μια γενετική παραλλαγή που σημειώνεται με το γράμμα Α στο προγονικό χρωμόσωμα αυξάνει τον κίνδυνο για κάποιο νόσημα, οι 2 απόγονοι της σύγχρονης γενιάς που κληρονομούν το συγκεκριμένο τμήμα του χρωμοσώματος, θα έχουν αυξημένο κίνδυνο. Πλησίον της παραλλαγής Α, υπάρχουν πολλοι SNPs που μπορούν να χρησιμοποιηθούν για τον εντοπισμό της Α.

Click HapMap Genome Browser (#24 - full data set)















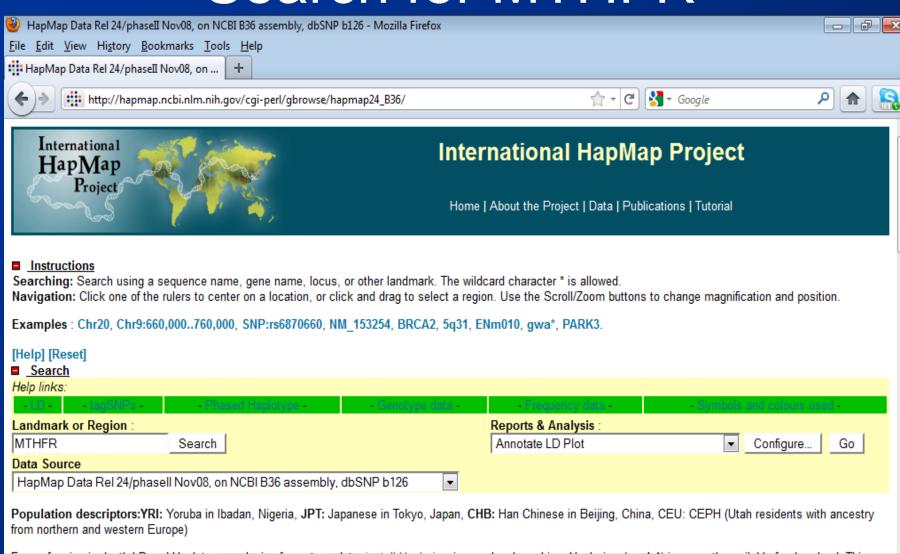








Search for MTHFR















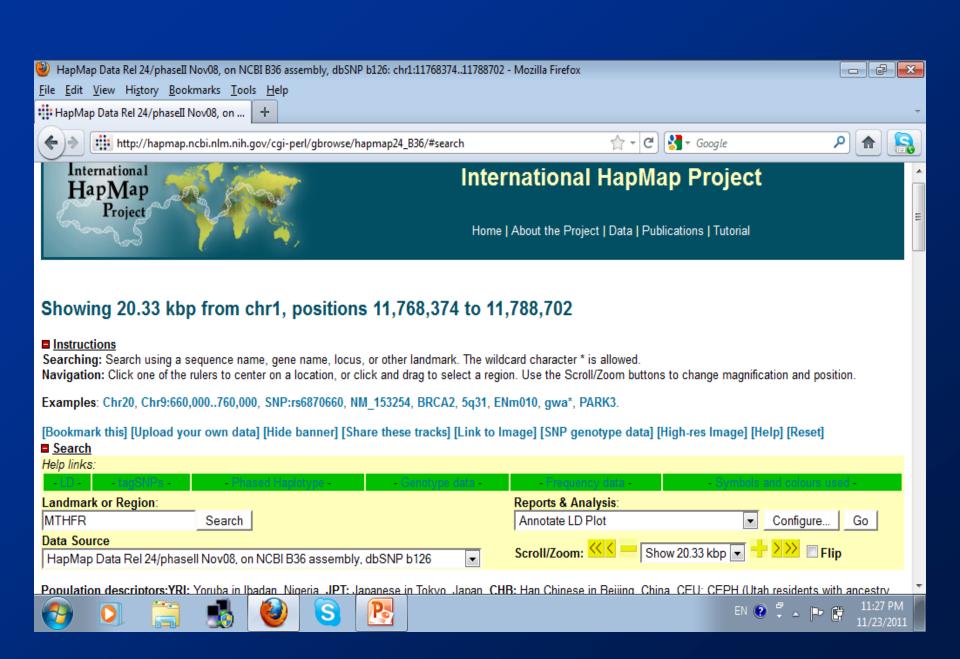


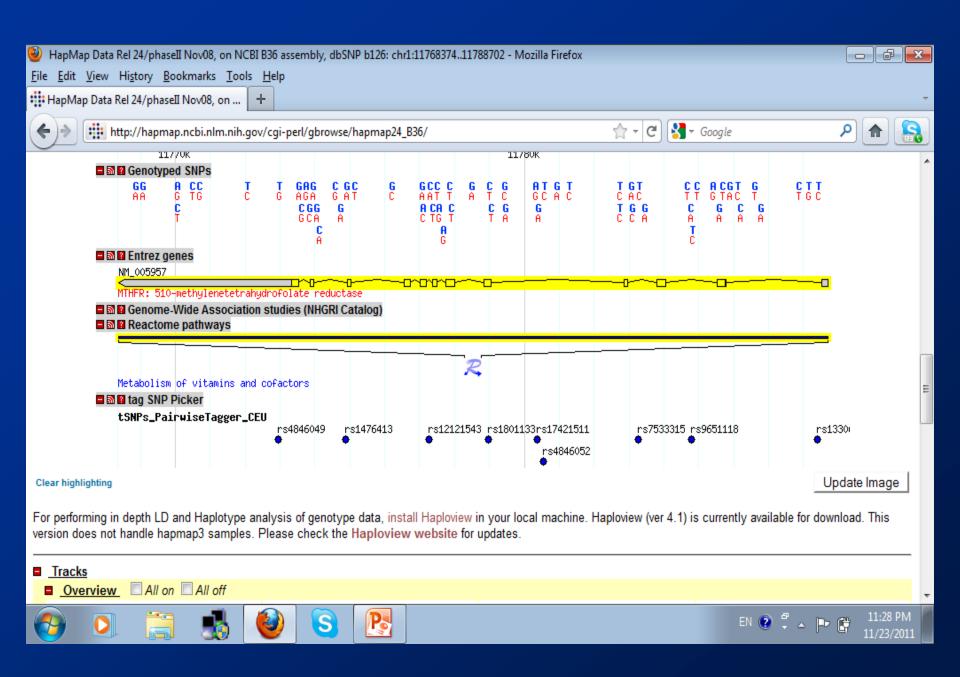




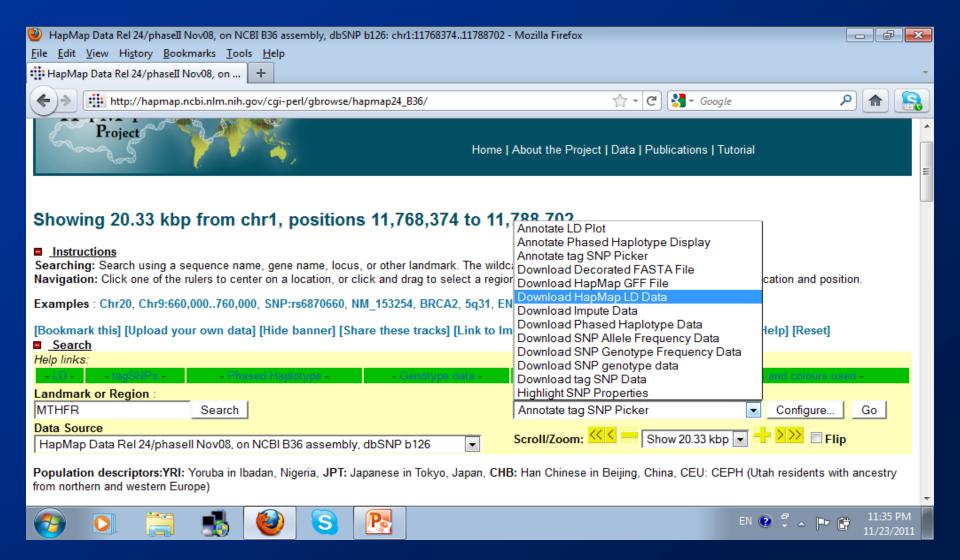




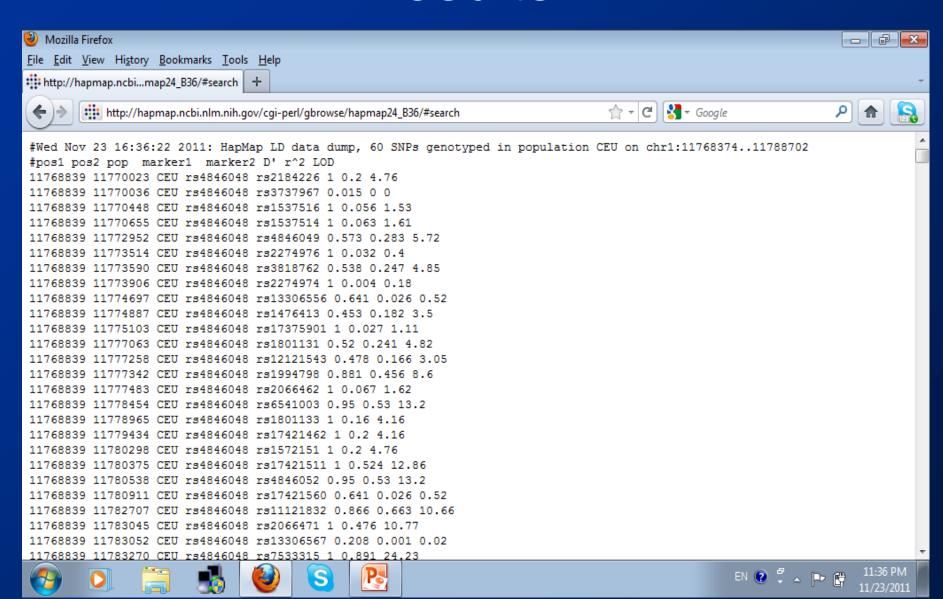




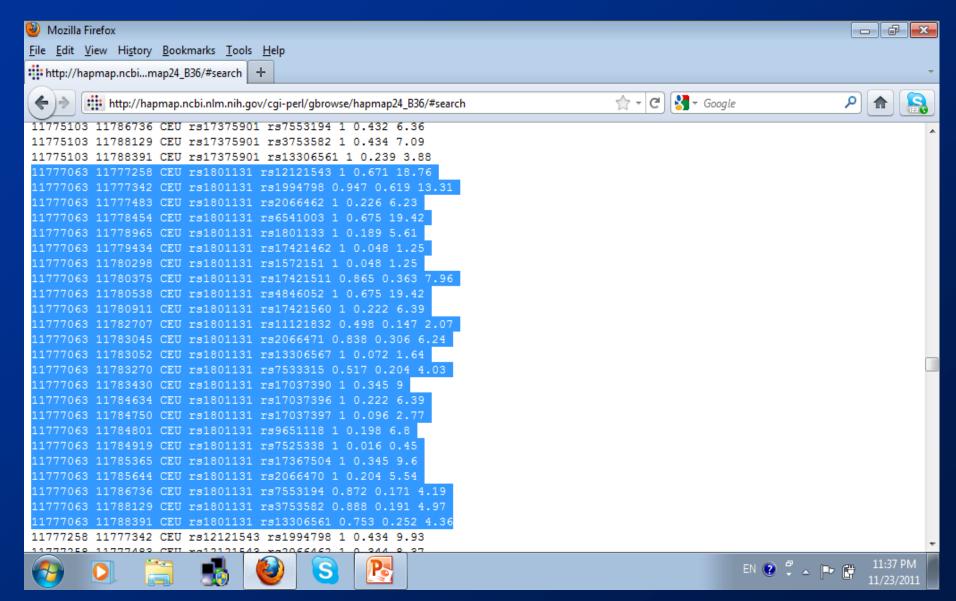
Click on Download HapMap data to find LD scores for all SNPs, including rs1801133



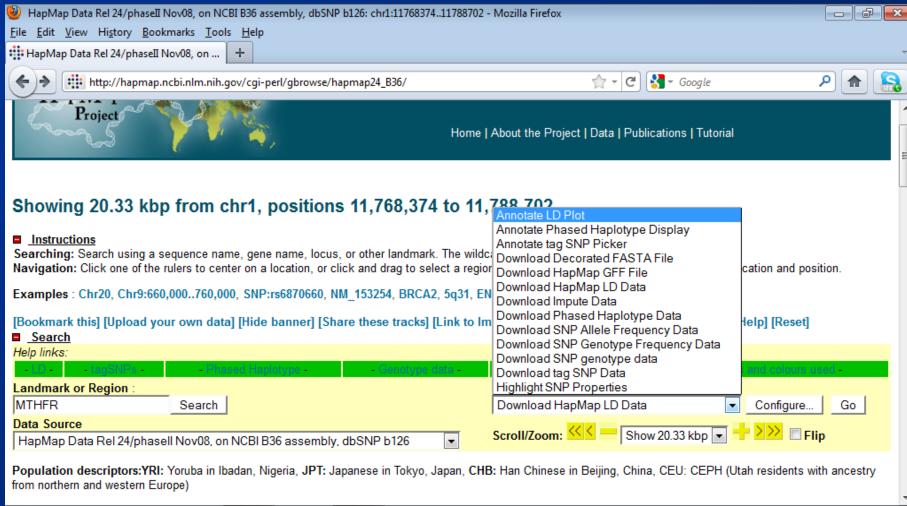
results



LD results for rs1801133



Back to MTHFR page, click Annotate LD plot















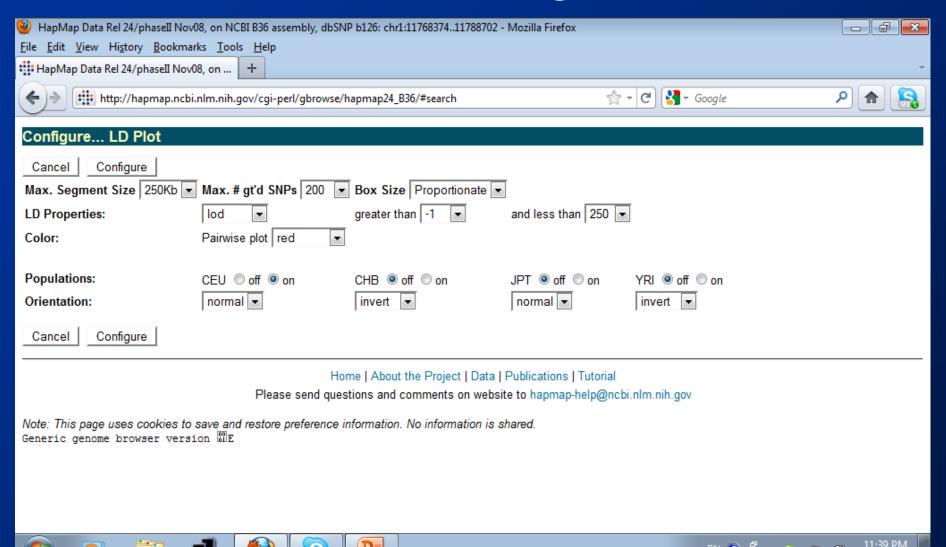




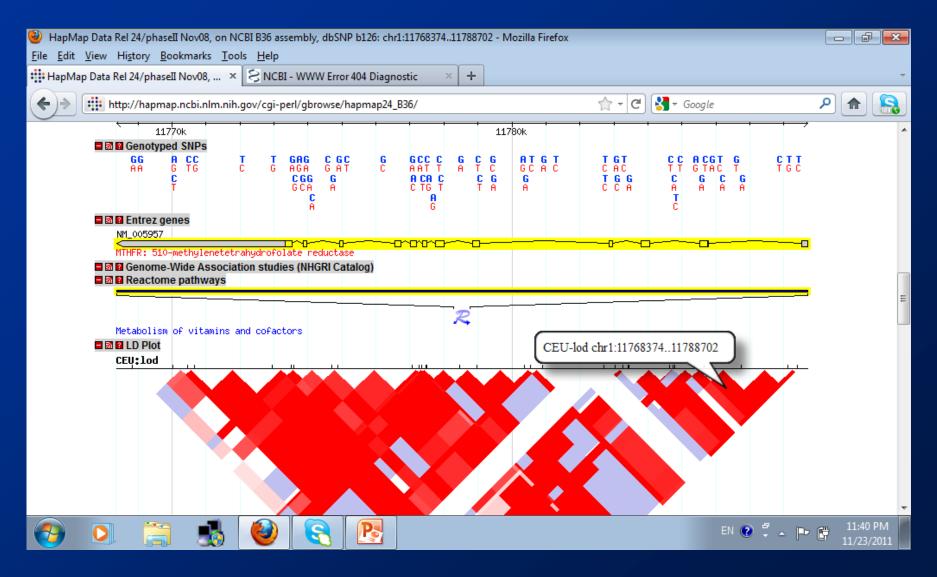


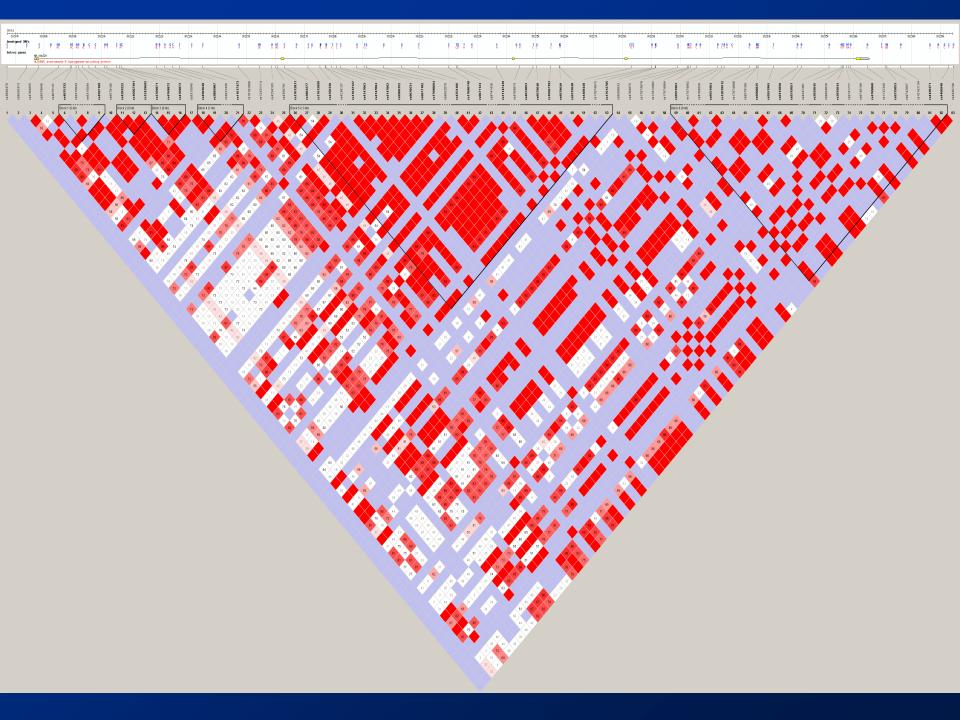


Click configure

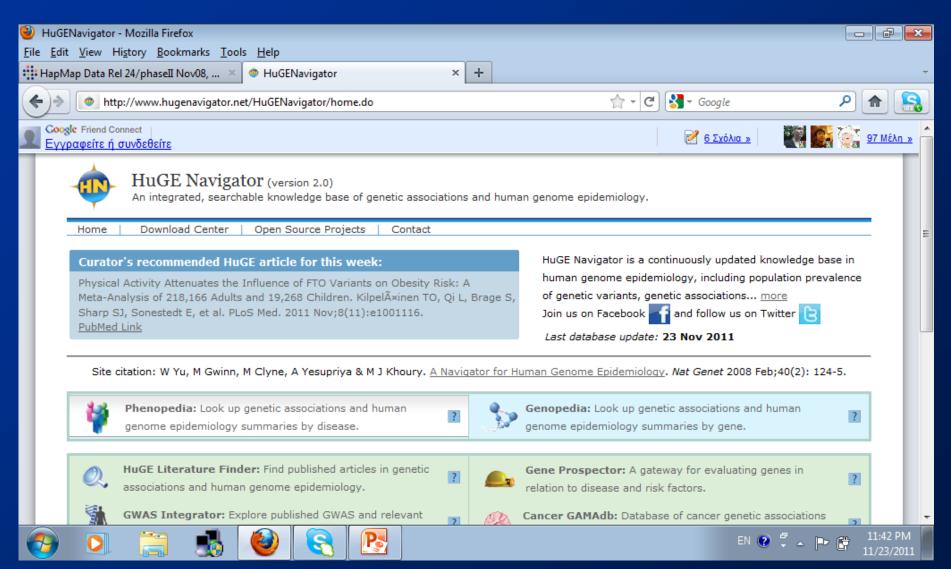


Red color shows regions of High LD

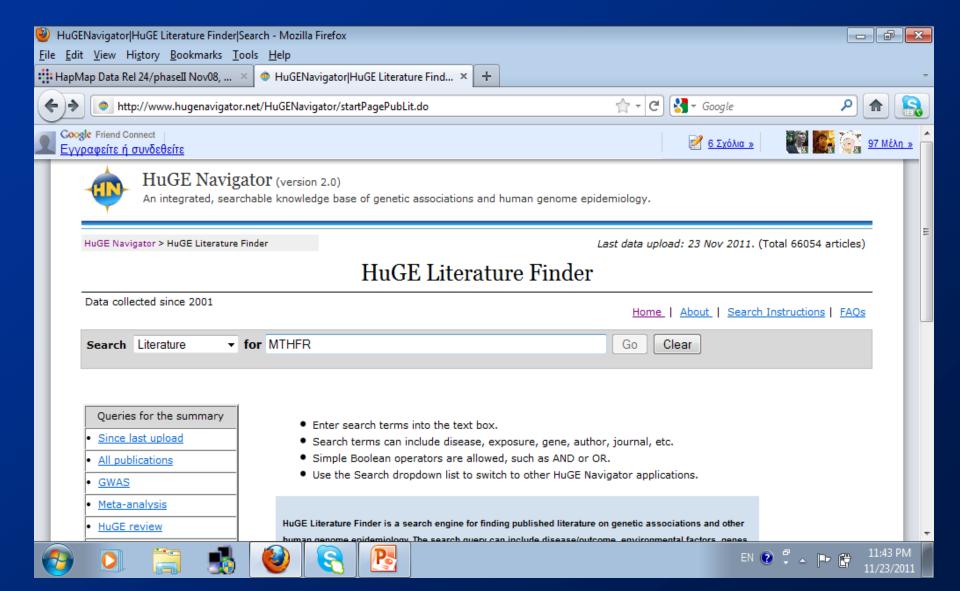




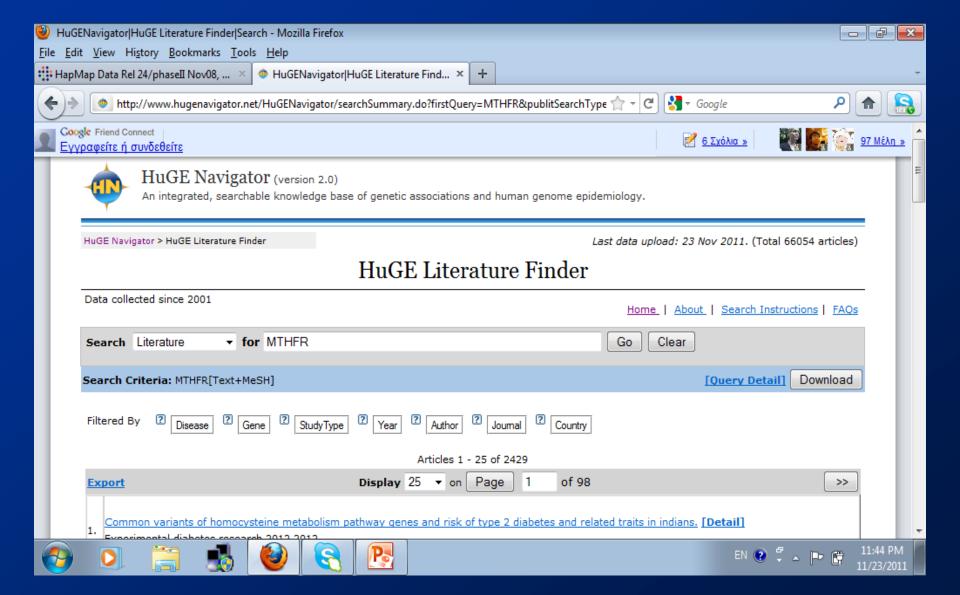
Search for the published literature of MTHFR gene. Go to HuGE Navigator (http://www.hugenavigator.net/), and click on HuGE literature finder



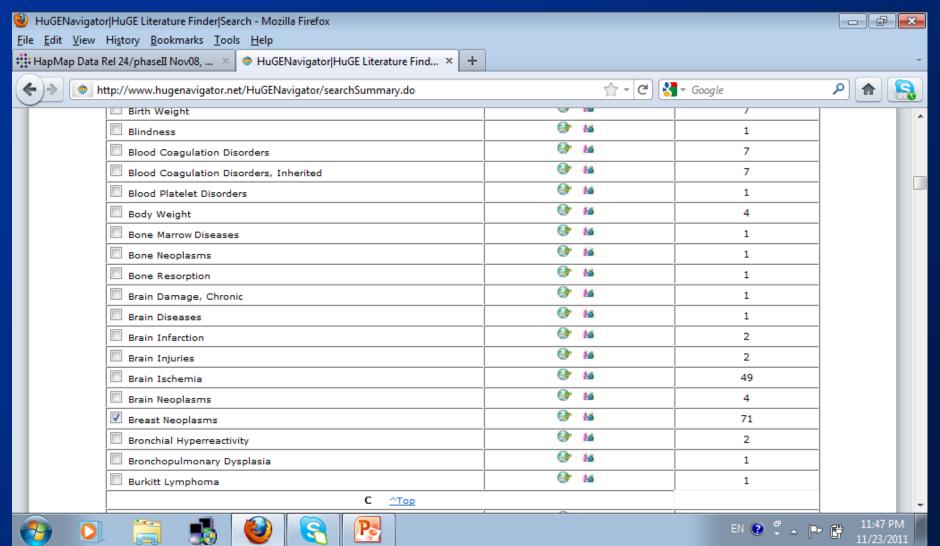
Search literature for MTHFR



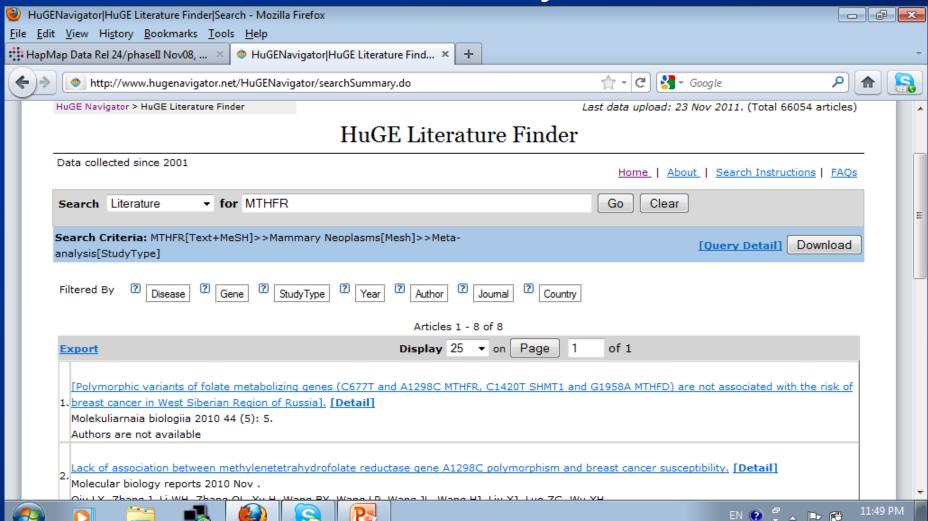
2429 results found!!!



Limit by disease and choose Breast Neoplasms...and click continue



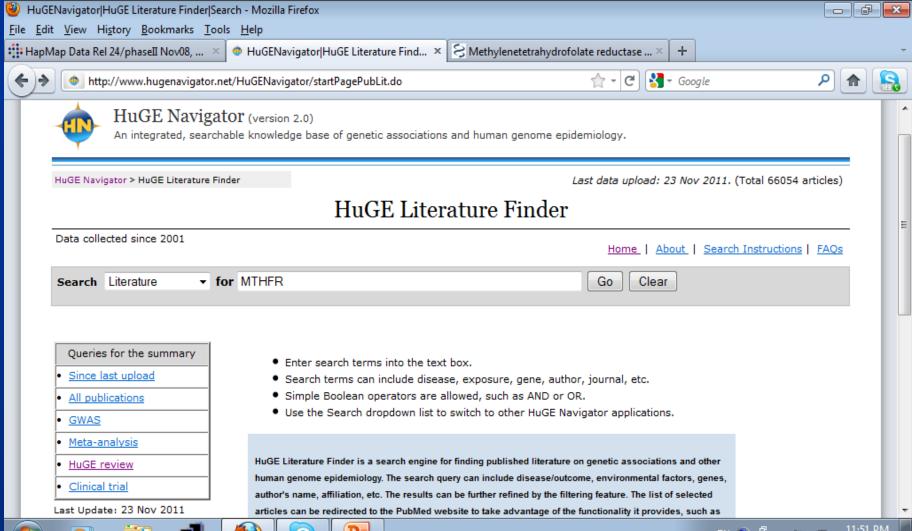
Limit by study type and choose meta-analysis



Click on the title to view the PubMed abstract



Back to HuGE Literature Finder, click on HuGE Reviews, to find state-of-the-art open-access reviews for gene-disease associations













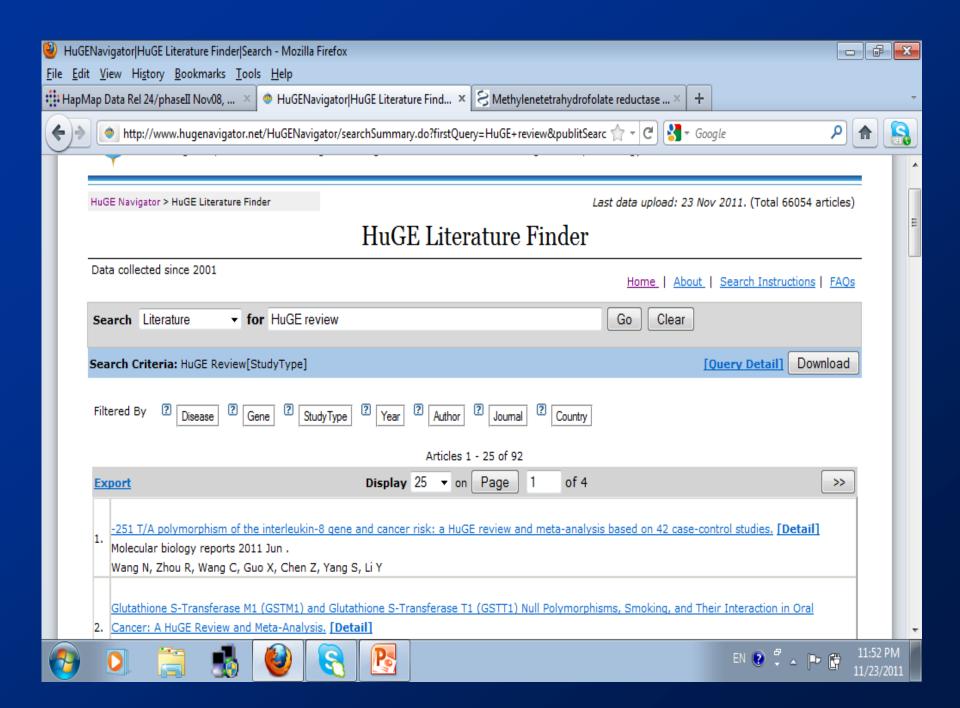










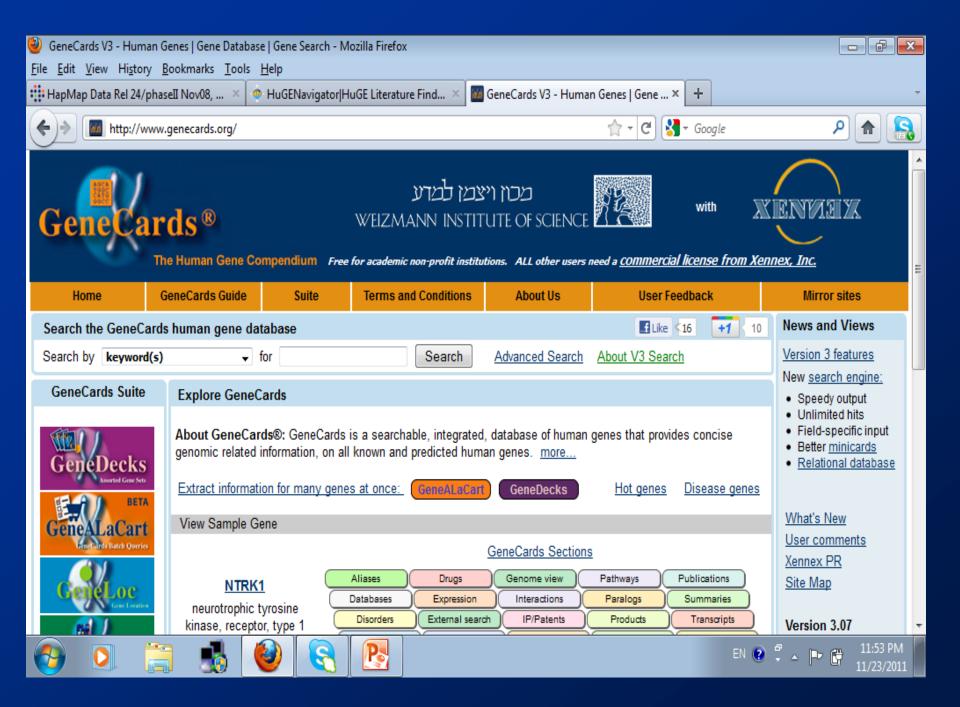




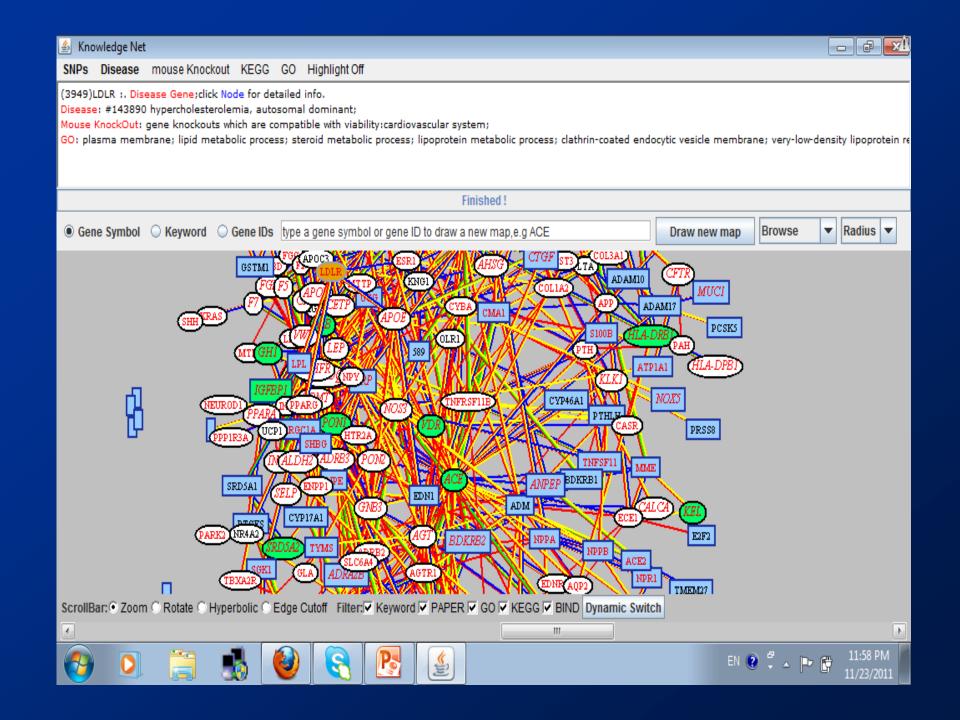
Useful links

http://www.genecards.org/

GeneCards is an integrated database of human genes that includes automatically-mined genomic, proteomic and transcriptomic information, as well as orthologies, disease relationships, SNPs, gene expression, gene function, and service links for ordering assays and antibodies.

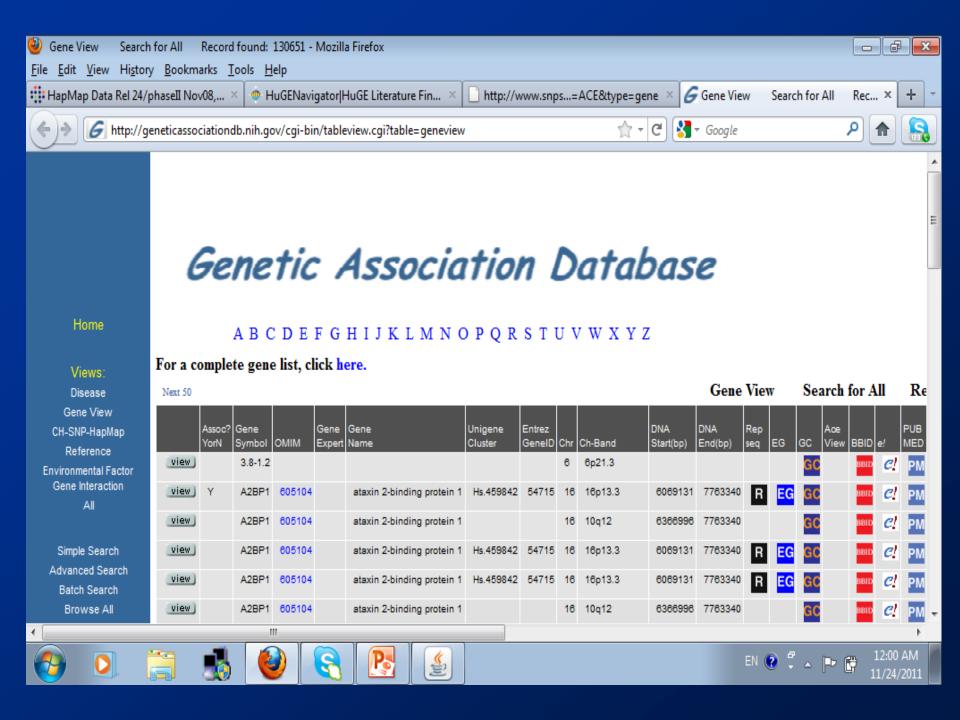


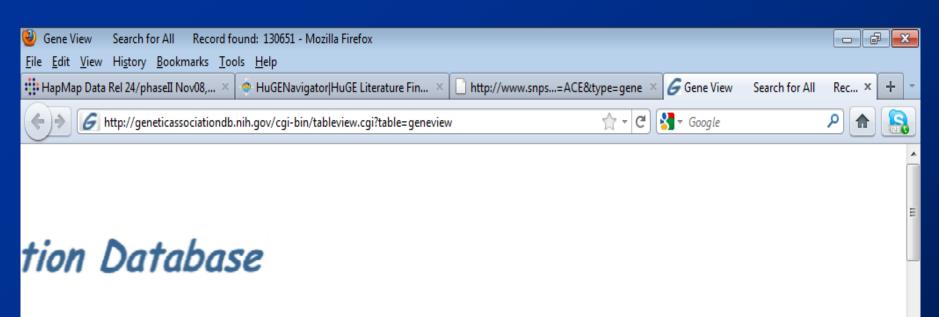
- http://www.snps3d.org/
- SNPs3D is a website which assigns molecular functional effects of nonsynonymous SNPs based on structure and sequence analysis.
- Gene-gene interactions networks



Genetic Association Database

- archive of human genetic association studies of complex diseases and disorders. The goal of this database is to allow the user to rapidly identify medically relevant polymorphism from the large volume of polymorphism and mutational data, in the context of standardized nomenclature.
- The data is from published scientific papers.
- additional molecular reference numbers and links.
- gene centered.
- http://geneticassociationdb.nih.gov/





PQRSTUVWXYZ

Gene View	Search for All	Record found:	130651

	Entrez GenelD	Chr				Rep seq	EG		Aœ View	BBID (PUB MED		Broad Phenotype (Disease)	Reference	ı
		6	6p21.3					GC		BBID	e!	PM	UNKNOWN	Atherosclerosis Hyperlipidemia	Felekis T et al. 2010	
Hs.459842	54715	16	16p13.3	6069131	7763340	R	EG	GC		BBID	e!	PM	METABOLIC	Osteoarthritis	G Zhai , et al. Journal of medi	C
		16	10q12	6366996	7763340			GC		BBID	e!	PM	METABOLIC	obesity	Ma L et al. 2010	
Hs.459842	54715	16	16p13.3	6069131	7763340	R	EG	GC		BBID	e!	PM	PSYCH	ADHD attention-deficit hyper	J Elia , et al. Molecular psychi	а
Hs.459842	54715	16	16p13.3	6069131	7763340	R	EG	GC		BBID	e!	PM	CHEMDEPENDENCY	Tobacco Use Disorder	George R Uhl , et al. Archives	0
		16	10q12	6366996	7763340			GC		BBID	<i>e</i> !	PM	METABOLIC	BILIARY CIRRHOSIS Liver Cirrho	Joshita S et al. 2010	Ŧ
←						III					I					

















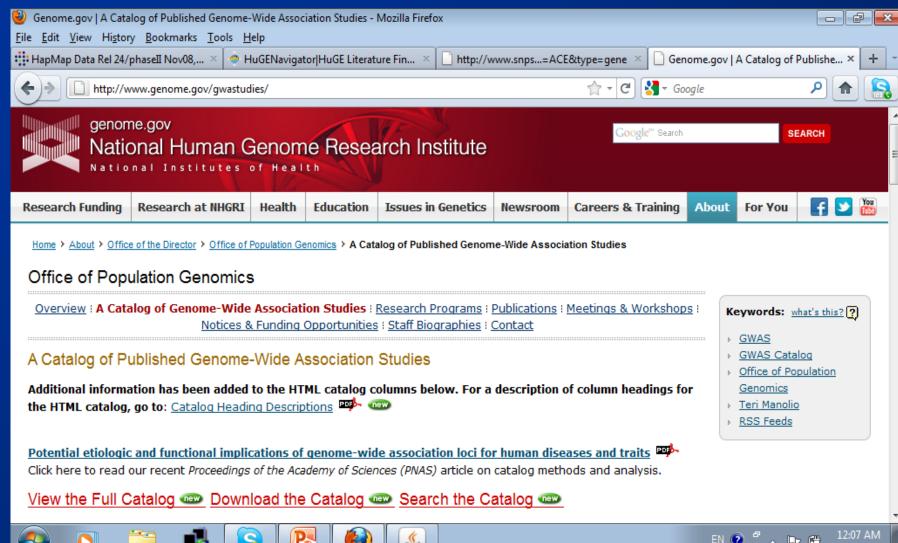






12:01 AM 11/24/2011

www.genome.gov





First Author/Date/ Disease/Trait Replication Reported Initial Date Region Mapped Strongest Con Journal/Study Sample Size Gene(s) Added to Sample Size SNP-Risk Gene(s) Catalog Allele (since 11/25/08) 1,333 European | NR 11/18/11 Alcohol Frank 1 4q23 ADH1B, ADH1C ADH1B rs1789891-? intergenic October 18, 2011 dependence ancestry cases, ADH1C 4p16.2 NR rs1000579-? intergenic Addict Biol 2,168 European STX18 - MSX1 Genome-wide significant ancestry 14q24.2 NR rs2810114-? intron association between controls PCNX alcohol dependence and a NR 2q35 rs1344694-? intergenic variant in the ADH gene MREG - PECR NR UTR-5 13q12.12 rs4770403-? cluster. SGCG GPD1L 3p22.3 rs9825310-? intergenic RPSAP11 -CMTM8 Hollingworth P Alzheimer's 11/19/11 1,299 European NR AC110611.1 4q12 NMU - EXOC1 rs753129-? intergenic October 18, 2011 disease ancestry, 2p16.3 AC079250.1 KCNK12 rs2969775-? interaenic











Ш















GENETIC EPIDEMIOLOGY GLOSSARY M.Tevfik DORAK

Accompanying Genetic Epidemiology Lecture Note & Presentation

ACCE project (analytic validity, clinical validity, clinical utility, ELSI): A CDC sponsored project for evaluating data on emerging genetic tests. It takes its name from the four components of evaluation: analytic validity, clinical validity, clinical utility and associated ethical, legal and social implications (ELSI). For details, see ACCE Project website. See also Grosse & Khoury, 2006 for the clinical utility of genetic testing, Offit, 2008 for issues surrounding genomic disease profiling, and Pharoah, 2008 for the possible utility of genomic profiling in breast cancer risk assessment.

Additive genetic model: In a disease association study, if the risk conferred by an allele is increased r-fold for heterozygotes and 2r-fold for homozygotes, this corresponds to additive model (Lewis, 2002). These data are best analyzed using Armitage trend test for genotype frequencies or by logistic regression in which the genotypes are represented as (-1), 0, (+1). This genotype-based association test does not require the locus to be in Hardy-Weinberg equilibrium. In the case of an association with heterozygosity, the additive model test may be statistically non-significant despite the presence of an association. Thus, a non-significant additive model test does not rule out an association. It has been pointed out that "genes do not generally act in a simple additive manner but through complex networks involving gene-gene and gene-environment interactions" (Colhoun, 2003). The effect that cannot be explained by an additive (or heterogeneity / non-interactive) model in complex disease genetics is due to the dominance (epistatic / interactive) model. See also multiplicative genetic model. See MODEL-online tool for genetic association analysis for different models.

Additive variance: The component of genetic variance due to the additive effects of alleles segregating in the population. In evolutionary genetics, additive genetic variance is a measure for the potential amount of evolutionary change caused by natural selection. See Genetic Calculation Applets:















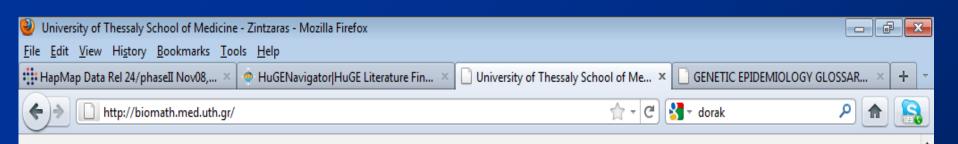












Department of Biomathematics







General

- Home
- Latest News
- Useful Links
- Contact Us

The

"PRIZE OF DISTINGUISHED RESEARCHER IN MEDICINE AT UNIVERSITY OF THESSALY"

was awarded to Elias Zintzaras

Research Interests

User Login - CUMAGAS

- A Create an account
- Forgot Password

- · Mathematical modelling of genetic systems
- · Genetic statistics and epidemiology
- Bioinformatics
- Meta-analysis
- Monte-Carlo modelling
- Bioequivalence testing























