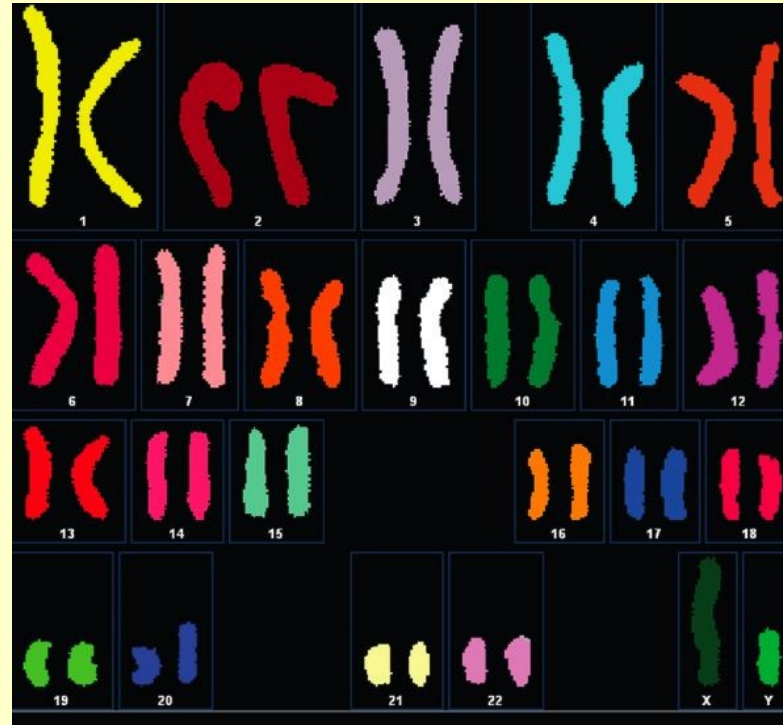


Computational Molecular Biology

Biochem 218 – BioMedical Informatics 231

<http://biochem218.stanford.edu/>

Simple Nucleotide Polymorphisms (SNPs)



Doug Brutlag

Professor Emeritus

Biochemistry & Medicine (by courtesy)



Simple Nucleotide Polymorphisms (SNPs)

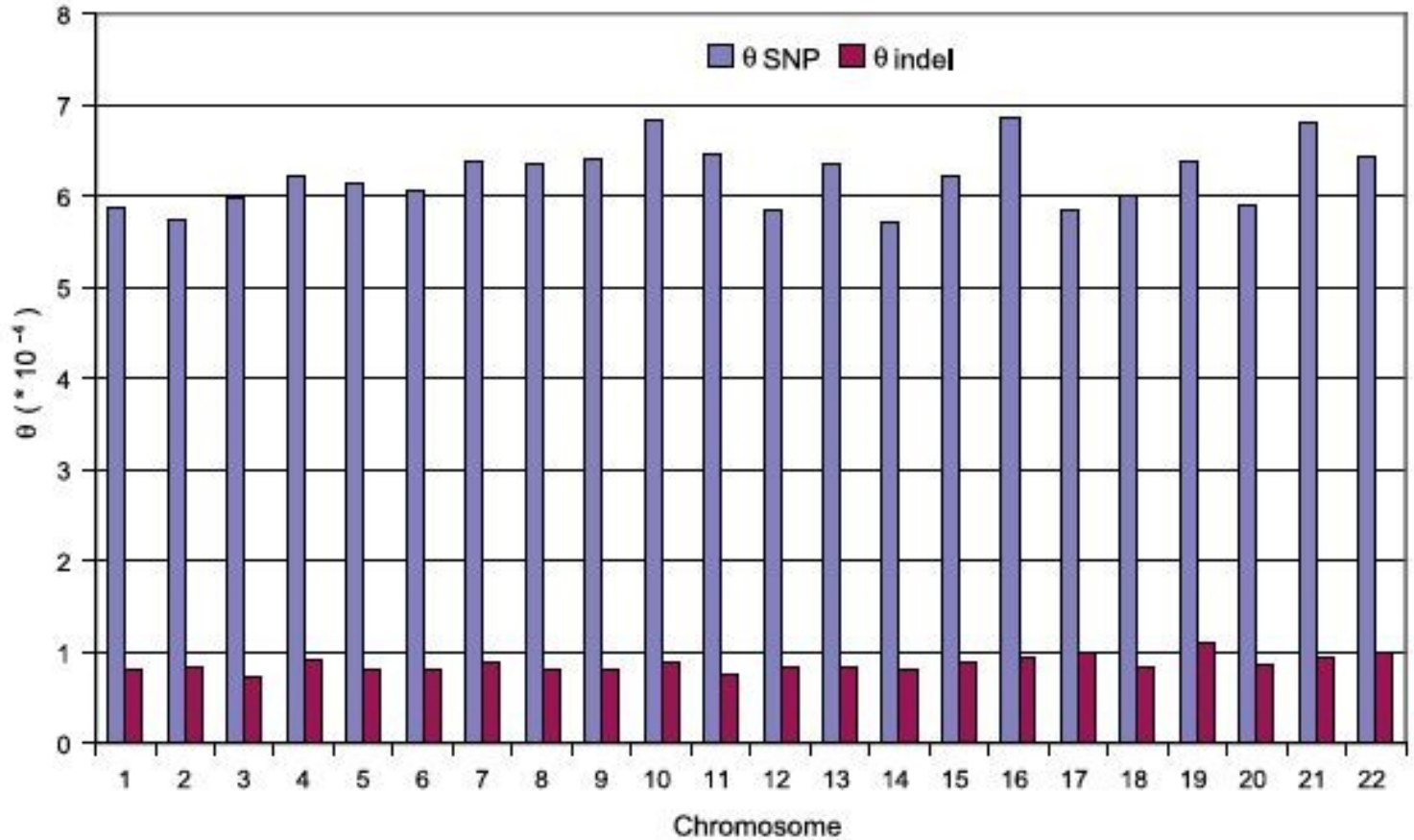
~ 15 million sites in the human genome where SNPs could occur (~ 10 million are common, MAF > 5%)

**~ 3.8 million SNPs catalogued in 270 individuals
(~ 1 SNP / 1000 bases)**

(International HapMap Consortium, October 2007)

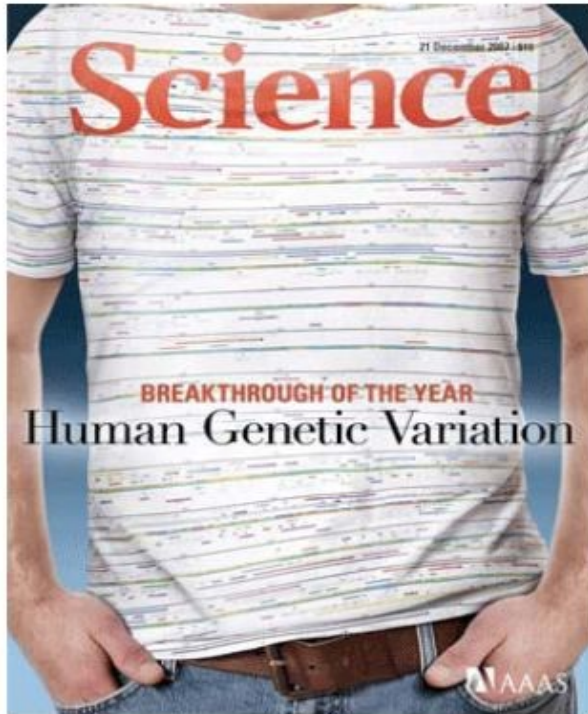
**~ Each person has ~ 3-5 million common SNPs (inherited) and ~30 new mutations have arisen
{mutation rate of 10^{-8} / base / generation}**

SNPs & InDels in HuRef Autosomes



2007 SCIENTIFIC BREAKTHROUGH OF THE YEAR

Science Magazine, December 21, 2007



“It’s all about me!”

Single Nucleotide Polymorphisms (SNPs)

SNP



SNP



Individual 1

A A C A **C** G C C A T T C G **G** G G T C

Individual 2

A A C A **C** G C C A T T C G **A** G G T C

Individual 3

A A C A **T** G C C A T T C G **G** G G T C

Individual 4

A A C A **C** G C C A T T C G **G** G G T C

Single (Simple) Nucleotide Polymorphisms (SNPs)

GCTGTATGACTAGAGAAGATCGAT
GCTGTATGACGAGAAGATCGAT

- SNPs are used for identification and forensics
- SNPs are used for mapping and genome-wide association studies of complex diseases
- SNPs are used for estimating predisposition to disease
- SNPs are used for immigration & citizenship in the UK
- SNPs are used to predict specific genetic traits
- SNPs are used for classifying patients in clinical trials



the Home of DNA testing
the UK's most recommended DNA testing service

- [Home](#)
- [About Cellmark](#)
- [Our customers](#)
- [DNA testing](#)
- [Your questions](#)
- [Forensics & DNA](#)
- [Contact us](#)
- [Register a case online](#)
- [Login](#)
- [Check case progress](#)
- [Privacy](#)
- [Sitemap](#)

DNA testing

Cellmark is proud to be the UK's most recommended DNA testing service. We have been providing a confidential, conclusive & **court approved** DNA testing service for over 20 years.



Fully accredited

We are accredited by the Ministry of Justice as a body that may carry out parentage tests (otherwise called DNA Testing or Paternity Testing) as directed by the civil courts in England and Wales under section 20 of the Family Law Reform Act 1969.

To view our most recent information download our "[Where's the place for DNA testing?](#)" mailer today.



Our Customers

We provide our customers with confidential advice, we can take **DNA testing** registrations over the phone, or **online**, and we will manage the whole sampling and testing process.



People just like you

Information on DNA testing
Why choose Cellmark
Register a case online

Solicitors

Register a case online
Instructing a DNA company
Preventing DNA fraud

Doctors & Nurses












Information for GP's
The ethics of DNA testing
Become a registered DNA sampler

Social workers

Information for Social Workers
Register a case online

Video Answers

Select any of the links below to see a short video answer.

-  [What is DNA testing?](#)
-  [What is the process?](#)
-  [Consent for DNA testing](#)
-  [How to choose a DNA tester](#)
-  [How do you take a DNA sample?](#)
-  [Who needs to give consent?](#)
-  [How long will it take?](#)
-  [Types of DNA testing](#)
-  [How accurate is DNA testing?](#)
-  [Will my DNA be safe?](#)
-  [About Cellmark](#)

Send an SMS

Need some advice on DNA testing? Contact us using our confidential SMS service. Just enter your mobile phone number and your question and we will answer your question.

Your mobile number

Your question

© 2009 Orchid Cellmark Ltd.

Registered Office: 16 Blacklands Way,
Abingdon Business Park, Abingdon,
Oxon, UK, OX14 1DY.
Registered in England No. 4045527.

A SNP Primer at NCBI

<http://www.ncbi.nlm.nih.gov/About/primer/snps.html>



National Center for Biotechnology Information

About NCBI	NCBI at a Glance	A Science Primer	Databases and Tools
Human Genome Resources	Model Organisms Guide	Outreach and Education	News

About NCBI
Site Map

Science Primer:

Bioinformatics

Genome Mapping

Molecular Modeling

ESTs

Microarray
Technology

Molecular Genetics

Pharmacogenomics

Phylogenetics

Just the Facts: A Basic Introduction to the Science Underlying NCBI Resources

SNPs: VARIATIONS ON A THEME

Wouldn't it be wonderful if you knew exactly what measures you could take to stave off, or even prevent, the onset of disease? Wouldn't it be a relief to know that you are not allergic to the drugs your doctor just prescribed? Wouldn't it be a comfort to know that the treatment regimen you are undergoing has a good chance of success because it was designed just for you? With the availability of millions of SNPs, biomedical researchers now believe that such exciting medical advances are not that far away.

What Are SNPs and How Are They Found?

Department of Energy (DOE) SNP Page

http://www.ornl.gov/sci/techresources/Human_Genome/faq/snps.shtml

genomics.energy.gov

Human Genome Project Information • Genomics:GTL • Microbial Genome Program • Home

[skip navigation](#)



Home

Site
Index

Human Genome Project Information

News

About HGP

Research

Education

Ethics

Medicine

Media

SNP Fact Sheet

[Subject Index](#)

[Send the url of this page to a friend](#)

Basic Information

- [FAQs](#)
- [Glossary](#)
- [Acronyms](#)
- [Links](#)
- [Genetics 101](#)
- [Publications](#)
- [Meetings Calendar](#)
- [Media Guide](#)

About the Project

- [What is it?](#)
- [Goals](#)
- [Progress](#)
- [History](#)
- [Ethical Issues](#)
- [Benefits](#)
- [Genetics 101](#)

Medicine & the New Genetics

- [Home](#)
- [Gene Testing](#)
- [Gene Therapy](#)
- [Pharmacogenomics](#)
- [Disease Information](#)
- [Genetic Counseling](#)

Ethical, Legal, Social Issues

- [Home](#)
- [Privacy Legislation](#)
- [Gene Testing](#)
- [Patenting](#)
- [Forensics](#)
- [Genetically Modified Food](#)
- [Behavioral Genetics](#)
- [Minorities, Race, Genetics](#)
- [Genetics in Courtroom](#)



Quick links for this page:

- [What are SNPs?](#)
- [How can SNPs be used as risk factors in disease development?](#)
- [Human Genome Project SNP Mapping Goals](#)
- [What is the SNP consortium?](#)
- [Who are members of the SNP consortium?](#)
- [Why should private companies fund a publicly accessible genome map?](#)
- [Whose DNA was analyzed to create the consortium's SNP map?](#)
- [Are SNP data available to the public?](#)
- [Related Links](#) - SNP basics, articles
- [Meeting Proceedings and Reports](#)

What are SNPs?

Single nucleotide polymorphisms or SNPs (pronounced "snips") are DNA sequence variations that occur when a single nucleotide (A,T,C,or G) in the genome sequence is altered. For example a SNP might change the DNA sequence AAGGCTAA to ATGGCTAA. For a variation to be considered a SNP, it must occur in at least 1% of the population. SNPs, which make up about 90% of all human genetic variation, occur every 100 to 300 bases along the 3-billion-base human genome. Two of every three SNPs involve the replacement of cytosine (C) with thymine (T). SNPs can occur in both coding (gene) and noncoding regions of the genome. Many SNPs have no effect on cell function, but scientists believe others could predispose people to disease or influence their response to a drug.

Although more than 99% of human DNA sequences are the same across the population, variations in DNA sequence can have a major impact on how humans respond to disease; environmental insults such as bacteria, viruses, toxins, and chemicals; and drugs and other therapies. This makes SNPs of great value for biomedical research and for developing pharmaceutical products or medical diagnostics. SNPs are also evolutionarily stable --not changing much from generation to generation --making them easier to follow in population studies.

Simple Nucleotide Polymorphisms (SNPs)

- Most SNPs are genetically neutral
 - Forensics
 - Paternity tests
 - Immigration in the United Kingdom
 - Follow ethnic migrations
 - Clinical trials
- However SNPs can reflect distinguishing characteristics
 - Predisposition to disease
 - Insurability and employment => GINA, December 7, 2009
 - Often the basis for discrimination or other stigma
- Some variations cause disease. Unlike SNPs, these variations are usually very rare.
- SNPs can serve as genetic markers for other traits
 - Clinical trials associate SNPs with drug efficacy
 - Clinical trials associate SNPs adverse drug reactions
 - Consumer genomics associate SNPs with many other traits
 - [23andMe](#), [Navigenics](#), [DNADirect](#)

Carrier Frequency of Cystic Fibrosis Alleles

<http://www.ncbi.nlm.nih.gov/bookshelf/br.fcgi?book=gene&part=cf>

Table 5. Carrier Frequency for Mutant CFTR Alleles

Population Group	Approximate <u>Carrier</u> Frequency	Reference
<u>Ashkenazi Jewish</u>	1:29	<u>Kerem et al [1997]</u>
North American Caucasian	1:28	<u>Hamosh et al [1998]</u>
African American	1:61	<u>Hamosh et al [1998]</u>

Carrier Frequency of Gaucher's Alleles

<http://www.ncbi.nlm.nih.gov/bookshelf/br.fcgi?book=gene&part=gaucher>

Prevalence

A study from Australia reported a disease frequency of 1:57,000 [Meikle et al 1999]; a similar study from the Netherlands reported 1.16:100,000 [Poorthuis et al 1999].

A founder effect for specific alleles underlies the observed occurrence of GD in specific populations:

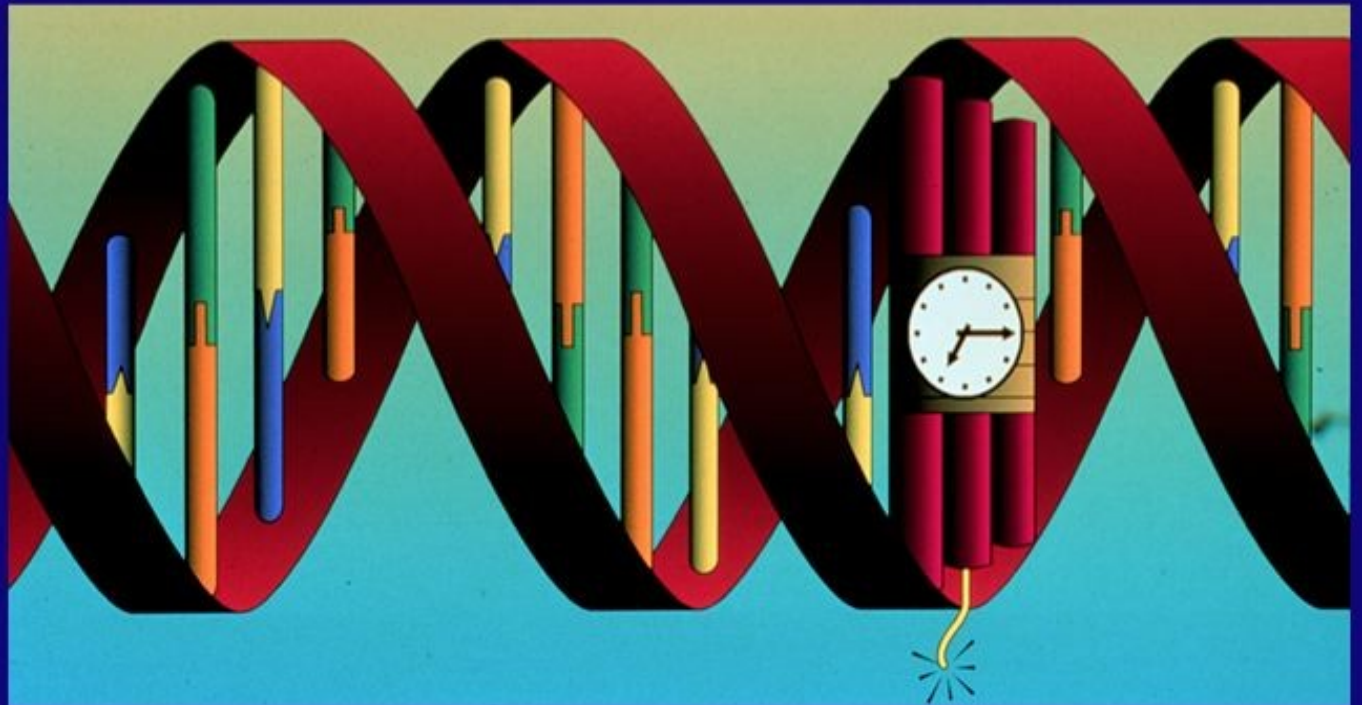
- Ashkenazi Jewish, Spanish, and Portuguese (N370S)
- Swedish (L444P)
- Jenin Arab, Greek, and Albanian (D409H)

Non-neuropathic GD (type 1) is prevalent in the Ashkenazi Jewish population, with a disease prevalence of 1:855 and an estimated carrier frequency of 1:18.

The prevalence of neuropathic GD (types 2 and 3) varies across ethnic groups but appears to be higher among non-Caucasians.

Eric Green's View of SNPs (Director of NHGRI)

All humans are ~99.7% identical at the DNA sequence level, and yet...



all of us carry a significant number of 'glitches' in our genomes.

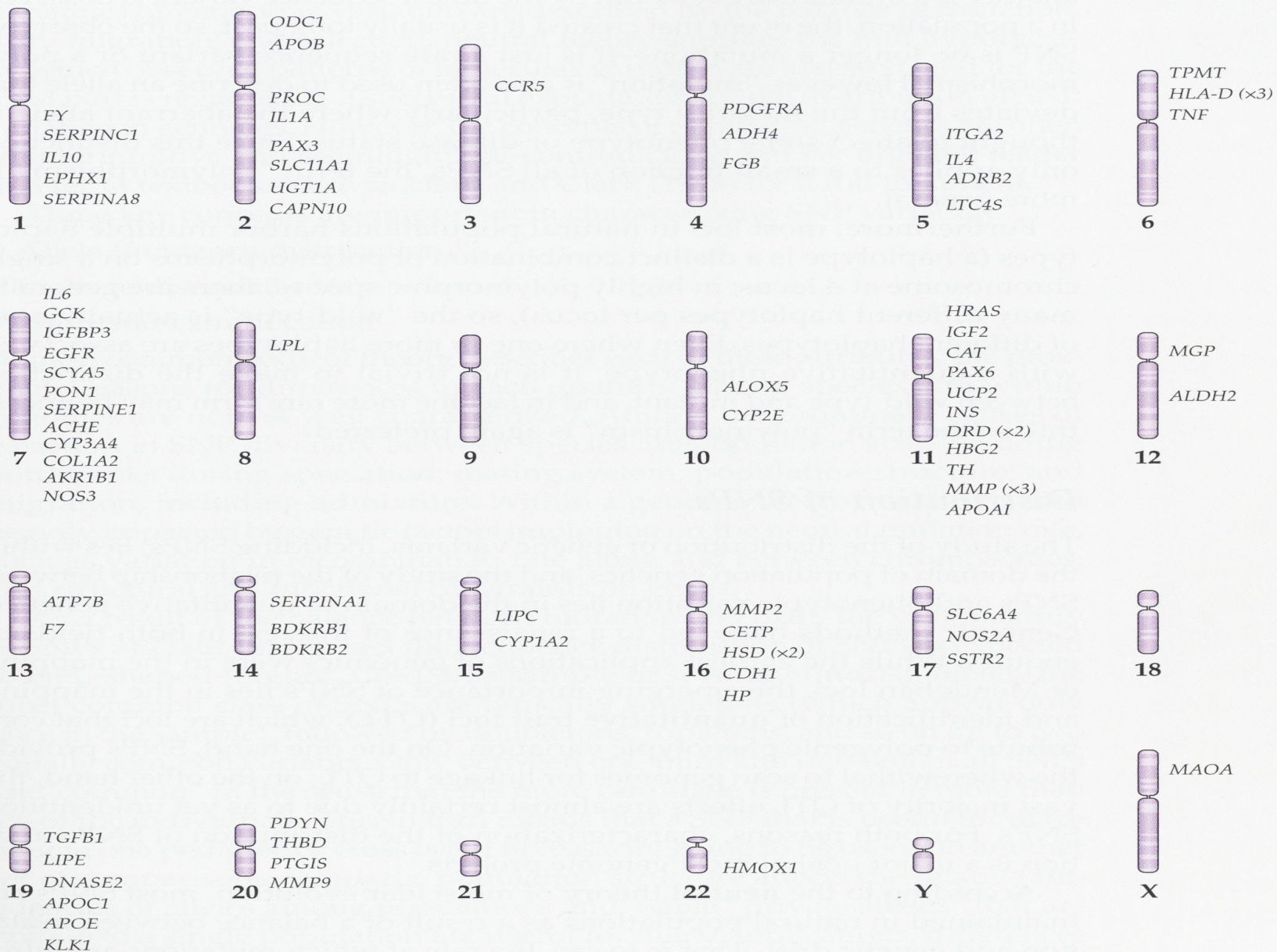


Types of SNPs

<http://www.ncbi.nlm.nih.gov/sites/entrez?db=snp>

- Noncoding SNPs
 - 5' UTR
 - 3' UTR
 - Introns
 - Intergenic Regions
 - Pseudogenes
 - Regulatory
 - Splicing
 - Transcriptional regulation (promoter & TF binding sites)
 - Translational regulation (initiation or termination)
 - Regulatory miRNA target sites
- Coding SNPs
 - Synonymous SNPs (third position variation)
 - Replacement SNPs (change Amino acid)
 - Functional SNPs (acceptable amino acid replacement)
 - Non-functional SNPs (traits & diseases)

Human Promoter SNPs





Search for

Display

1: HBB hemoglobin, beta [*Homo sapiens*]

GeneID: 3043 updated 17-Feb-2010

Summary ↑ ?

Official Symbol	HBB	provided by HGNC
Official Full Name	hemoglobin, beta	provided by HGNC
Primary source	HGNC:4827	
See related	Ensembl:ENSG00000223609 ; HPRD:00786 ; MIM:141900	
Gene type	protein coding	
RefSeq status	REVIEWED	
Organism	Homo sapiens	
Lineage	<i>Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo</i>	
Also known as	CD113t-C; beta-globin; HBB	
Summary	The alpha (HBA) and beta (HBB) loci determine the structure of the 2 types of polypeptide chains in adult hemoglobin, Hb A. The normal adult hemoglobin tetramer consists of two alpha chains and two beta chains. Mutant beta globin causes sickle cell anemia. Absence of beta chain causes beta-zero-thalassemia. Reduced amounts of detectable beta globin causes beta-plus-thalassemia. The order of the genes in the beta-globin cluster is 5'-epsilon -- gamma-G -- gamma-A -- delta -- beta--3'. [provided by RefSeq]	

Genomic regions, transcripts, and products ↑ ?

(minus strand) Go to [reference sequence details](#) [Try our new Sequence Viewer](#)



Genomic context ↑ ?

chromosome: 11; Location: 11p15.5 [See HBB in MapViewer](#)



Table Of Contents

- Summary
- Genomic region
- products
- Genomic context
- Bibliography
- Interactions
- General gene info
- General protein
- Reference sequ
- Related sequen
- Additional links

Links

- Order cDNA clo
- BioAssay, by G
- Books
- CCDS
- Conserved Dom
- Full text in PMC
- GEO Profiles
- Genome
- HomoloGene
- Map Viewer
- Nucleotide
- OMIM
- Probe
- Protein
- PubChem Comp
- PubChem Subs
- PubMed
- PubMed (Gene
- PubMed (OMIM
- ✓ **SNP**
- ✓ **SNP: GeneView**
- ✓ **SNP: Genotype**
- ✓ **SNP: VarView**
- Taxonomy

Human Beta Hemoglobin Gene SNPs

NCBI ENTREZ **SNP** Single Nucleotide Polymorphism

All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC

Search SNP for [] Go Clear

Limits Preview/Index History Clipboard Details

Display Graphic Summary Show 20 Sort By Send to

All: 374 1000 Genomes: 10 Cited in PubMed: 1 Clinical/LSDB Submissions: 338 Human: 374

Items 1 - 20 of 374 Page 1 of 19

1: rs41486348 [*Homo sapiens*]

GCAAGGTGAACGTGGATGAAGTTGGT [(44 BP DELETED) /-] TAAGGAGACCAATAGAACTGGGCA

MapView No VarVu No PubMed GeneView SeqView No 3D No OMIM

2: rs41442644 [*Homo sapiens*]

CCTGTGGGGCAAAGTGAACGTGGATG [(7398 BP DELETED) /-] CAAGACAGGTTAAGGAGACCAATA

MapView No VarVu No PubMed GeneView Not on mRNA No 3D No OMIM

3: rs41399745 [*Homo sapiens*]

GCACTGACTCTCTCTGCCTATTGGTC [(17 BP DELETED) /-] GCTGCTGGTGGTCTACCCTTGGACC

MapView No VarVu No PubMed GeneView SeqView No 3D No OMIM

4: rs36107977 [*Homo sapiens*]

TGGGCAACCCTAAGGTGAAGGCTCAT [-/G] GCAAGAAAGTGCTCGGTGCCTTTAG

MapView VarView No PubMed GeneView SeqView No 3D No OMIM

Gene Model (mRNA alignment) information from genome sequence

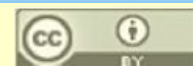
Total gene model (contig mRNA transcript):				3		
mrna	transcript	protein	mrna orientation	Contig	Contig Label	List SNP
NM_000518.4	minus strand	NP_000509.1	reverse	NT_009237.17	reference	<- currently shown
NM_000518.4	minus strand	NP_000509.1	reverse	NW_925006.1	Celera	View snp on GeneModel
NM_000518.4	minus strand	NP_000509.1	reverse	NW_001838021.1	HuRef	View snp on GeneModel

Include clinically associated
 in gene region
 cSNP
 has frequency
 double hit

gene model (contig mRNA transcript):	Contig Label	Contig	mrna	protein	mrna orientation	transcript	snp count
reference	NT_009237.17	NM_000518.4	NP_000509.1		reverse	minus strand	12, coding



Region	Contig position	mRNA pos	dbSNP rs# cluster id	Heterozygosity	Validation	3D	Clinically Associated	Function	dbSNP allele	Protein residue	Codon pos	Amino acid pos	
exon_3	4034095	465	rs41405449	N.D.				frame shift	-/AGC		1	139	
								frame shift	AGC	Ser [S]	1	139	
	4034117	430	rs41511744	N.D.					contig reference	GCTA	[AN]	1	139
									frame shift	(15bp)	[QAAYQ]	2	127
									frame shift	(17 BP DELETED)-		2	127
contig reference	(17bp)	[VQAAYQ]	2	127									
intron_2	4034199	365	rs71811954	N.D.				frame shift	(14bp)	Thr [T]	3	105	
								frame shift	(15bp)		3	105	
								frame shift	(LARGEDELETION)-		3	105	
exon_2	4035096	317	rs11549405	N.D.				synonymous	C	Leu [L]	3	89	
								contig reference	G	Leu [L]	3	89	
	4035119	294	rs11549406	0.005					missense	G	Val [V]	1	82
									contig reference	C	Leu [L]	1	82
	4035143	270	rs36209341	N.D.					missense	A		1	74
									missense	T		1	74
									contig reference	G		1	74
	4035210	203	rs17850156	N.D.					synonymous	C	Thr [T]	3	51
									contig reference	T	Thr [T]	3	51
	4035211	202	rs34676051	N.D.					missense	A	Asn [N]	2	51
contig reference									C	Thr [T]	2	51	





Search for

Limits Preview/Index History Clipboard Details

5: rs36092904 [*Homo sapiens*]

Links

CCTAAGGTGAAGGCTCATGGCAAGAA [A/T] GTGCTCGGTGCCTTTAGTGATGGCC

HGVS Names: [*NG_000007.3:g.70925A>T*] [*NM_000518.4:c.201A>T*] [*NP_000509.1:p.Lys67Asn*] [*NT_009237.17:g.4035162T>A*]

6: rs36081208 [*Homo sapiens*]

Links

ACTGAGTGAGCTGCACTGTGACAAGC [C/T] GCACGTGGATCCTGAGAACTTCAGG

HGVS Names: [*NG_000007.3:g.71014T>C*] [*NM_000518.4:c.290T>C*] [*NP_000509.1:p.Leu97Pro*] [*NT_009237.17:g.4035073A>G*]

7: rs36038739 [*Homo sapiens*]

Links

ACACTGAGTGAGCTGCACTGTGACAA [C/G/T] CTGCACGTGGATCCTGAGAACTTCA

HGVS Names: [*NG_000007.3:g.71012G>C*] [*NG_000007.3:g.71012G>T*] [*NM_000518.4:c.288G>C*] [*NM_000518.4:c.288G>T*] [*NP_000509.1:p.Lys96Asn*] [*NT_009237.17:g.4035075C>A*] [*NT_009237.17:g.4035075C>G*]

GeneView via analysis of contig annotation: [HBB](#) hemoglobin, beta

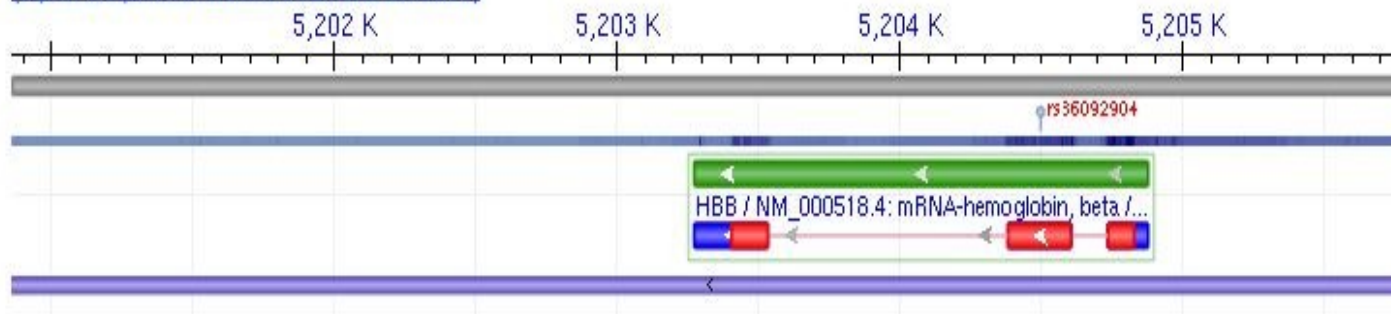
▼ View more variation on this gene (click to hide).

Include clinically associated: in gene region cSNP has frequency double hit

Assembly	SNP to Chr	Chr	Chr position	Contig	Contig position	Allele
reference	-	11	5204497	NT_009237.17	4035162	T

Function	mRNA				Protein		
	mRNA to Chr	Accession	Position	Allele change	Accession	Position	Residue change
missense	-	NM_000518.4	251	AAA ⇒ AAT	NP_000509.1	67	K [Lys] ⇒ N [Asn]

(Open sequence viewer in a new window.)



GeneView via direct blast against RefSeq sequences (used when no gene model is available): N/A

Genome Build	Chr	Chr Pos	Contig	Contig Pos	SNP to Chr	Contig allele	Group term	Group label	Contig label	Neighbor SNP
36.3	11	5366592	NW_925006.1	866916	-	T	alt_assembly_1	Celera	Celera	view
36.3	11	4907094	NW_001838021.1	876469	-	T	alt_assembly_8	HuRef	HuRef	view
36.3	11	5204497	NT_009237.17	4035162	-	T	ref_assembly	reference	reference	view
Provisional	37.1	5247921	NT_009237.18	5187921	-	T	ref_assembly	Primary_Assembly	Primary_Assembly	view

Diversity of One β -Hemoglobin SNP

Population Diversity

ss#	Sample Ascertainment				<u>Genotype Detail</u> <small>NEW</small>			Alleles	
	Population	Individual Group	Chrom. Sample Cnt.	Source	C/C	C/G	HWP	C	G
ss16249024	HapMap-CEU	European	112	IG	0.982	0.018		0.991	0.009
	HapMap-HCB	Asian	86	IG	1.000			1.000	
	HapMap-JPT	Asian	86	IG	1.000			1.000	
	HapMap-YRI	Sub-Saharan African	114	IG	1.000			1.000	

HBB Variation Viewer

Gene	HBB; hemoglobin, beta
Description	beta globin beta globin chain hemoglobin beta chain Also known as: CD113t-C
Species	Homo sapiens
Cyto	11p15.5

Gene Reference Sequences	NG_000007.1 genomic NM_000518.4 transcript NP_000509.1 protein <i>variation locations are based on</i>
Links	HGMD , Panther , Gene , ...

Observed Variation Displaying results

Var Class	Genomic	Transcript	Protein	Clinical interpretation	Freq	Pub...	MIM AI Var
DIP		c.-11A>AAAC					2	2	
DIP		c.*108A>AATAA					2	2	141900.0417
SNP	g.70453C>T						2	2	141900.0371
SNP	g.70455C>T						2	2	141900.0425
SNP	g.70457C>A g....						4	4	141900.0372 , ...
SNP	g.70458C>A g....						6	6	141900.0374
SNP	g.70459C>A g....						4	4	141900.0375
SNP	g.70513C>A						3	3	141900.0406
SNP	g.70514A>C g....						4	4	141900.0376
SNP	g.70515T>A g....						4	4	141900.0377 , ...
SNP	g.70516A>G						2	2	141900.0379
SNP	g.70517A>C g....						4	4	141900.0380 , ...
SNP	g.70545A>C	c.-50A>C					2	2	141900.0387
SNP	g.70566G>A	c.-29G>A					2	2	
SNP	g.70577C>G	c.-18C>G					2	2	
SNP	g.70595A>G	c.1A>G	p.M1V				2	2	
SNP	g.70596T>C g....	c.2T>C c.2T>G	p.M1T p.M1R				5	5	141900.0344 , ...
SNP	g.70597G>A g....	c.3G>A c.3G>...	p.M1I p.M1I p...				6	6	141900.043
SNP	g.70598G>A g....	c.4G>A c.4G>T	p.V2M p.V2L				4	4	141900.0266 , ...
MIX	g.70599T>A g....	c.5T>A c.5T>...	p.V2E p.V2A ...				7	7	141900.0069 , ...

.0077 HEMOGLOBIN FUKUOKA [HBB, HIS2TYR] dbSNP

See [Harano et al. \(1990\)](#).

.0078 HEMOGLOBIN FUKUYAMA [HBB, HIS77TYR] dbSNP

See [Hidaka et al. \(1988\)](#).

.0079 HEMOGLOBIN G (ACCRA) [HBB, ASP79ASN]

There is no clinical or hematologic abnormality in the homozygote. See [Edington](#)

.0080 HEMOGLOBIN G (COPENHAGEN) [HBB, ASP47ASN] dbSNP

See [Sick et al. \(1967\)](#), [Schiliro et al. \(1981\)](#), and [Chen et al. \(1985\)](#).

.0081 HEMOGLOBIN G (COUSHATTA) [HBB, GLU22ALA] dbSNP

HEMOGLOBIN G (SASKATOON)

HEMOGLOBIN G (HSIN-CHU)

HEMOGLOBIN G (TAEGU)

See [Schneider et al. \(1964\)](#), [Bowman et al. \(1967\)](#), [Vella et al. \(1967\)](#), [Blackwe](#)
[\(1989\)](#).

.0082 HEMOGLOBIN G (FERRARA) [HBB, ASN57LYS] dbSNP

See [Giardina et al. \(1978\)](#).

.0083 HEMOGLOBIN G (GALVESTON) [HBB, GLU43ALA] dbSNP

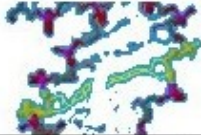

HEMOGLOBIN G (PORT ARTHUR)

HEMOGLOBIN G (TEXAS)

See [Bowman et al. \(1962, 1964\)](#).

dbSNP at NCBI

<http://www.ncbi.nlm.nih.gov/SNP/>



Single Nucleotide Polymorphism

PubMed Nucleotide Protein Genome Structure PopSet Taxonomy OMIM Books SNP

Search for SNP on NCBI Reference Assembly

Search Entrez for

BUILD 127

Have a question about dbSNP? Try searching the SNP FAQ Archive!

GENERAL

Contact Us
Site Map NEW
dbSNP Homepage
Announcements
dbSNP Summary
FTP Download

SNP SUBMISSION

DOCUMENTATION

SEARCH

HAPLOTYPE
RELATED SITES

dbSNP Search Options

Entrez SNP	ID Numbers	Submission Info	Batch	Locus Info	Between Markers
------------	------------	-----------------	-------	------------	-----------------

ANNOUNCEMENT

03/18/2007: b127 XML EMERGENCY UPDATE

Attention dbSNP user:

We have discovered an error in our assignment of functional class as it appears in the b127 XML files. The processing error affects approximately 380,000 SNPs.

Search by IDs on All Assemblies

Note: rs# and ss# must be prefixed with "rs" or "ss", respectively (i.e. rs25, ss25)

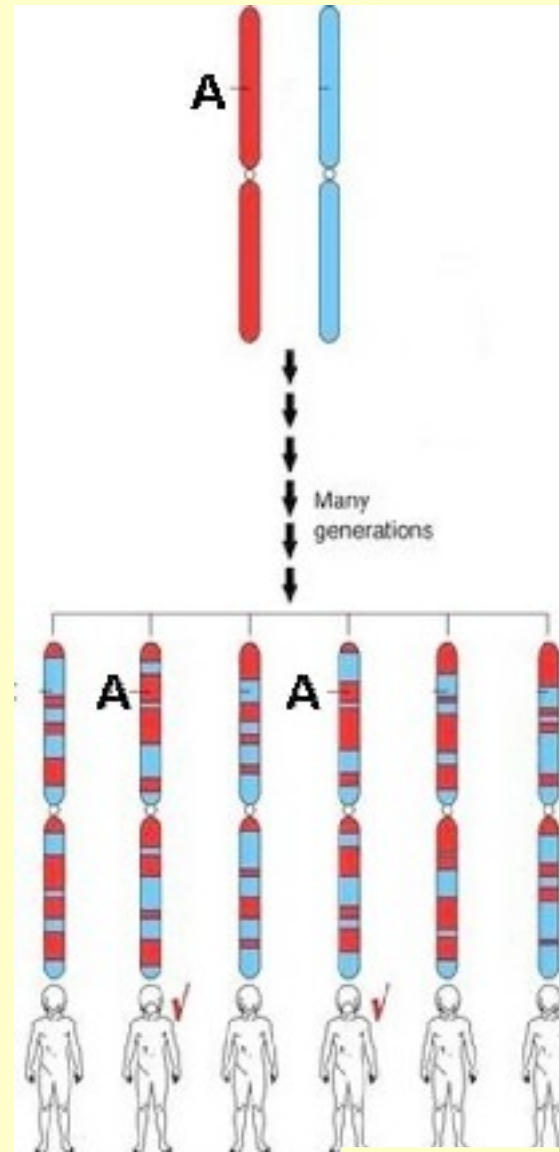
Reference cluster ID(rs#)

Submission Information

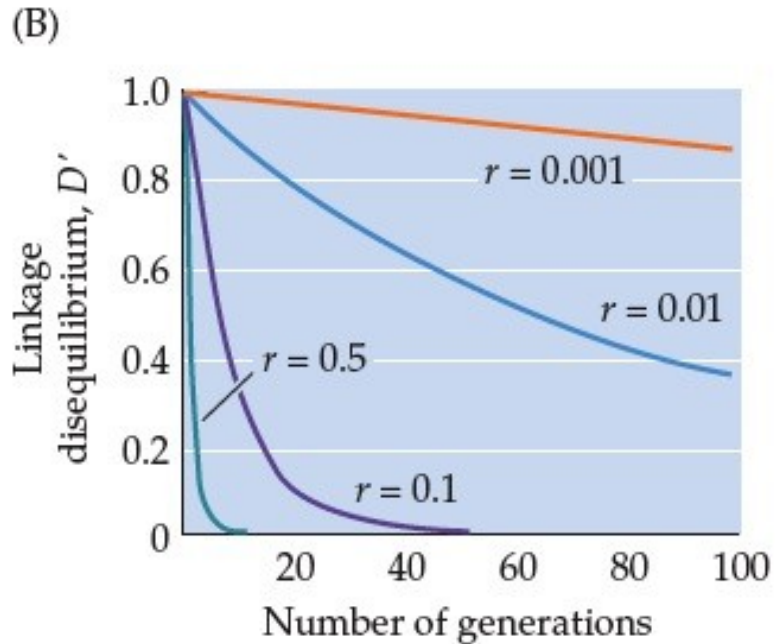
- [By Submitter](#)
- [New Batches](#)
- [Method](#)
- Population
 - [Detail](#) (Description, Handle, and ID)
 - [Class](#) (Based on geographic location)
- [Publication](#)



Origin of Haplotypes



Linkage Disequilibrium and Recombination Rate



Linkage Disequilibrium (LD) Across the Human LPL Gene

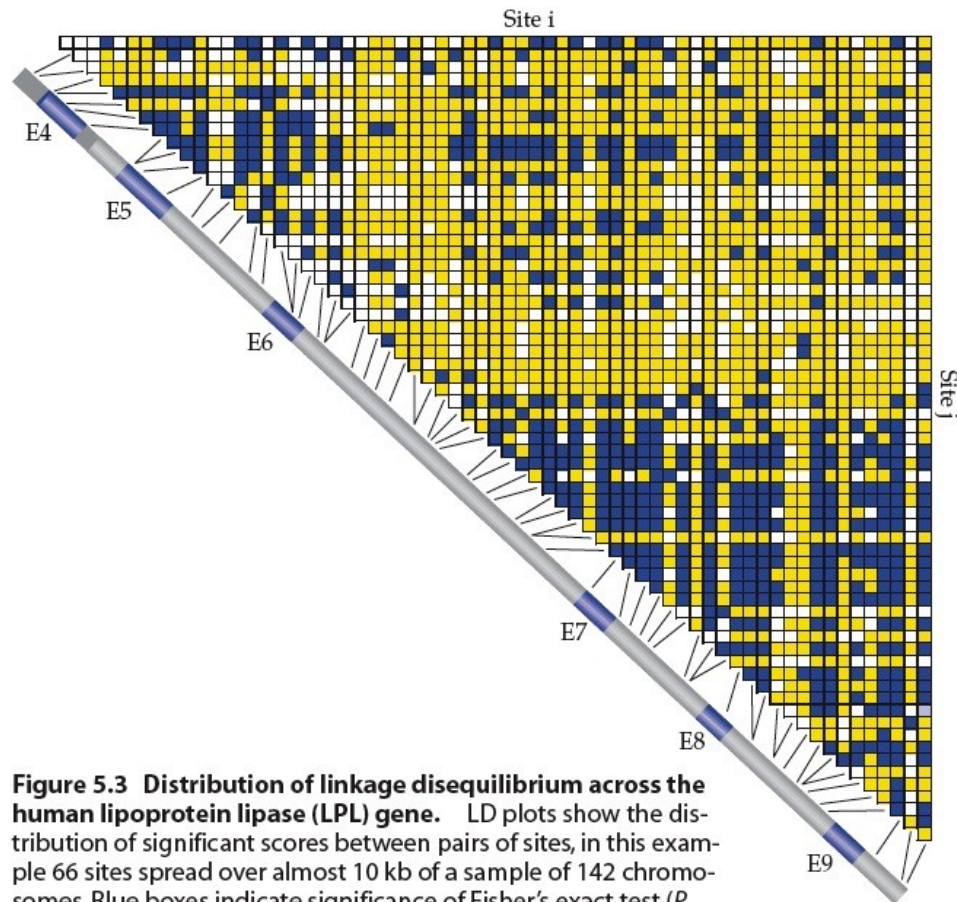
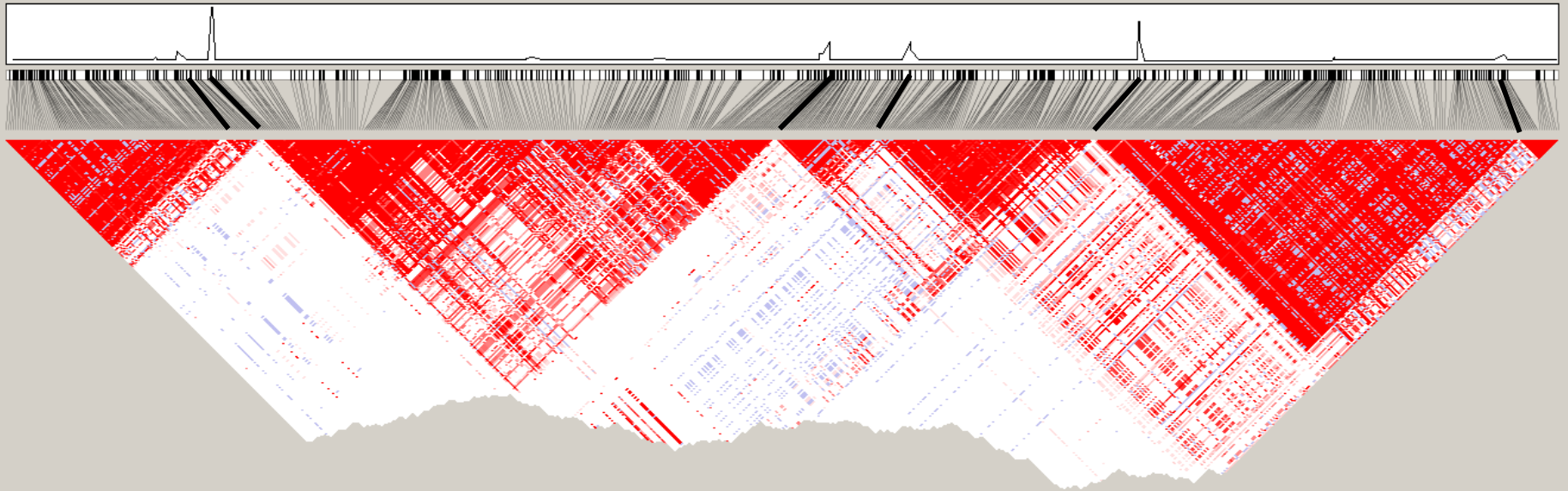


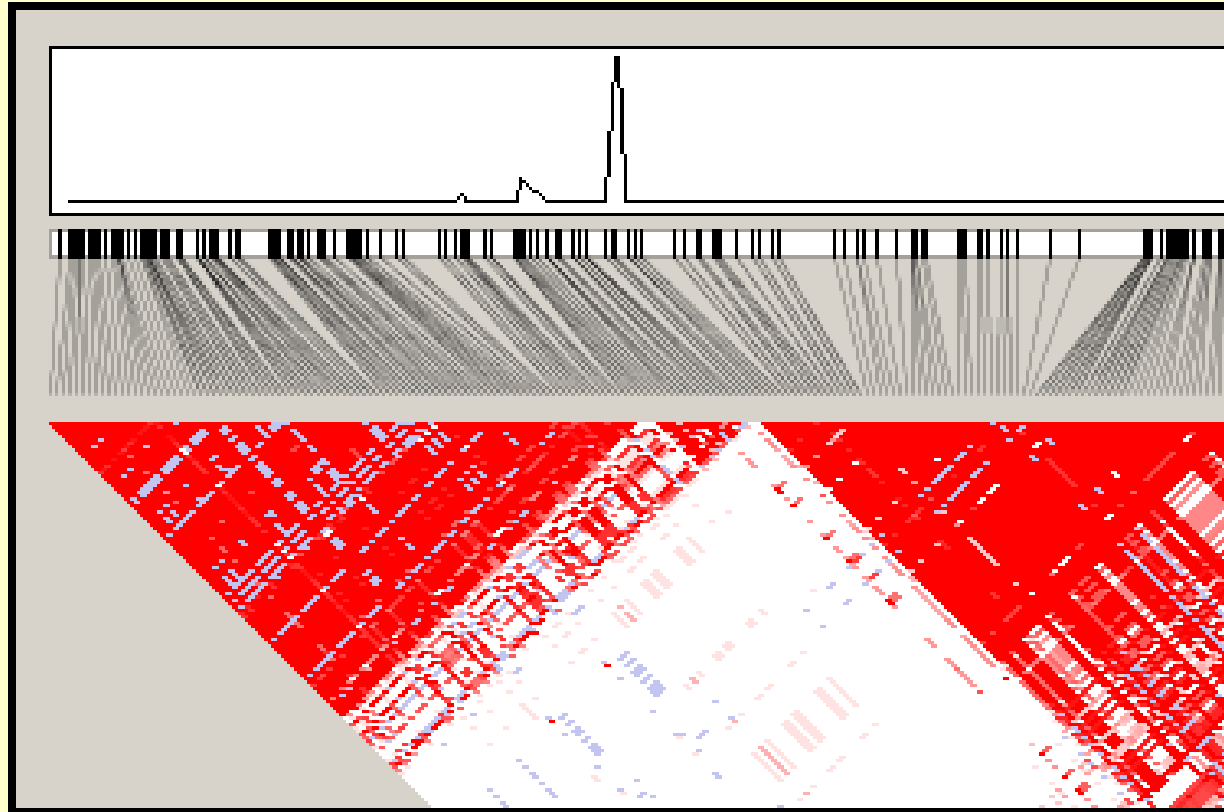
Figure 5.3 Distribution of linkage disequilibrium across the human lipoprotein lipase (LPL) gene. LD plots show the distribution of significant scores between pairs of sites, in this example 66 sites spread over almost 10 kb of a sample of 142 chromosomes. Blue boxes indicate significance of Fisher's exact test ($P < 0.001$), yellow boxes indicate nonsignificance, and white boxes are cases where there was insufficient power to test for LD at this level. Note that the extent of LD varies across the locus, and is not restricted to exon sequences. (Redrawn from Clark et al. 1998.)

Recombination hotspots are widespread and account for LD structure



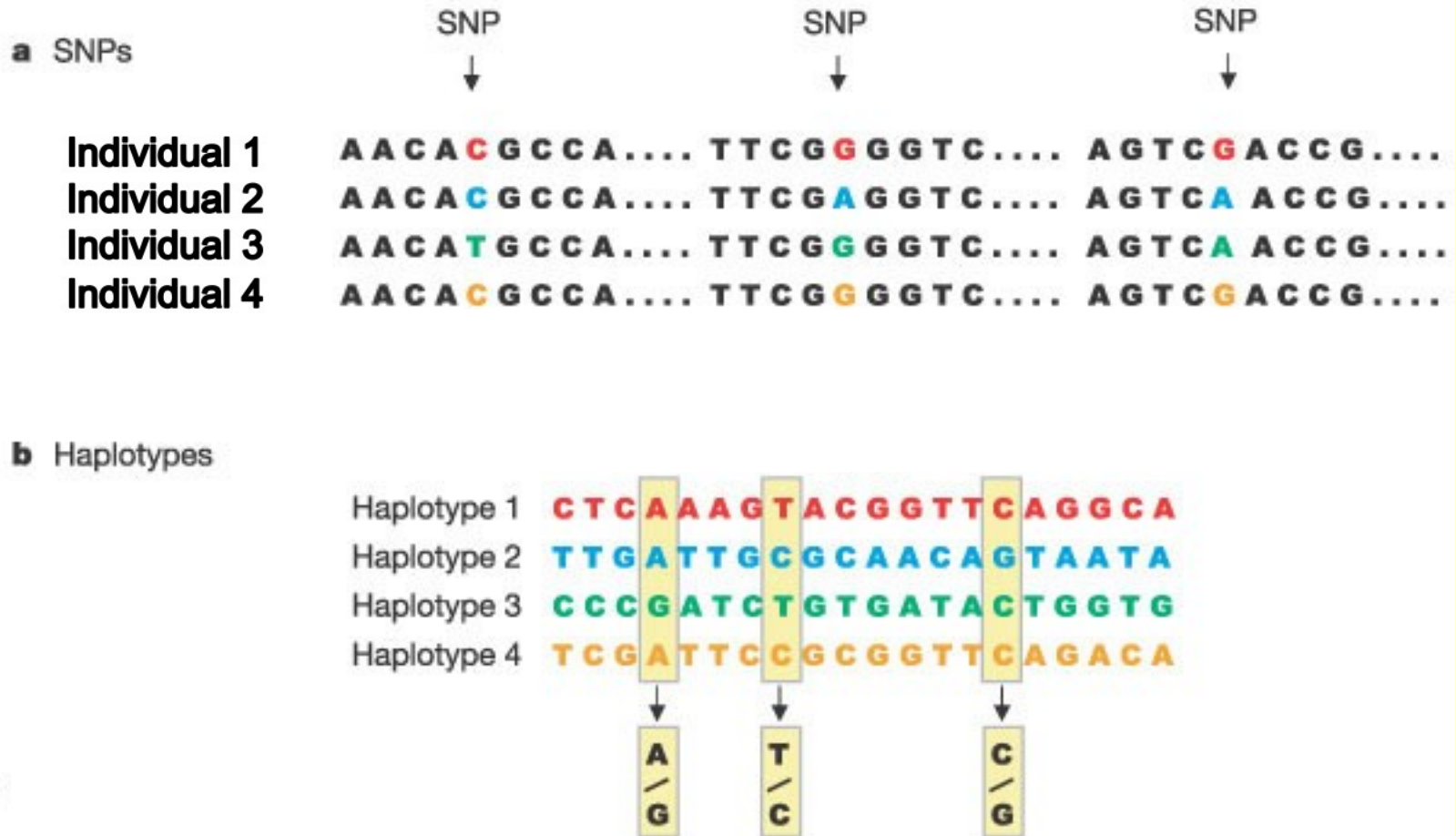
7q21

Recombination hotspots are widespread and account for LD structure

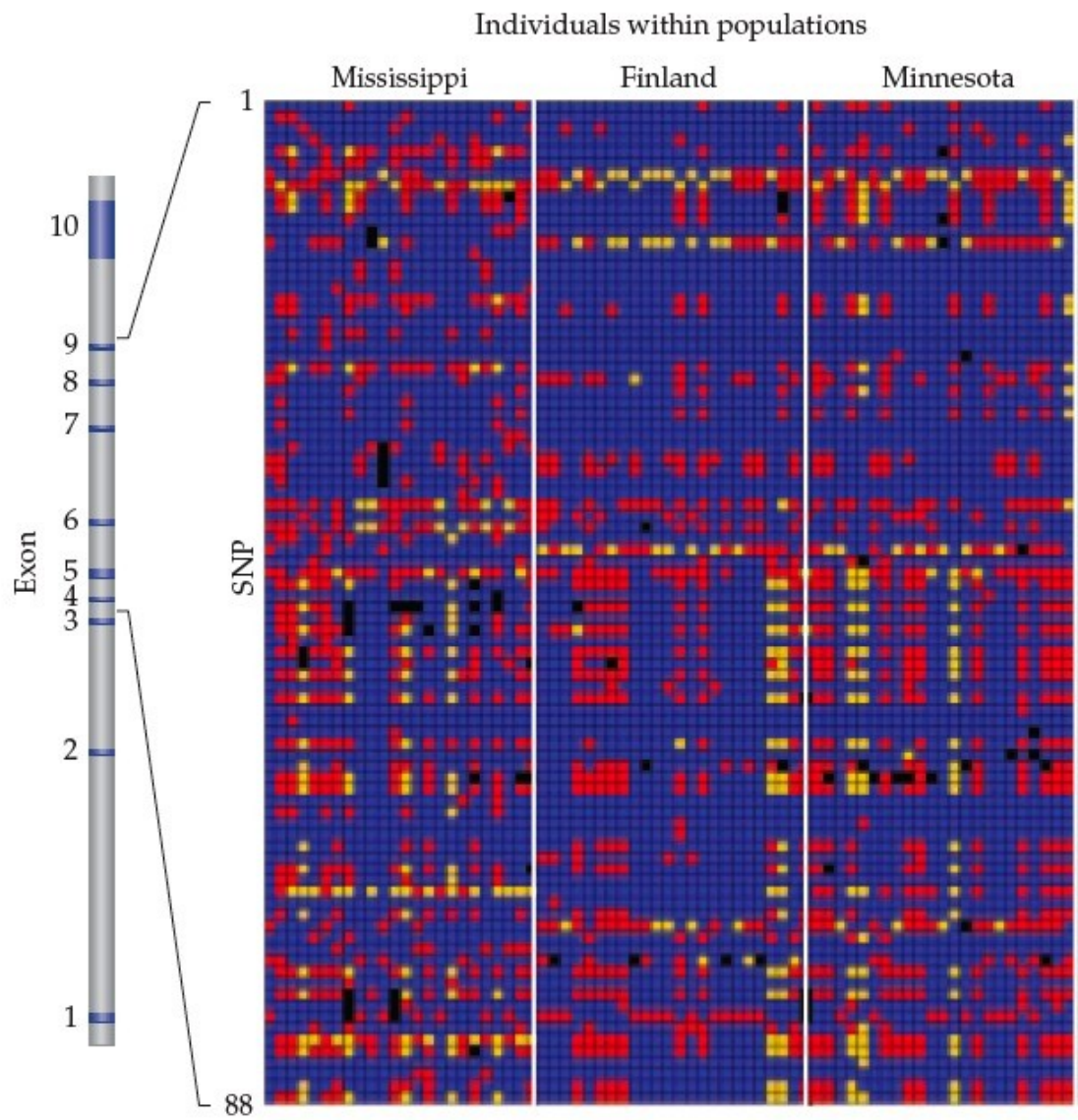


7q21

Observation of Haplotypes

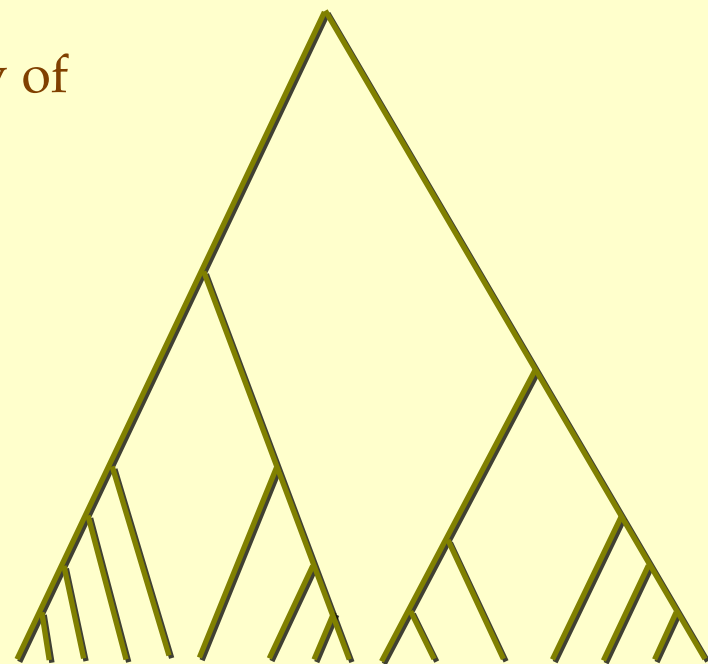


SNPs in Populations



Sequence and Distance-Based Phylogenies (evolutionary trees)

- Sequence-Based Methods (Parsimony)
 - Assigns mutations to branches
 - Minimize number of changes
 - Topology maximizes similarity of neighboring leaves
- Distance-based methods
 - Branch lengths = $D(i,j) / 2$ for sequences i, j
 - Distances must be metric
 - Distances can reflect time or number of changes
 - Distances must be relatively constant per unit branch length



nature



THE HAPMAP PROJECT

Chapter and verse on
human genetic variation





International HapMap Project

<http://www.hapmap.org/>

International HapMap Project

[Home](#) | [About the Project](#) | [Data](#) | [Publications](#) | [Tutorial](#)

[中文](#) | [English](#) | [Français](#) | [日本語](#) | [Yoruba](#)

The International HapMap Project is a partnership of scientists and funding agencies from Canada, China, Japan, Nigeria, the United Kingdom and the United States to develop a public resource that will help researchers find genes associated with human disease and response to pharmaceuticals. See "[About the International HapMap Project](#)" for more information.

Project Information

[About the Project](#)
[HapMap Publications](#)
[HapMap Tutorial](#)
[HapMap Mailing List](#)
[HapMap Project Participants](#)
[HapMap Mirror Site in Japan](#)

Project Data

[HapMap Genome Browser \(Phase 1, 2 & 3 - merged genotypes & frequencies\)](#)
[HapMap Genome Browser \(Phase 3 - genotypes, frequencies & LD\)](#)
[HapMap Genome Browser \(Phase 1 & 2 - full dataset\)](#)
[GWAs Karyogram](#)
[HapMart](#)
[HapMap FTP](#)
[Bulk Data Download](#)
[Data Freezes for Publication](#)
[ENCODE Project](#)
[Guidelines For Data Use](#)

News

• 2009-12-14: **Notice to Haploview users**

Recently, there are several questions about Haploview data format errors, and these errors were observed when users tried to analyze HapMap release 27 data dumped from HapMap. The current Haploview version (4.1) does not work with release 27 data. Haploview will generate a software error similar to "Hapmap data format error: NA06984" when trying to open the data.

The r27 data format will be supported by next Haploview version. There is a beta test version that is supposed to work and it can be obtained from <http://www.broadinstitute.org/haploview/haploview-downloads>. But since it is NOT an official release version, please use it base on your own judgment.

• 2009-12-10: **Corrected HapMap3 phased haplotypes available for chromosome X**

Phased haplotypes for consensus HapMap3 release 2 data for chromosome X has been corrected and the new data are now [available for bulk download](#). Sorry for any inconvenience this might have caused.

• 2009-12-02: **HapMap3 phased haplotypes available for chromosome X**

Phased haplotypes for consensus HapMap3 release 2 data has been phased for chromosome X and are now available for bulk download. [Update: The downloading was disabled because several users have found that there are repeating data in some of the chrX phasing data files. The data source is being contacted and the downloading will be enabled as soon as the problem is cleared.]



International HapMap Project

<http://www.hapmap.org/>



International HapMap Project

Home | About the Project | Data | Publications | Tutorial

■ Instructions

Searching: Search using a sequence name, gene name, locus, or other landmark. The wildcard character * is allowed.

Navigation: Click one of the rulers to center on a location, or click and drag to select a region. Use the Scroll/Zoom buttons to change magnification and position.

Examples : [Chr20](#), [Chr9:660,000..760,000](#), [SNP:rs6870660](#), [NM_153254](#), [BRCA2](#), [5q31](#), [ENm010](#), [gwa*](#), [PARK3](#).

[\[Help\]](#) [\[Reset\]](#)

■ Search

Help links:

- [- LD -](#)
- [- tagSNPs -](#)
- [- Phased Haplotype -](#)
- [- Genotype data -](#)
- [- Frequency data -](#)
- [- Symbols and colours used -](#)

Landmark or Region :

Reports & Analysis :

Data Source

Population descriptors: **ASW:** African ancestry in Southwest USA, **CEU:** Utah residents with Northern and Western European ancestry from the CEPH collection, **CHB:** Han Chinese in Beijing, China, **CHD:** Chinese in Metropolitan Denver, Colorado, **GIH:** Gujarati Indians in Houston, Texas, **JPT:** Japanese in Tokyo, Japan, **LWK:** Luhya in Webuye, Kenya, **MEX:** Mexican ancestry in Los Angeles, California, **MKK:** Maasai in Kinyawa, Kenya, **TSI:** Tuscans in Italy, **YRI:** Yoruban in Ibadan, Nigeria.

For performing in depth LD and Haplotype analysis of genotype data, install [Haploview](#) in your local machine. Haploview (ver 4.1) is currently available for download. This version does not handle hapmap3 samples. Please check the [Haploview website](#) for updates.

■ Tracks

■ Overview All on All off

<input checked="" type="checkbox"/> dbSNP SNPs/500Kb	<input checked="" type="checkbox"/> GWA studies (NHGRI Catalog)	<input checked="" type="checkbox"/> NT contigs	
<input checked="" type="checkbox"/> gt'd SNPs/500Kb	<input checked="" type="checkbox"/> Ideogram	<input checked="" type="checkbox"/> OMIM disease associations	

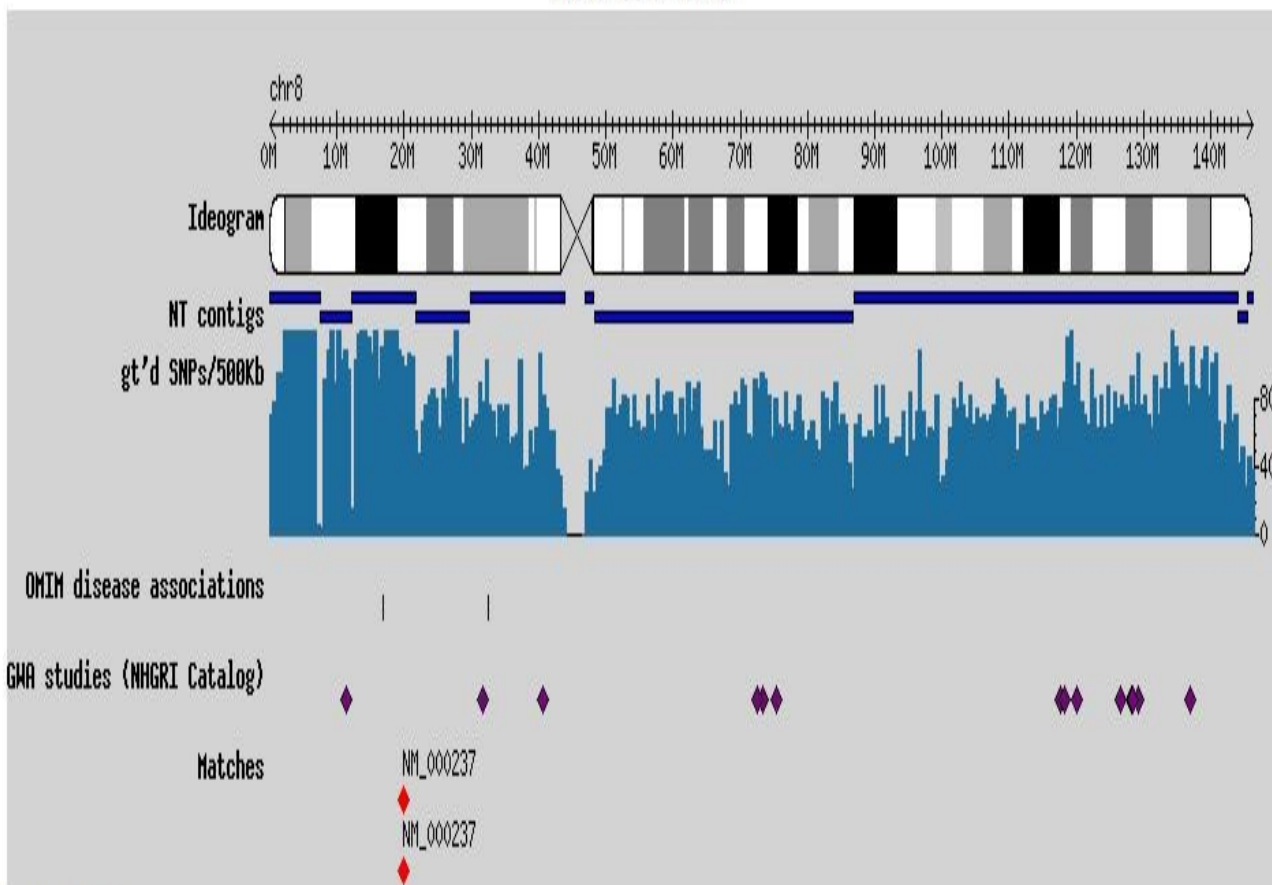
■ Region All on All off

<input checked="" type="checkbox"/> Copy Number Variation	<input type="checkbox"/> Entrez genes	<input type="checkbox"/> GWA studies (NHGRI Catalog)	
<input type="checkbox"/> dbSNP SNPs/20Kb	<input checked="" type="checkbox"/> gt'd SNPs/20Kb	<input type="checkbox"/> OMIM disease associations	

■ Copy Number Variation All on All off

<input type="checkbox"/> Deletions (Conrad et al.)	<input type="checkbox"/> Genomic Variants (Iafate et al.)	<input type="checkbox"/> Genomic Variants (Redon et al.)	<input type="checkbox"/> Genomic Variants (Simon-Sanchez et al.)
<input type="checkbox"/> Deletions (Hinds et al.)	<input type="checkbox"/> Genomic Variants (Locke et al.)	<input type="checkbox"/> Genomic Variants (Sebat et al.)	<input type="checkbox"/> Genomic Variants (Tuzun et al.)
<input type="checkbox"/> Deletions (McCarroll et al.)	<input type="checkbox"/> Genomic Variants (Mills et al.)	<input type="checkbox"/> Genomic Variants (Sharp et al.)	<input type="checkbox"/> Genomic Variants (Wong et al.)

Matches on chr8



NM_000237 lipoprotein lipase precursor

chr8:19.84..19.87 score=27.49
Mbp (27.99 kbp)

NM_000237 LPL encodes **lipoprotein lipase**, which is expressed in heart, muscle, and adipose tissue. LPL functions as a homodimer, and has the dual functions of triglyceride hydrolase and ligand/bridging factor for receptor-mediated **lipoprotein** uptake. Severe mutations that cause LPL deficiency result in type I hyper**lipoproteinemia**, while less extreme mutations in LPL are linked to many disorders of **lipoprotein** metabolism. Publication Note: This RefSeq record includes a subset of the publications that are available for this gene. Please see the Entrez Gene record to access additional publications.

chr8:19.84..19.87 score=26.49
Mbp (27.99 kbp)

Overview

Ideogram

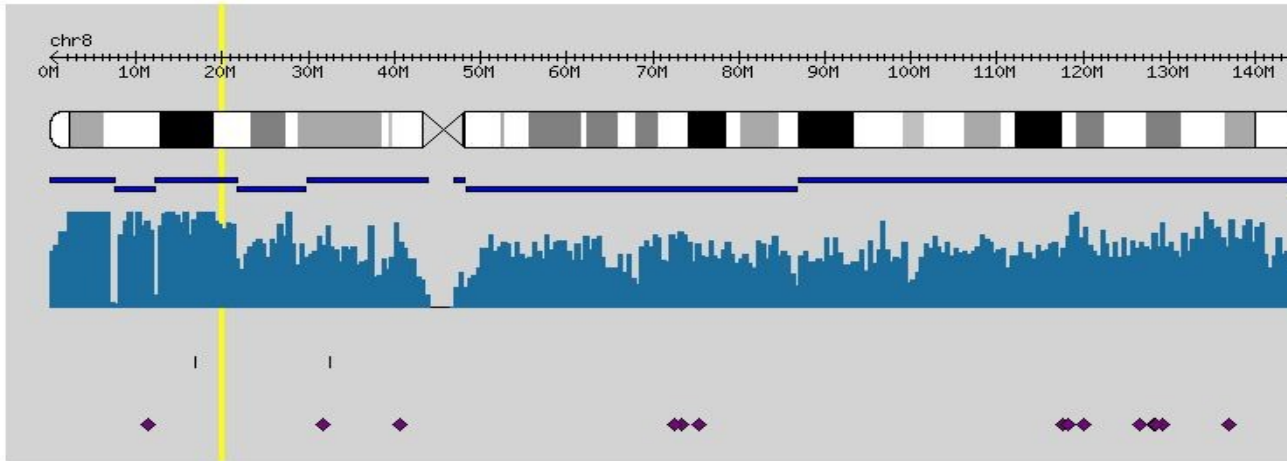
NT contigs

gt'd SNPs/500Kb

OMIM disease associations

GWA studies (NHGRI Catalog)

dbSNP SNPs/500Kb

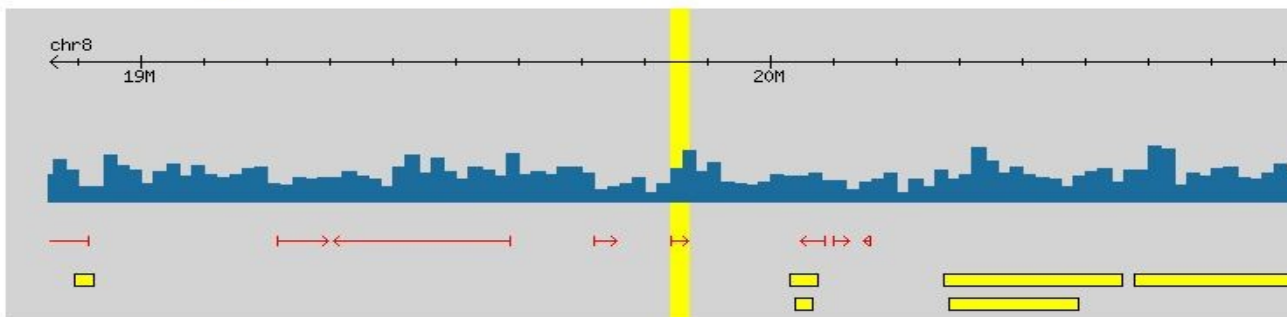


Region

gt'd SNPs/20Kb

Entrez genes

Copy Number Variation

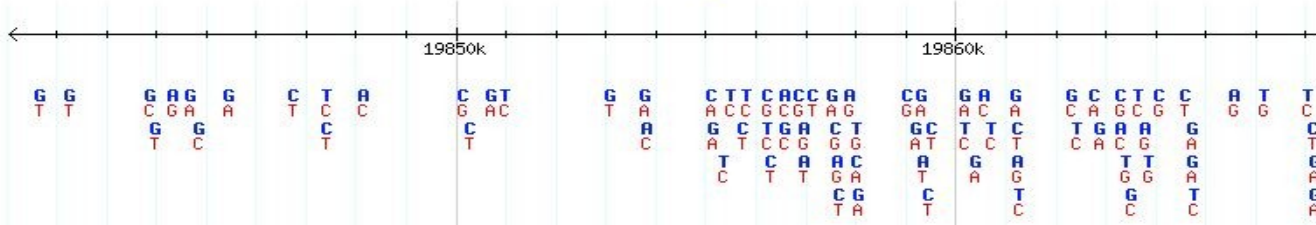


Details

Genotyped SNPs

Entrez genes

Reactome pathways



NM_000237

LPL: lipoprotein lipase precursor



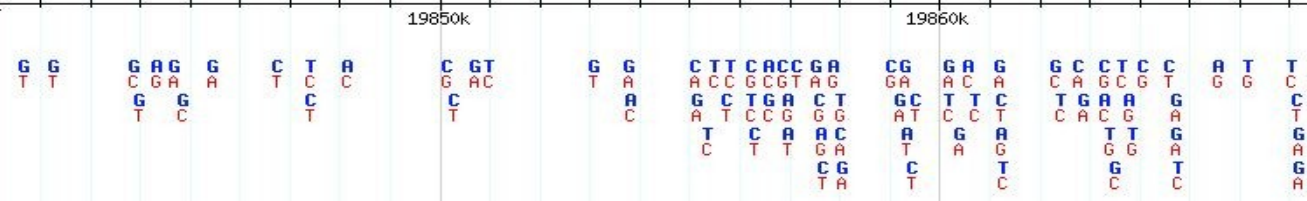
Genome-Wide Association studies (NHGRI Catalog)

LD Plot

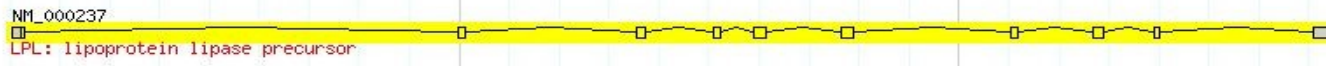
CEU:lod



Genotyped SNPs



Entrez genes

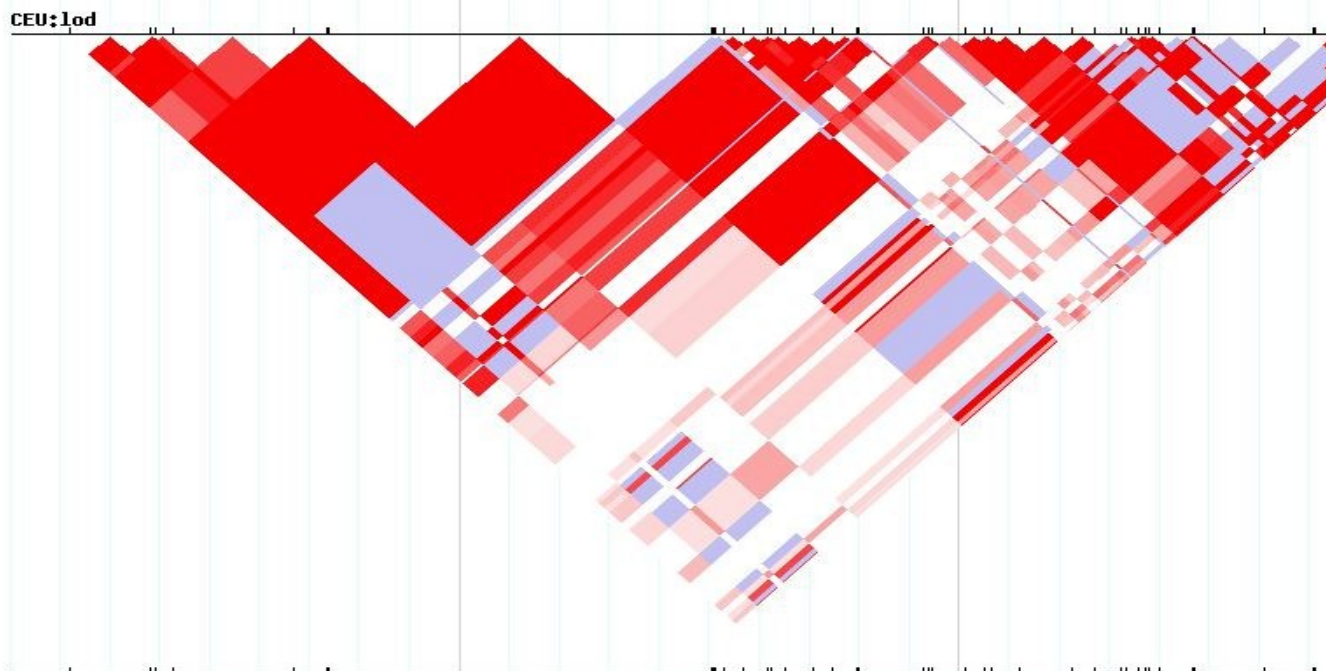


Reactome pathways

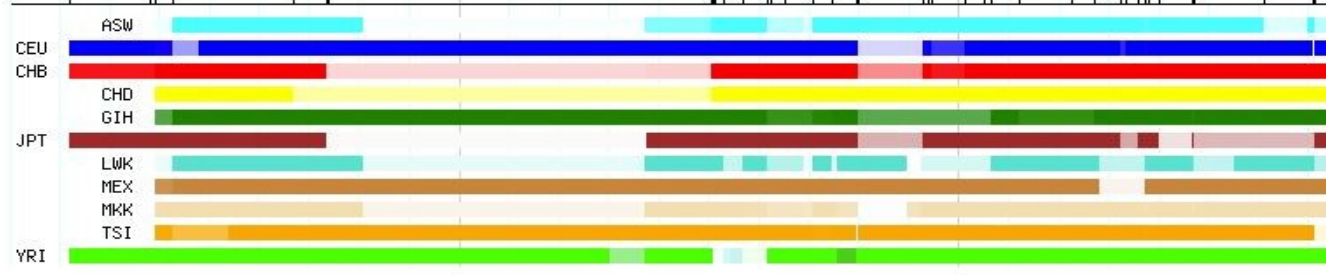


Genome-Wide Association studies (NHGRI Catalog)

LD Plot



LD Heat Plot



1000 Genomes

A Deep Catalog of Human Genetic Variation

[Home](#) [About](#) [Partners](#) [Data](#) [Contact](#) [Wiki](#)

1000 GENOMES PROJECT DATA RELEASE

SNP data downloads and genome browser representing four high coverage individuals

The first set of SNP calls representing the preliminary analysis of four genome sequences are now available to download through the [EBI FTP site](#) and the [NCBI FTP site](#). The README file dealing with the FTP structure will help you find the data you are looking for.

The data can also be viewed directly through the 1000 Genomes browser at <http://browser.1000genomes.org>. Launch the browser and [view a sample region here](#).

More information about the data release can be found in the [data section](#) of this web site.

Download the 1000 Genomes Browser Quick Start Guide

[Quick start \(pdf\)](#)

LOG IN

Username:

Password:

([I forgot my password](#))

LINKS



[Download the meeting report](#)



[View the participants](#)

PRESS RELEASE

WEDNESDAY JUN. 11, 2008

[Three Sequencing Companies Join 1000 Genomes Project](#)

TUESDAY JAN. 22, 2008

[International Consortium Announces the 1000 Genomes Project](#)

1000 Genomes

A Deep Catalog of Human Genetic Variation

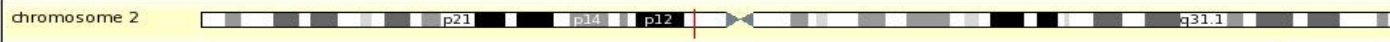
Location: 2:85,333,054-85,433,054

Location-based displays

- Whole genome
- Chromosome summary
- Region overview
- Region in detail**
- Comparative Genomics
 - Genomic alignments (0)
 - Synteny (0)
- Genetic Variation
 - Resequencing (10)
 - Linkage Data
- Markers

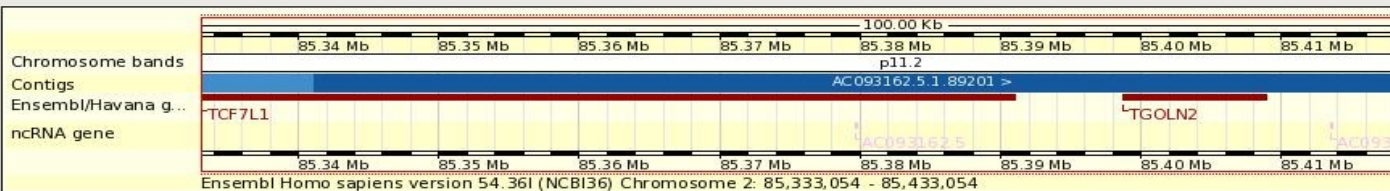
- Configure this page
- Add custom data to page
- Export data
- Bookmark this page

Chromosome 2: 85,333,054-85,433,054

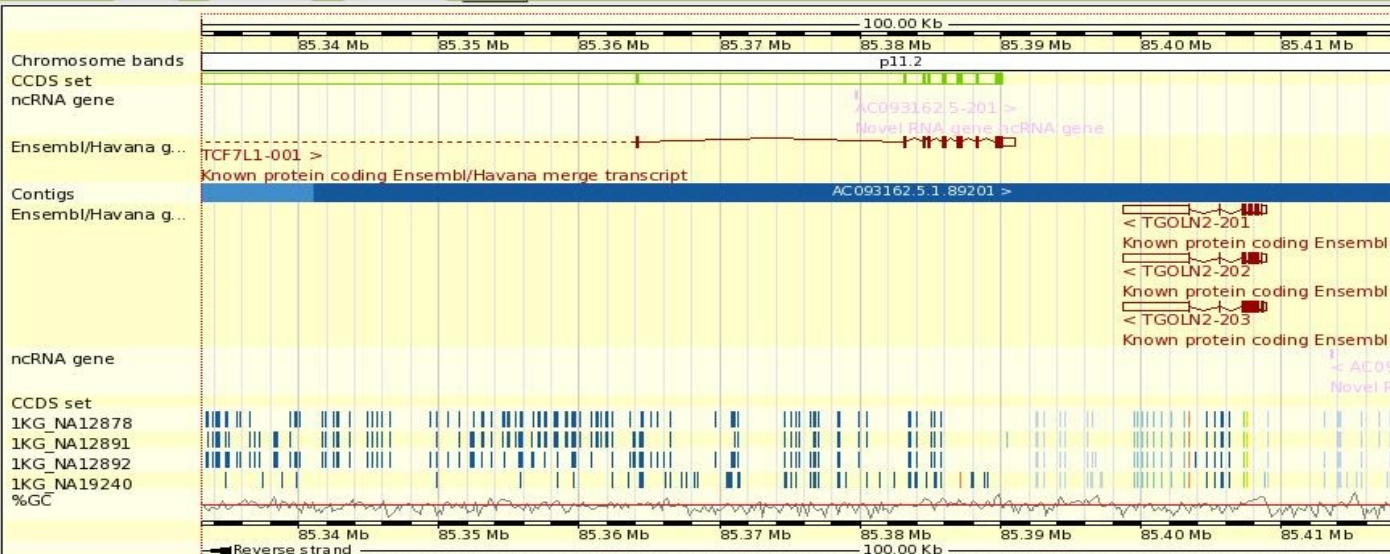


Region overview

Region in detail



Location: 2 : 85333054 - 85433054



Gene Legend

- Known protein coding
- Novel RNA gene

Variation Legend

- Intronic
- Splice site SNP
- Upstream
- Downstream
- Synonymous coding
- 5' UTR
- 3' UTR
- Non-synonymous

There are currently 69 tracks turned off.
Ensembl Homo sapiens version 54.361 (NCBI36) Chromosome 2: 85,333,054 - 85,433,054

10,000 Genomes Project Evolutionary Biology

<http://www.genome.gov/>

genome.gov
National Human Genome Research Institute
National Institutes of Health

Google™ Search Search

Home | About NHGRI | Newsroom | Staff

Research Grants Health Policy & Ethics Educational Resources Careers & Training

Research

Intramural Research
Extramural Research
Population Genomics

Grants

Funding Opportunities
Active Grants Database
Minority Funding

Health

Genetic Disorders FAQ
Clinical Research FAQ
GARD Center

Policy & Ethics

Genetic Discrimination
Genetic Testing
Legislative Database

Educational Resources

Fact Sheets
National DNA Day
Talking Glossary

Careers & Training

Research Training
Educational Opps
Working at NHGRI

Newsroom

About NHGRI

Enhancing Peer Review Application Changes PDF

Read about shortened and restructured application forms and instructions



Recovery Act at NHGRI

ADONIS Highlights NOW

10K Genomes Project to Create Vertebrate Genome Zoo

As DNA sequencing costs decline, a group of research biologists have pondered the possibility of reading evolution's notebooks in the genomes of thousands of vertebrate species. Using a workshop to organize their ideas, the group proposed The 10K Genomes Project in a paper published Nov. 5 in the *Journal of Heredity*. Co-authors include NHGRI's Eric Green M.D., Ph.D. and Adam Felsenfeld, Ph.D.



Play Video

Drs. Green and Felsenfeld discuss the project in this video interview

Scientists propose a "genome zoo" of 10,000 vertebrate species

Press release from the University of California at Santa Cruz



Hay! First Analysis of Horse Genome

An international research team, supported in part by NHGRI, has found that the genome of the domestic horse, *Equus caballus*, has a structure remarkably similar to our own genome. Published in the Nov. 6 issue of *Science*, the team's landmark analysis also sheds new light on the centromere evolution.

[Horse genome sequence and analysis published in Science](#)
[From NHGRI's Talking Glossary of Genetic Terms: Centromeres](#)

The Human Genome Project



[Read the story behind the Human Genome Project](#)

Newsroom

RSS Help

[Study Conclusively Ties Rare Disease Gene to Parkinson's](#)
October 21, 2009

[NHGRI Launches Improved Online Talking Glossary of Genetic Terms](#)
October 20, 2009

[NIH Funds Four Centers of Excellence in Genomic Science](#)
September 28, 2009

Calendar

RSS Help

[DIR Seminar Series](#)
Rhoda Alani, M.D.
November 19, 2009

[Evolutionary conserved pathways suppress](#)

