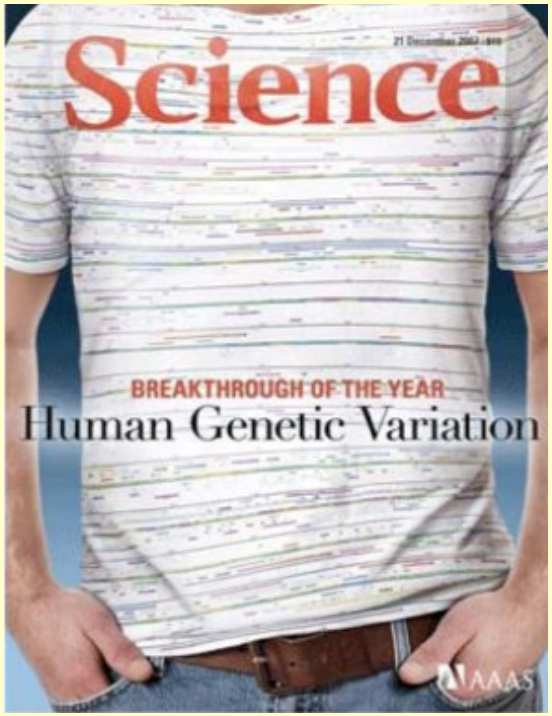
Genomics, Bioinformatics & Medicine

http://biochem158.stanford.edu/

Simple Nucleotide Polymorphisms

http://biochem158.stanford.edu/SNPs.html

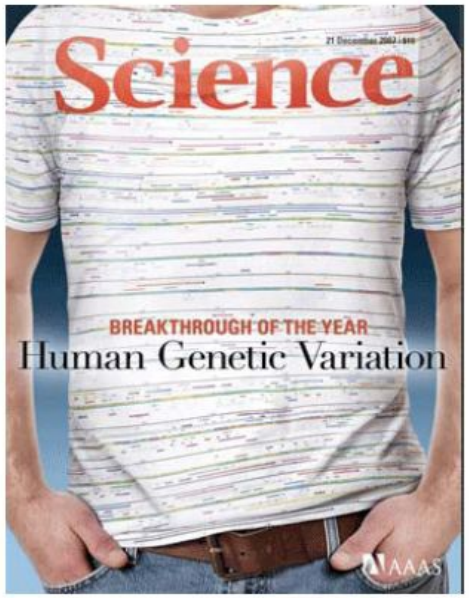


Doug Brutlag
Professor Emeritus of Biochemistry & Medicine
Stanford University School of Medicine

Human Genetic Variation
2007 Scientific Breakthrough of the Year

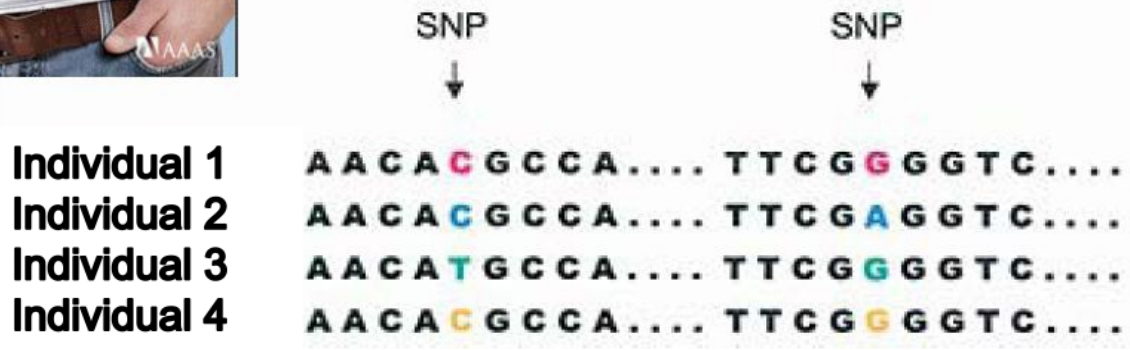
2007 SCIENTIFIC BREAKTHROUGH OF THE YEAR

Science Magazine, December 21, 2007



"It's all about me!"

Single Nucleotide Polymorphisms (SNPs)

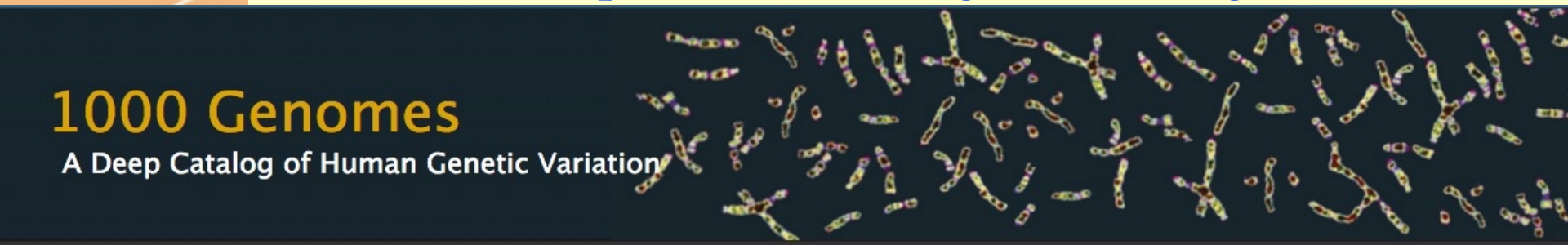


International HapMap Project

http://www.hapmap.org/

Thousand Genomes Project

http://www.1000genomes.org/



Home About Data Analysis Participants Contact Browser Wiki FTP search Search

LATEST ANNOUNCEMENTS

WEDNESDAY SEPTEMBER 30, 2015

A global reference for human genetic variation

The Phase 3 publication, A global reference for human genetic variation and the Phase 3 Structural variation publication, An integrated map of structural variation in 2,504 human genomes are now available from Nature alongside a celebration of 25 years of the Human Genome Project

The variants from the Phase 3 analysis are available in ftp/release/20130502/ and extended information about the SV dataset can be found in ftp/phase3/integrated_sv_map/.

Both these papers are open access and should be free for everyone to read and download.

If you have any questions about the data these papers are based on or how to access it please email info@1000genomes.org

http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/

Recent project announcements

FRIDAY OCTOBER 16, 2015

GRCh38 mapping of the Illumina Platinum Genomes CEU pedigree

NAVIGATION

- Frequently Asked Questions

LINKS

All Project Announcements

Sample and Project Information

Media Archive

Find the 1000 Genomes Project Publications

1000 Genomes
A Deep Catalog of Human Genetic Variation

A Global Reference for Human Genetic Variation

http://www.nature.com/nature/journal/v526/n7571/full/nature15393.html

A global reference for human genetic variation

The 1000 Genomes Project Consortium*

The 1000 Genomes Project set out to provide a comprehensive description of common human genetic variation by applying whole-genome sequencing to a diverse set of individuals from multiple populations. Here we report completion of the project, having reconstructed the genomes of 2,504 individuals from 26 populations using a combination of low-coverage whole-genome sequencing, deep exome sequencing, and dense microarray genotyping. We characterized a broad spectrum of genetic variation, in total over 88 million variants (84.7 million single nucleotide polymorphisms (SNPs), 3.6 million short insertions/deletions (indels), and 60,000 structural variants), all phased onto high-quality haplotypes. This resource includes >99% of SNP variants with a frequency of >1% for a variety of ancestries. We describe the distribution of genetic variation across the global sample, and discuss the implications for common disease studies.

Nature **526, 68-74 (October 1, 2015)**

1000 Genomes
A Deep Catalog of Human Genetic Variation

An Integrated Map of Structural Variation in 2,504 Human Genomes

An integrated map of structural variation in 2,504 human genomes

A list of authors and their affiliations appears at the end of the paper.

Structural variants are implicated in numerous diseases and make up the majority of varying nucleotides among human genomes. Here we describe an integrated set of eight structural variant classes comprising both balanced and unbalanced variants, which we constructed using short-read DNA sequencing data and statistically phased onto haplotype blocks in 26 human populations. Analysing this set, we identify numerous gene-intersecting structural variants exhibiting population stratification and describe naturally occurring homozygous gene knockouts that suggest the dispensability of a variety of human genes. We demonstrate that structural variants are enriched on haplotypes identified by genome-wide association studies and exhibit enrichment for expression quantitative trait loci. Additionally, we uncover appreciable levels of structural variant complexity at different scales, including genic loci subject to clusters of repeated rearrangement and complex structural variants with multiple breakpoints likely to have formed through individual mutational events. Our catalogue will enhance future studies into structural variant demography, functional impact and disease association.

Nature **526, 75-81 (October 1, 2015)**

NHS 100,000 Genomics England Project

http://www.genomicsengland.co.uk/

Genomics england

Home | About Genomics England | The 100,000 Genomes Project | GeCIP | Library and resources | News | Contact us

Genomics England, with the consent of participants and the support of the public, is creating a lasting legacy for patients, the NHS and the UK economy through the sequencing of 100,000 genomes: the 100,000 Genomes Project.

Genomics England was set up by the Department of Health to deliver the 100,000 Genomes Project. Initially the focus will be on rare disease, cancer and infectious disease. The project is currently in its pilot phase and will be completed by the end of 2017.

Read more...

NIH Precision Medicine Initiative

http://www.nih.gov/precisionmedicine/

National Institutes of Health
Turning Discovery Into Health

Search

For Employees | Staff Directory | En Español

Health Information | Grants & Funding | News & Events | Research & Training | Institutes at NIH | About NIH

NIH Home > Research & Training

PRECISION MEDICINE INITIATIVE

Precision Medicine Initiative

- What are the near-term goals?
- What are the longer-term goals?
- How is it different?
- Who will participate?
- NIH Workshop

Precision Medicine Initiative

Far too many diseases do not have a proven means of prevention or effective treatments. We must gain better insights into the biology of these diseases to make a difference for the millions of Americans who suffer from them. Precision medicine is an emerging approach for disease treatment and prevention that takes into account individual variability in genes, environment, and lifestyle for each person. While significant advances in precision medicine have been made for select cancers, the practice is not currently in use for most diseases. Many efforts are underway to help make precision medicine the norm rather than the exception. To accelerate the pace, President Obama has now unveiled the Precision Medicine Initiative – a bold new enterprise to revolutionize medicine and generate the scientific evidence needed to move the concept of precision medicine into every day clinical practice.

Email Updates

To sign up for updates please enter your e-mail address.

Submit

Related Links

- NEJM Perspective: A New Initiative on Precision Medicine
- White House Precision Medicine Web Page
- White House Fact Sheet: President Obama's Precision Medicine Initiative
- Precision Medicine Initiative and Cancer Research

Single Nucleotide Polymorphisms (SNPs) in the Human Genome

GCTGTATGACTAGAAGATCGAT
GCTGTATGACGAGAAGATCGAT

- About 38 million sites in the human genome where sequence variations have occurred
- About 15 million sites where variation exceeds 1% of a particular population (MAF > 1%)
- Each ethnicity has its own distribution of SNPs
- About 3 million sites where any individual varies from the consensus human genome.
- Each person differs from others in 3 million places (about 0.1% of the genome)
- SNP sequence variations are common, unlike disease causing mutations which are rare.

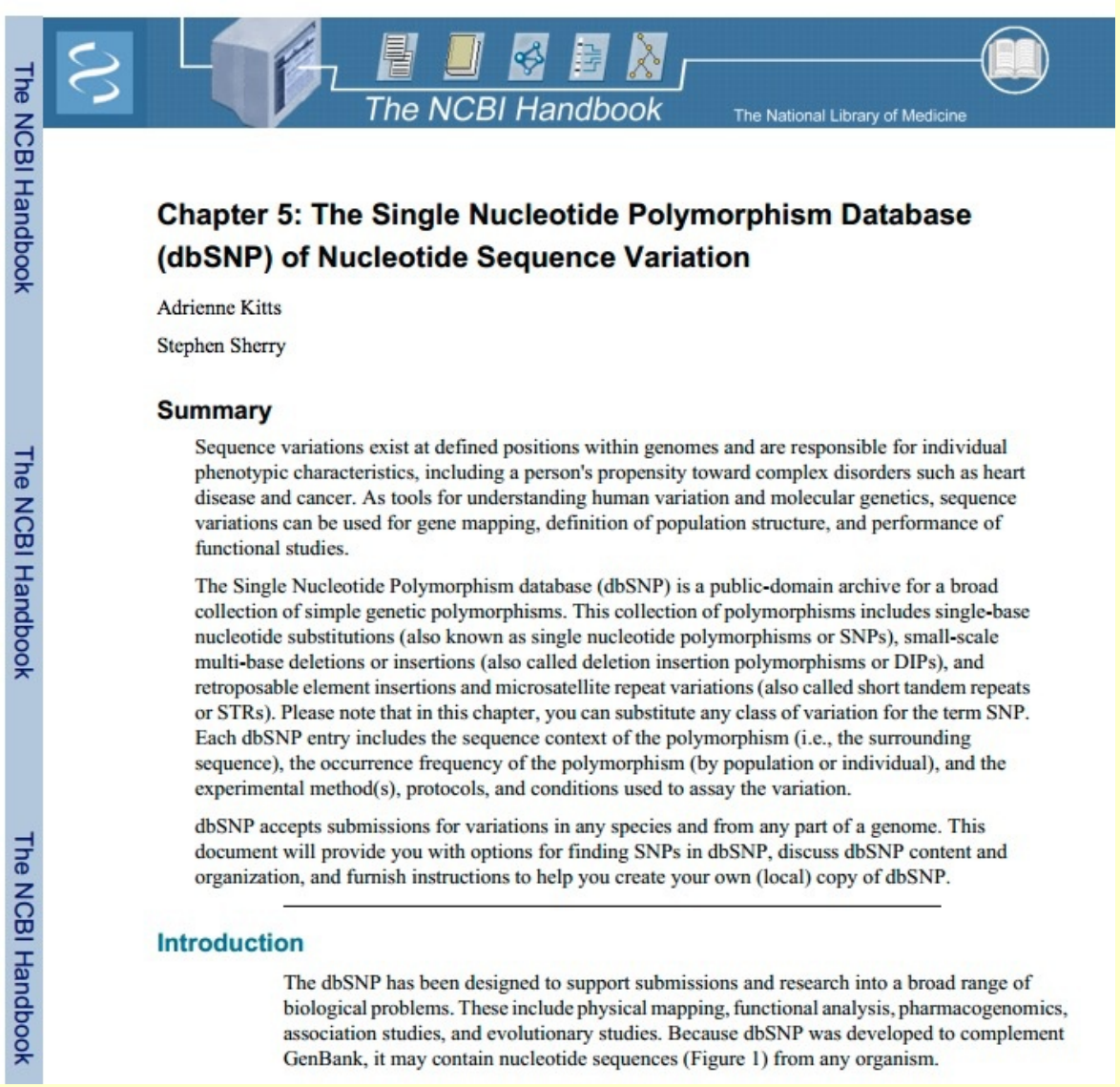
Single Nucleotide Polymorphisms (SNPs) in the Human Genome

GCTGTATGACTAGAAGATCGAT
GCTGTATGACGAGAAGATCGAT

- SNPs can be used for identifying individuals and forensics
- SNPs are used for mapping & genome-wide association studies of complex diseases
- SNPs are used for ancestry tracking & family relationships
- SNPs are used to predict risk of common genetic diseases
- SNPs are used for classifying patients in clinical trials
- SNPs are used to predict drug sensitivity an adverse reactions
- SNPs are used for personalized medicine & pharmacogenomics
- While SNPs are linked with disease, they do not cause disease
- In short, SNPs are used as genetic markers

The dbSNP Database

http://www.ncbi.nlm.nih.gov/books/NBK21088/pdf/ch5.pdf



The NCBI Handbook

The National Library of Medicine

The NCBI Handbook

The NCBI Handbook

The NCBI Handbook

Chapter 5: The Single Nucleotide Polymorphism Database (dbSNP) of Nucleotide Sequence Variation

Adrienne Kitts

Stephen Sherry

Summary

Sequence variations exist at defined positions within genomes and are responsible for individual phenotypic characteristics, including a person's propensity toward complex disorders such as heart disease and cancer. As tools for understanding human variation and molecular genetics, sequence variations can be used for gene mapping, definition of population structure, and performance of functional studies.

The Single Nucleotide Polymorphism database (dbSNP) is a public-domain archive for a broad collection of simple genetic polymorphisms. This collection of polymorphisms includes single-base nucleotide substitutions (also known as single nucleotide polymorphisms or SNPs), small-scale multi-base deletions or insertions (also called deletion insertion polymorphisms or DIPs), and retroposable element insertions and microsatellite repeat variations (also called short tandem repeats or STRs). Please note that in this chapter, you can substitute any class of variation for the term SNP. Each dbSNP entry includes the sequence context of the polymorphism (i.e., the surrounding sequence), the occurrence frequency of the polymorphism (by population or individual), and the experimental method(s), protocols, and conditions used to assay the variation.

dbSNP accepts submissions for variations in any species and from any part of a genome. This document will provide you with options for finding SNPs in dbSNP, discuss dbSNP content and organization, and furnish instructions to help you create your own (local) copy of dbSNP.

Introduction

The dbSNP has been designed to support submissions and research into a broad range of biological problems. These include physical mapping, functional analysis, pharmacogenomics, association studies, and evolutionary studies. Because dbSNP was developed to complement GenBank, it may contain nucleotide sequences (Figure 1) from any organism.

The dbSNP Database

http://www.ncbi.nlm.nih.gov/books/NBK174586/

The Database of Short Genetic Variation (dbSNP)

Kitts A, Phan L, Ward M, et al.

Scope

Sequence variation is of scientific interest to population geneticists, genetic mappers, and those investigating relationships among variation and phenotype. These variations can be of several types, from simple substitutions that do not affect sequence length, to those that result in minor length differences, to those that affect multiple genes and multiple chromosomes. Variations can also be categorized with respect to their frequency within a population, from a variation with a single allele to a variation that is highly polymorphic.

Although SNP is the abbreviation for "single nucleotide polymorphism," dbSNP is a public archive of all short sequence variation, not just single nucleotide substitutions that occur frequently enough in a population to be termed polymorphic. dbSNP includes a broad collection of simple genetic variations such as single-base nucleotide substitutions, small-scale multi-base deletions or insertions, and microsatellite repeats. Data submitted to dbSNP can be from any organism, from any part of a genome, and can include genotype and allele frequency data if those data are available. dbSNP accepts submissions for all classes of simple sequence variation, and provides access to variations of germline or somatic origin that are clinically significant.

In order to emphasize the comprehensive nature of dbSNP's content, the full name of the database was changed from "database of Single Nucleotide Polymorphism" to the more inclusive "database of Short Genetic Variation" in July of 2011. The acronym that represents the database will remain "dbSNP" to avoid any confusion that might arise from a complete name change.

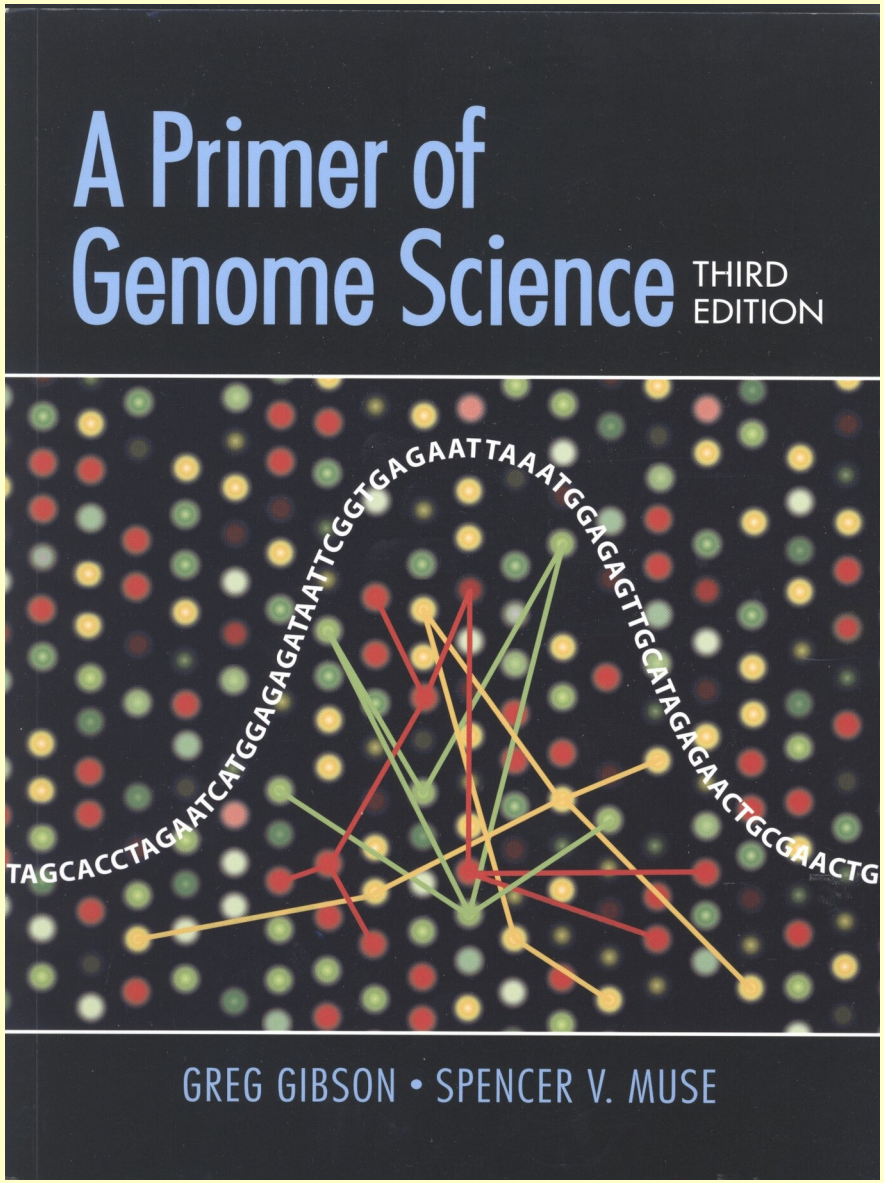
Each record in dbSNP includes the sequence context of the variant, the frequency of the polymorphism in a population if available, its zygosity if available, and the experimental method(s), protocols, and conditions used to assay the variation by each submitter. Individual submissions are clustered into dbSNP reference records (rs#) that contain summary data which may include clinical significance from ClinVar, association with phenotype from dbGaP, variation false positive status, allele origin (germline or somatic), and submitter attributes.

The dbSNP has been designed to support submissions and research into a broad range of biological problems that include the identification of genotype-phenotype relationships, genetic and physical mapping, functional analysis, pharmacogenomics, and association studies.

Medical Genetics

Advances in next-generation sequencing technologies allow

A Primer of Genome Science
Chapter 3 Genomic Variation



Single Nucleotide Polymorphisms (SNPs)

- SNPs are common variations in the genome (minor allele frequency or MAF between 50% and 1%)
- Most SNPs are genetically neutral
  - Used in DNA fingerprints - forensics
  - Paternity tests
  - Immigration in the US and United Kingdom
  - Used to track ethnic migrations and ancestry
- Some SNPs reflect distinguishing characteristics
  - Often the basis for racial & genetic discrimination or other stigma
- Rarer variations cause disease. Unlike SNPs, these variations are rare, often called mutations.
- Some SNPs linked to predisposition to disease
- SNPs can serve as genetic markers for other traits
  - Clinical trials associate SNPs with drug efficacy
  - Clinical trials associate SNPs adverse drug reactions
  - Personal genomics associate SNPs with traits

Types of SNPs

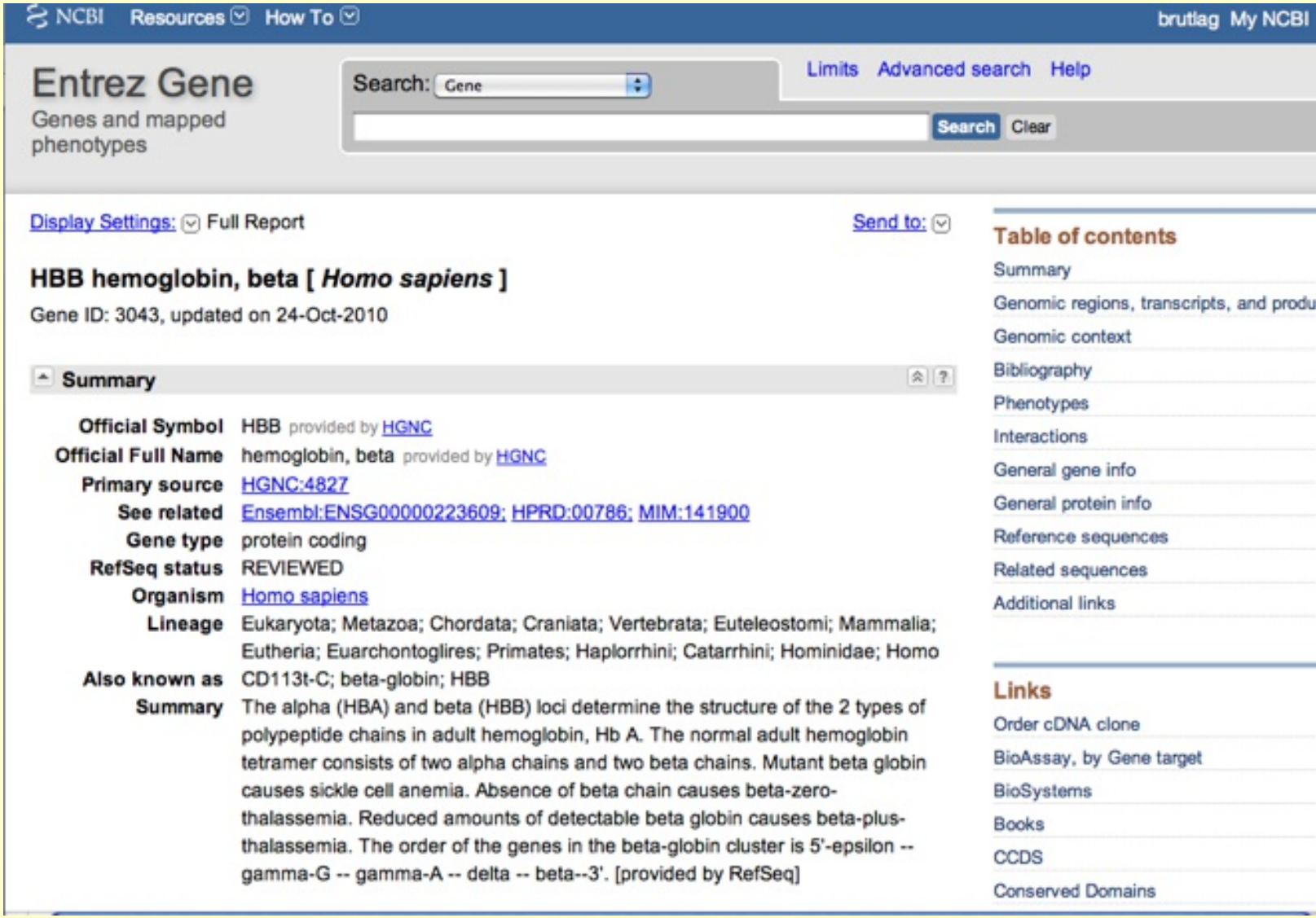
http://www.ncbi.nlm.nih.gov/sites/entrez?db=snp

- Non protein coding SNPs
  - Promoters
  - 5′ UTR
  - 3′ UTR
  - Introns
  - Intergenic Regions
  - Pseudogenes
  - Regulatory
    - Splicing
    - Transcriptional regulation (promoter & transcription factor 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

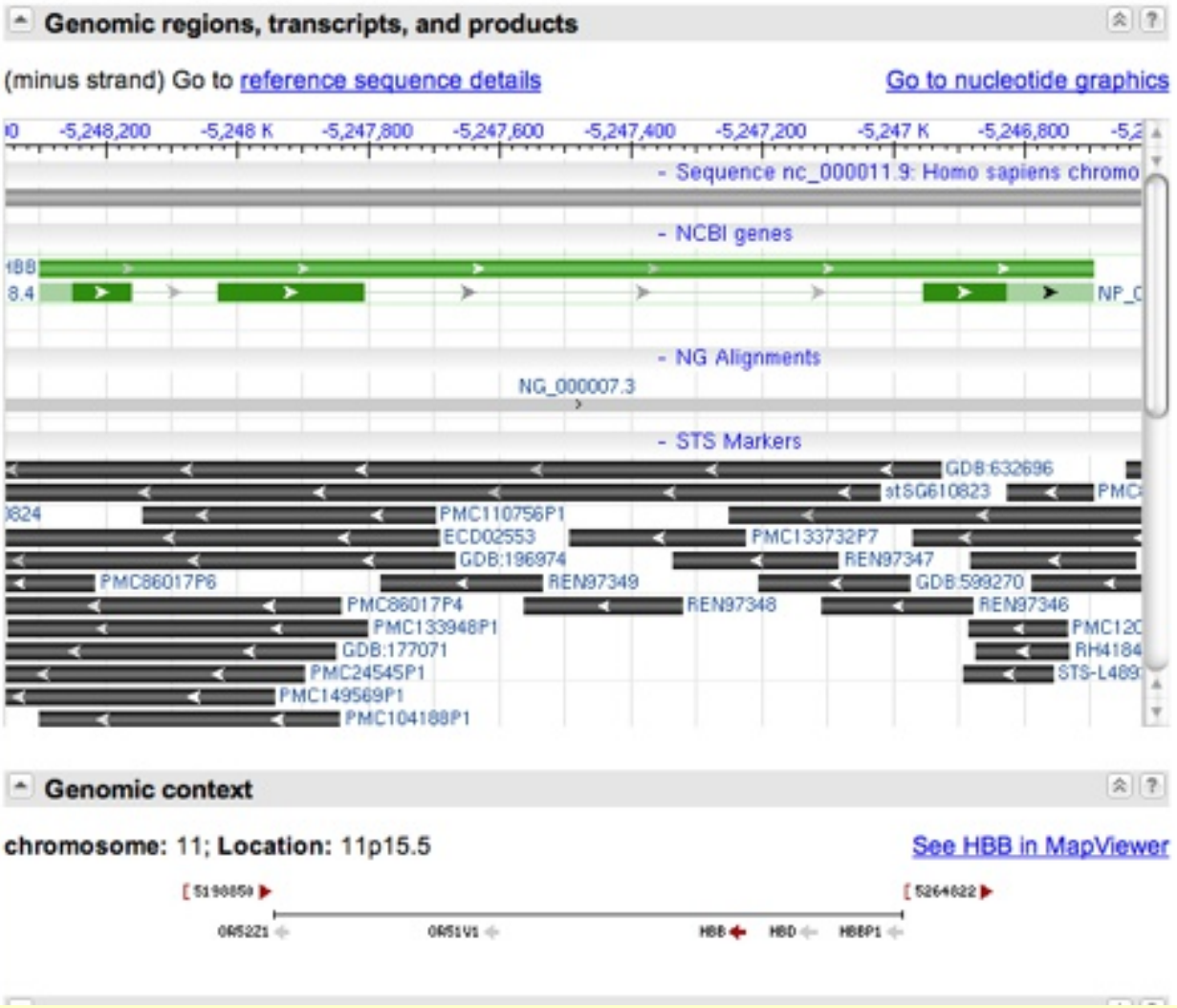
Human β-Hemoglobin Gene

http://www.ncbi.nlm.nih.gov/gene/3043



Human β-Hemoglobin Gene

http://www.ncbi.nlm.nih.gov/gene/3043



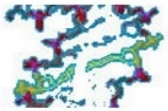
Full text in PMC
GEO Profiles
Genome
HomoloGene
Map Viewer
Nucleotide
OMIM
Peptidome
Probe
Protein
PubChem Compound
PubChem Substance
PubMed
PubMed (GeneRIF)
PubMed (OMIM)
RefSeq Proteins
RefSeq RNAs
RefSeqGene
SNP
SNP: GeneView
SNP: Genotype
SNP: VarView
Taxonomy
UniSTS

Human β-Hemoglobin Gene SNPs

http://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?locusId=3043

NCBI

dbSNP
Short Genetic Variations



PubMed Nucleotide Protein Genome Structure PopSet Taxonomy OMIM Books SNP

Search for SNP on NCBI Reference Assembly

Search Entrez SNP for Go

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

GENERAL
RSS Feed
Contact Us
Site Map
dbSNP Homepage
Announcements
dbSNP Summary
FTP Download
HUMAN VARIATION
SNP SUBMISSION
DOCUMENTATION
SEARCH
RELATED SITES

SNP linked to Gene (geneID:3043) Via Contig Annotation

The SNP GeneView page only reports human variation on GRCh38. A new Variation Viewer is available to view the gene HBB variations in GRCh37p13 or GRCh38, and will replace SNP GeneView later this year. Please visit the Help Page or YouTube for available features and send your comments and suggestions to NCBI helpdesk.

Send rs# on all gene models to Batch Query Download all rs# to file.

Gene Model (mRNA alignment) information from genome sequence

Total gene model (contig mRNA transcript):				1		
mrna	transcript	protein	mrna orientation	Contig	Contig Label	List SNP
NM_000518.4	minus strand	NP_000509.1	reverse	NT_009237.19	GRCh38	<- currently shown

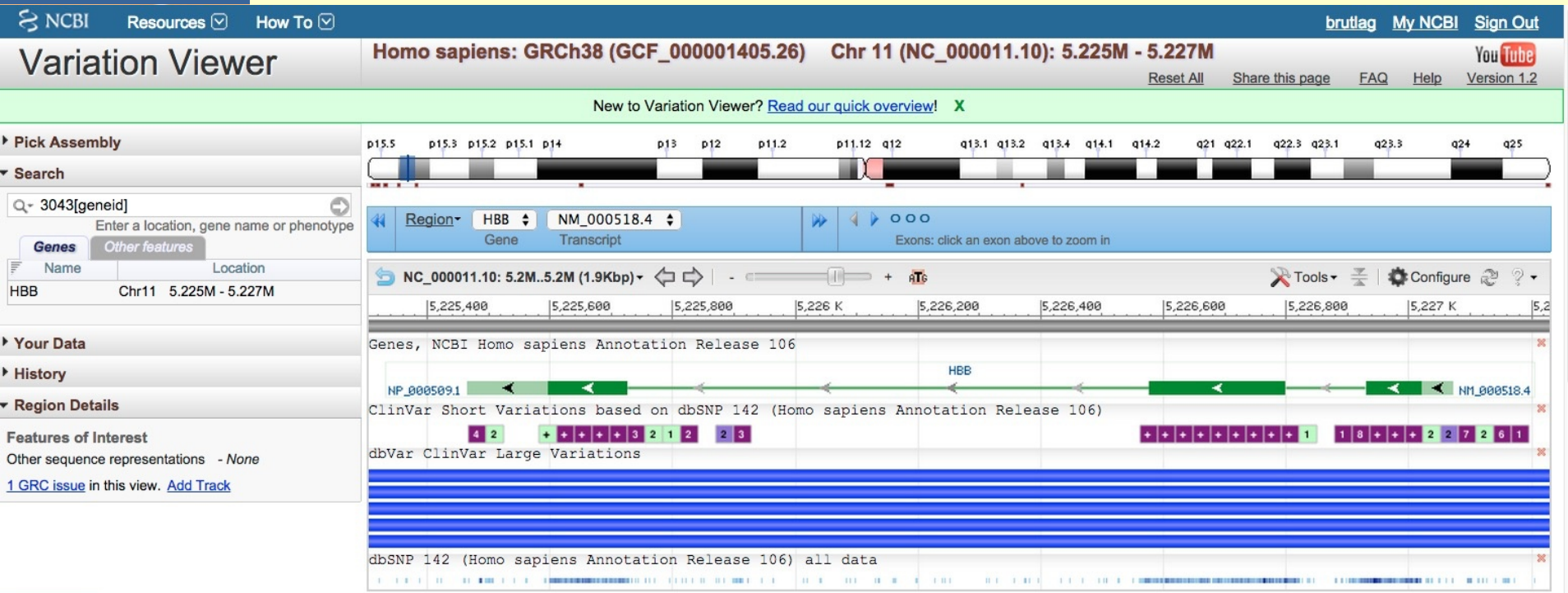
Clinical Source in gene region cSNP has frequency double hit refresh

gene model (contig mRNA transcript):	Contig Label	Contig	mrna	protein	mrna orientation	transcript	snp count
	GRCh38	NT_009237.19	NM_000518.4	NP_000509.1	reverse	minus strand	15, coding

Region	Chr. position	mRNA pos	dbSNP rs# cluster id	Hetero-zygosity	Validation	MAF	Allele origin	3D	Linkout	Function	dbSNP allele	Protein residue	Codon pos	Amino acid pos	PubMed
	5225653	439	rs369582912	N.D.				Yes		frame shift	AA	Glu [E]	2	130	
										frame shift	AC	Asp [D]	2	130	
										frame shift	CC	Leu [L]	2	131	
	5226625	317	rs11549405	N.D.				Yes		synonymous	C	Leu [L]	3	89	
										contig reference	G	Leu [L]	3	89	
	5226648	294	rs11549406	0.005	H			Yes		missense	G	Val [V]	1	82	
										contig reference	C	Leu [L]	1	82	
	5226685	257	rs112287010	0.002				Yes		synonymous	T	Leu [L]	3	69	
										contig reference	C	Leu [L]	3	69	
	5226739	203	rs17850156	N.D.				Yes		synonymous	C	Thr [T]	3	51	
										contig reference	T	Thr [T]	3	51	
	5226740	202	rs34676051	N.D.				Yes		missense	A	Asn [N]	2	51	
										contig reference	C	Thr [T]	2	51	
	5226759	183	rs373212989	N.D.				Yes		missense	C	Pro [P]	1	45	
										contig reference	T	Ser [S]	1	45	

Human β-Hemoglobin Variation Viewer

http://www.ncbi.nlm.nih.gov/variation/view/



Variation Data

Filter by

Source database
- dbSNP (632)
- dbVar (37)

In ClinVar
- Yes (378)
- No (291)

Worst clinical significance
- Pathogenic (109)
- Likely pathogenic (11)
- drug response (0)
- other (248)
- risk factor (0)

More...

Download | Edit columns | Items 1 - 30 of 669 | << First | < Prev | Page 1 of 23 | Next > | Last >>

Variant ID	Location	Variant type	Gene	Molecular consequences	Worst clinical significance	1000G MAF	GO-ESP MAF	Publications
nsv931147	61,793 - 10,727,969	copy number variation	PNPLA2 and 268 more		Pathogenic			1
nsv915986	196,855 - 5,321,874	copy number variation	PNPLA2 and 153 more		Pathogenic			1
nsv984845	198,510 - 135,074,876	copy number variation	SPTBN2 and 1515 more					1
nsv532276	202,758 - 31,726,224	copy number variation	TRIM5 and 387 more		Pathogenic			1
nsv1054121	205,983 - 6,415,299	copy number variation	TRIM5 and 192 more					1
nsv1048536	205,983 - 17,160,103	copy number variation	TRIM5 and 304 more					1
nsv1037023	205,983 - 30,840,538	copy number variation	TRIM5 and 382 more					1

β-Hemoglobin Gene SNP rs111645889

http://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?rs=111645889

Reference SNP(refSNP) Cluster Report: rs111645889

RefSNP		Allele		HGVS Names
Organism:	human (*Homo sapiens*)	**Variation Class:**	SNP: single nucleotide polymorphism	
Molecule Type:	Genomic	**RefSNP Alleles:**	C/T	
Created/Updated in build:	132/132	**Ancestral Allele:**	C	
Map to Genome Build:	37.1	**Clinical Association:**	unknown	

SNP Details are organized in the following sections:

GeneView | Map | Submission | Fasta | Resource | Diversity | Validation

Integrated Maps (Hint: click on 'Chr Pos' or 'Contig Pos' column value to see variation in NCBI sequence viewer)

Genome Build	Chr	Chr Pos	Contig	Contig Pos	SNP to Chr	Contig allele	Contig to Chr	Group term	Group label
37.1	11	5365554	NW_925006.1	865878	-	G	+	Celera	Celera
37.1	11	4906056	NW_001838021.1	875431	-	G	+	HuRef	HuRef
37.1	11	5246883	NT_009237.18	5186883	-	G	+	GRCh37	GRCh37

GeneView

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 Go

Assembly	SNP to Chr	Chr	Chr position	Contig	Contig position	Allele
GRCh37	-	11	5246883	NT_009237.18	5186883	G

RefSeqGene	Gene (ID)	SNP to RefSeqGene	Position	Allele
NG_000007.3	HBB (3043)	+	71963	C

Function	mRNA				Protein		
	SNP to mRNA	Accession	Position	Allele change	Accession	Position	Residue change
missense	+	NM_000518.4	439	GCC ➡ GTC	NP_000509.1	130	A [Ala] ➡ V [Val]

β-Hemoglobin Gene SNP rs111645889

http://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?rs=111645889

Submitter records for this RefSNP Cluster

The submission **ss212960479** has the longest flanking sequence of all cluster members and was used to instantiate sequence for **rs111645889** during BLAST analysis for the current

NCBI Assay ID	Handle|Submitter ID	Validation Status	ss to rs Orientation /Strand	Alleles	5' Near Seq 30 bp	3' Near Seq 30 bp	Entry Date	Update Date	Build Added
ss212960479	RSG_UW|HBB-3419		fwd/B	C/T	gcaaagaattcaccccaccagtgcaggctg	ctatcagaaagtggtggctggtgtggctaa	04/02/10	04/02/10	132

Fasta sequence (Legend)

>gnl|dbSNP|rs111645889|allelePos=256|totalLen=511|taxid=9606|snpclass=1|alleles='C/T'|mol=Genomic|build=132

```
TGCATATAAA TATTTCTGCA TATAAATTGT AACTGATGTA AGAGGTTTCA TATTGCTAAT
AGCAGCTACA ATCCAGCTAC CATTCTGCTT TTATTTTATG GTTGGGATAA GGCTGGATTA
TTCTGAGTCC AAGCTAGGCC CTTTTGCTAA TCATGTTCAT ACCTCTTATC TTCCTCCCAC
AGCTCCTGGG CAACGTGCTG GTCTGTGTGC TGGCCCATCA CTTTGGCAAA GAATTCACCC
CACCAGTGCA GGCTG
Y
CTATCAGAAA GTGGTGGCTG GTGTGGCTAA TGCCCTGGCC CACAAGTATC ACTAAGCTCG
CTTTCTTGCT GTCCAATTTC TATTAAAGGT TCCTTTGTTC CCTAAGTCCA ACTACTAAAC
TGGGGGATAT TATGAAGGGC CTTGAGCATC TGGATTCTGC CTAATAAAAA ACATTTATTT
TCATTGCAAT GATGTATTTA AATTATTTCT GAATATTTTA CTAAAAAGGG AATGTGGGAG
GTCAGTGCAT TTAAA
```

NCBI Resource Links

Resource

Submitter-Referenced	dbSNP Blast Analysis	3D structure mapping
GenBank		NP_000509
GU324922		

Population Diversity

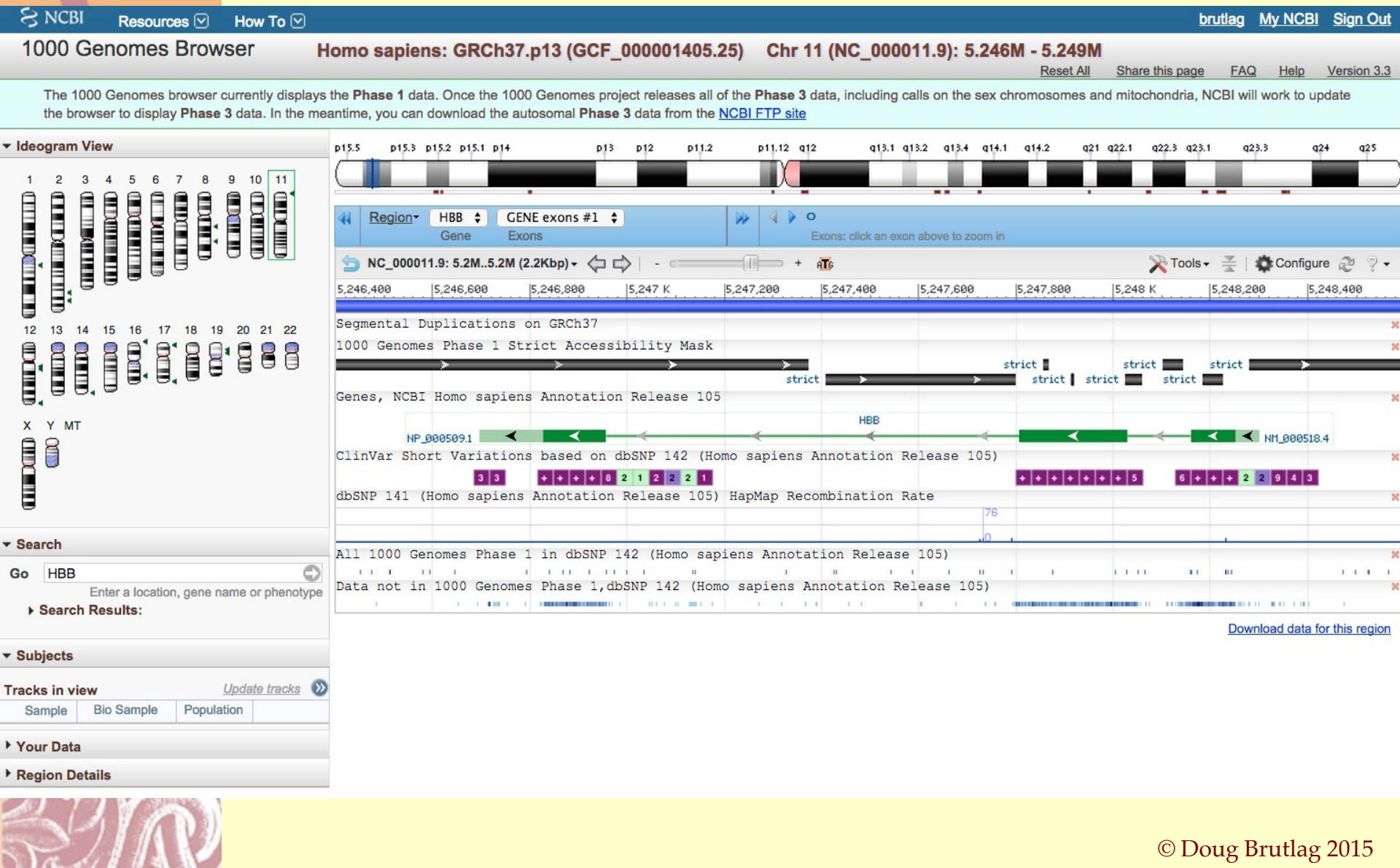
There is no frequency data.

Validation Summary:

Validation status	Marker displays Mendelian segregation	PCR results confirmed in multiple reactions	Homozygotes detected in individual genotype data
	UNKNOWN	UNKNOWN	YES

Thousand Genomes Browser

http://www.ncbi.nlm.nih.gov/variation/tools/1000genomes/



1000 Genomes
A Deep Catalog of Human Genetic Variation

Thousand Genomes Browser

http://www.ncbi.nlm.nih.gov/variation/tools/1000genomes/

Genotypes

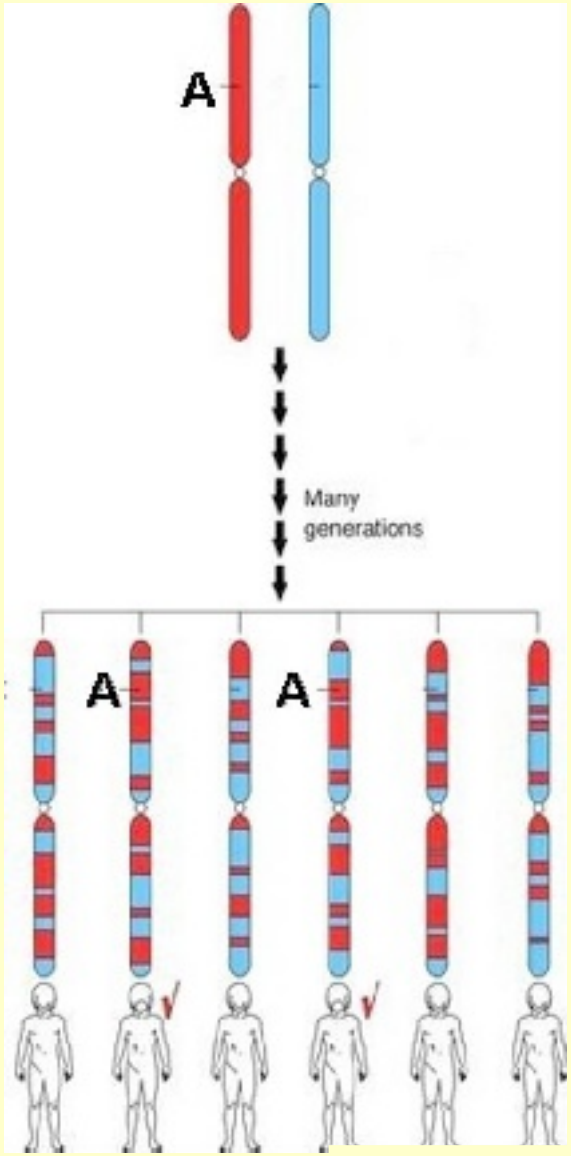
Drag ruler or use the arrow buttons to scroll the visible range.
Click or Shift-click the ruler to select a column. Alt-click or Shift-Alt-click to show on sequence.

Hide populations with unchecked samples

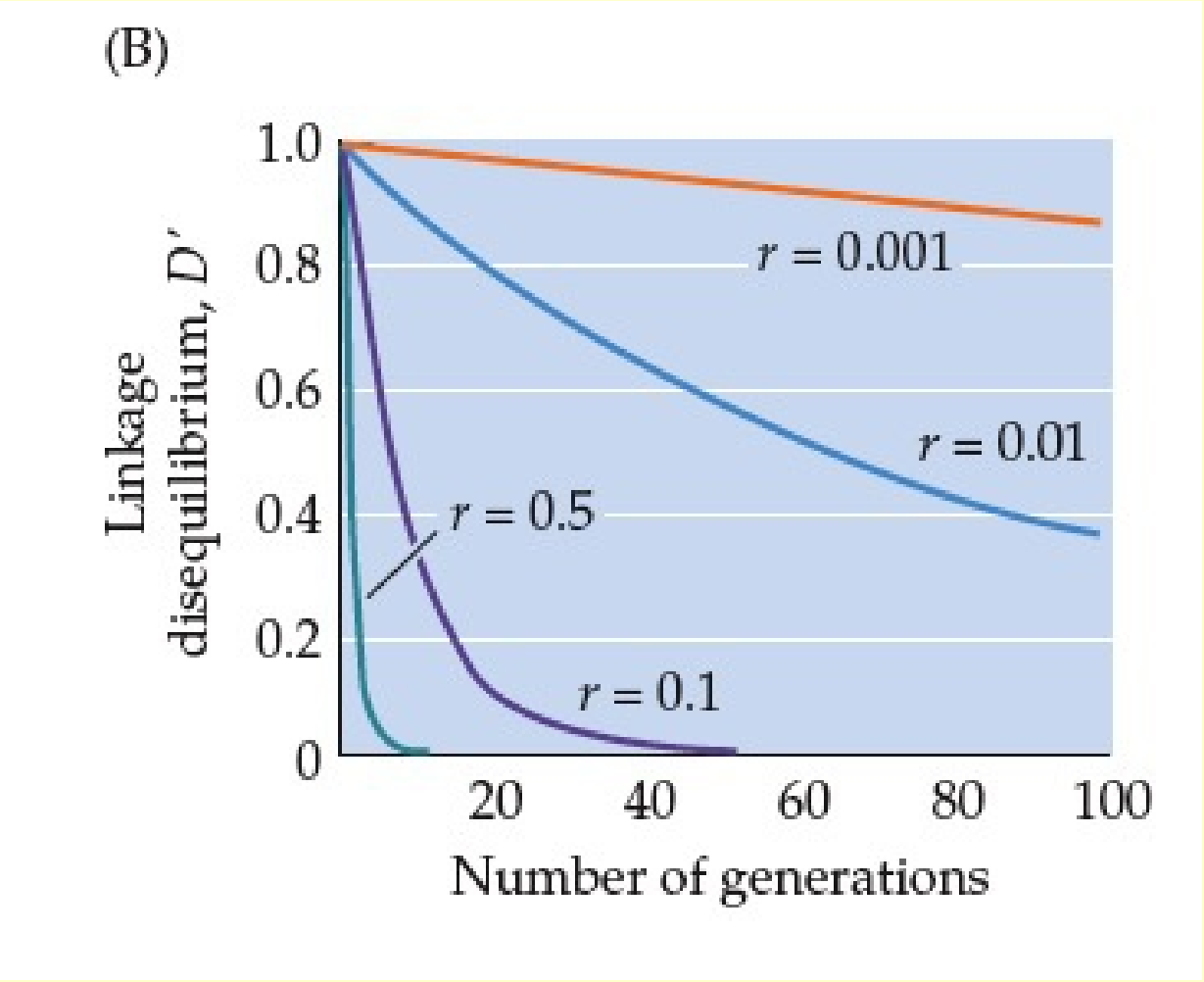
Go to Selection | Scroll Region

Populations / Samples Show: Allele frequencies	5,246,840 s36020563	5,246,870 rs113082294	5,246,883 .	5,246,923 rs34049764	5,246,958 .	5,246,975 rs191535077	5,247,001 rs140033163	5,247,035 rs181743523	5,247,135 rs185607297	5,247,141 rs1609812	5,247,329 rs78815705	5,247,427 rs190369729	5,247,430 rs182729393	5,247,543 rs187507944	5,247 rs1131
	G=0.9995	**C=0.9977**	**G=0.9995**	**G=1.0000**	**T=0.9995**	**A=0.9991**	**C=0.9954**	**C=0.9995**	**T=0.9982**	G=0.2594	**G=0.9876**	**A=0.9995**	**G=0.9995**	**G=0.9995**	**G=0.**
	A=0.0005	G=0.0023	A=0.0005	A=0.0000	C=0.0005	T=0.0009	G=0.0046	T=0.0005	G=0.0018	**A=0.7406**	T=0.0124	G=0.0005	A=0.0005	C=0.0005	A=0.0
▸ **ASW** Americans of African…	**G=1.0000**	**C=1.0000**	**G=1.0000**	**G=1.0000**	**T=0.9918**	**A=1.0000**	**C=1.0000**	**C=1.0000**	**T=1.0000**	G=0.1967	**G=1.0000**	**A=1.0000**	**G=1.0000**	**G=1.0000**	**G=1.**
	A=0.0000	G=0.0000	A=0.0000	A=0.0000	C=0.0082	T=0.0000	G=0.0000	T=0.0000	G=0.0000	**A=0.8033**	T=0.0000	G=0.0000	A=0.0000	C=0.0000	A=0.0
▸ **CEU** Utah Residents (CEPH…	**G=1.0000**	**C=0.9882**	**G=1.0000**	**G=1.0000**	**T=1.0000**	**A=1.0000**	**C=1.0000**	**C=1.0000**	**T=1.0000**	G=0.1412	**G=0.9647**	**A=1.0000**	**G=1.0000**	**G=1.0000**	**G=1.**
	A=0.0000	G=0.0118	A=0.0000	A=0.0000	C=0.0000	T=0.0000	G=0.0000	T=0.0000	G=0.0000	**A=0.8588**	T=0.0353	G=0.0000	A=0.0000	C=0.0000	A=0.0
▸ **CHB** Han Chinese in Bejin…	**G=1.0000**	**C=1.0000**	**G=1.0000**	**G=1.0000**	**T=1.0000**	**A=1.0000**	**C=0.9948**	**C=1.0000**	**T=0.9948**	G=0.4794	**G=1.0000**	**A=1.0000**	**G=1.0000**	**G=1.0000**	**G=1.**
	A=0.0000	G=0.0000	A=0.0000	A=0.0000	C=0.0000	T=0.0000	G=0.0052	T=0.0000	G=0.0052	**A=0.5206**	T=0.0000	G=0.0000	A=0.0000	C=0.0000	A=0.0
▸ **CHS** Southern Han Chinese	**G=1.0000**	**C=1.0000**	**G=1.0000**	**G=1.0000**	**T=1.0000**	**A=1.0000**	**C=0.9900**	**C=0.9950**	**T=0.9950**	**G=0.5300**	**G=1.0000**	**A=1.0000**	**G=1.0000**	**G=1.0000**	**G=1.**
	A=0.0000	G=0.0000	A=0.0000	A=0.0000	C=0.0000	T=0.0000	G=0.0100	T=0.0050	G=0.0050	A=0.4700	T=0.0000	G=0.0000	A=0.0000	C=0.0000	A=0.0
▸ **CLM** Colombians from Mede…	**G=1.0000**	**C=1.0000**	**G=1.0000**	**G=1.0000**	**T=1.0000**	**A=1.0000**	**C=1.0000**	**C=1.0000**	**T=1.0000**	G=0.2083	**G=0.9667**	**A=1.0000**	**G=1.0000**	**G=1.0000**	**G=1.**
	A=0.0000	G=0.0000	A=0.0000	A=0.0000	C=0.0000	T=0.0000	G=0.0000	T=0.0000	G=0.0000	**A=0.7917**	T=0.0333	G=0.0000	A=0.0000	C=0.0000	A=0.0
▸ **FIN** Finnish in Finland	**G=1.0000**	**C=1.0000**	**G=1.0000**	**G=1.0000**	**T=1.0000**	**A=1.0000**	**C=1.0000**	**C=1.0000**	**T=1.0000**	G=0.2366	**G=0.9624**	**A=0.9946**	**G=1.0000**	**G=1.0000**	**G=0.**
	A=0.0000	G=0.0000	A=0.0000	A=0.0000	C=0.0000	T=0.0000	G=0.0000	T=0.0000	G=0.0000	**A=0.7634**	T=0.0376	G=0.0054	A=0.0000	C=0.0000	A=0.0
▸ **GBR** British in England a…	**G=1.0000**	**C=0.9944**	**G=1.0000**	**G=1.0000**	**T=1.0000**	**A=0.9888**	**C=1.0000**	**C=1.0000**	**T=1.0000**	G=0.1854	**G=0.9831**	**A=1.0000**	**G=1.0000**	**G=1.0000**	**G=1.**
	A=0.0000	G=0.0056	A=0.0000	A=0.0000	C=0.0000	T=0.0112	G=0.0000	T=0.0000	G=0.0000	**A=0.8146**	T=0.0169	G=0.0000	A=0.0000	C=0.0000	A=0.0
▸ **IBS** Iberian population i…	**G=1.0000**	**C=0.9643**	**G=1.0000**	**G=1.0000**	**T=1.0000**	**A=1.0000**	**C=1.0000**	**C=1.0000**	**T=1.0000**	G=0.1786	**G=1.0000**	**A=1.0000**	**G=1.0000**	**G=1.0000**	**G=1.**
	A=0.0000	G=0.0357	A=0.0000	A=0.0000	C=0.0000	T=0.0000	G=0.0000	T=0.0000	G=0.0000	**A=0.8214**	T=0.0000	G=0.0000	A=0.0000	C=0.0000	A=0.0
▸ **JPT** Japanese in Tokyo, J…	**G=1.0000**	**C=1.0000**	**G=1.0000**	**G=1.0000**	**T=1.0000**	**A=1.0000**	**C=0.9607**	**C=1.0000**	**T=0.9888**	**G=0.5112**	**G=1.0000**	**A=1.0000**	**G=1.0000**	**G=1.0000**	**G=1.**
	A=0.0000	G=0.0000	A=0.0000	A=0.0000	C=0.0000	T=0.0000	G=0.0393	T=0.0000	G=0.0112	A=0.4888	T=0.0000	G=0.0000	A=0.0000	C=0.0000	A=0.0
▸ **LWK** Luhya in Webuye, Ken…	**G=0.9948**	**C=1.0000**	**G=1.0000**	**G=1.0000**	**T=1.0000**	**A=1.0000**	**C=1.0000**	**C=1.0000**	**T=1.0000**	G=0.0625	**G=1.0000**	**A=1.0000**	**G=1.0000**	**G=1.0000**	**G=0.**
	A=0.0052	G=0.0000	A=0.0000	A=0.0000	C=0.0000	T=0.0000	G=0.0000	T=0.0000	G=0.0000	**A=0.9375**	T=0.0000	G=0.0000	A=0.0000	C=0.0000	A=0.0
▸ **MXL** Mexican Ancestry fro…	**G=1.0000**	**C=1.0000**	**G=1.0000**	**G=1.0000**	**T=1.0000**	**A=1.0000**	**C=1.0000**	**C=1.0000**	**T=1.0000**	G=0.3047	**G=0.9688**	**A=1.0000**	**G=0.9922**	**G=1.0000**	**G=1.**
	A=0.0000	G=0.0000	A=0.0000	A=0.0000	C=0.0000	T=0.0000	G=0.0000	T=0.0000	G=0.0000	**A=0.6953**	T=0.0313	G=0.0000	A=0.0078	C=0.0000	A=0.0
▸ **PUR** Puerto Ricans from P…	**G=1.0000**	**C=0.9909**	**G=1.0000**	**G=1.0000**	**T=1.0000**	**A=1.0000**	**C=1.0000**	**C=1.0000**	**T=1.0000**	G=0.2091	**G=0.9909**	**A=1.0000**	**G=1.0000**	**G=0.9909**	**G=1.**
	A=0.0000	G=0.0091	A=0.0000	A=0.0000	C=0.0000	T=0.0000	G=0.0000	T=0.0000	G=0.0000	**A=0.7909**	T=0.0091	G=0.0000	A=0.0000	C=0.0091	A=0.0
▸ **TSI** Toscani in Italia	**G=1.0000**	**C=1.0000**	**G=1.0000**	**G=1.0000**	**T=1.0000**	**A=1.0000**	**C=1.0000**	**C=1.0000**	**T=1.0000**	G=0.1378	**G=0.9898**	**A=1.0000**	**G=1.0000**	**G=1.0000**	**G=1.**
	A=0.0000	G=0.0000	A=0.0000	A=0.0000	C=0.0000	T=0.0000	G=0.0000	T=0.0000	G=0.0000	**A=0.8622**	T=0.0102	G=0.0000	A=0.0000	C=0.0000	A=0.0
▸ **YRI** Yoruba in Ibadan, Ni…	**G=1.0000**	**C=1.0000**	**G=0.9943**	**G=1.0000**	**T=1.0000**	**A=1.0000**	**C=1.0000**	**C=1.0000**	**T=1.0000**	G=0.1080	**G=1.0000**	**A=1.0000**	**G=1.0000**	**G=1.0000**	**G=0.**
	A=0.0000	G=0.0000	A=0.0057	A=0.0000	C=0.0000	T=0.0000	G=0.0000	T=0.0000	G=0.0000	**A=0.8920**	T=0.0000	G=0.0000	A=0.0000	C=0.0000	A=0.0

Origin of Haplotypes



Linkage Disequilibrium and Recombination Rate



Linkage Disequilibrium (LD) Across the Human LPL Gene

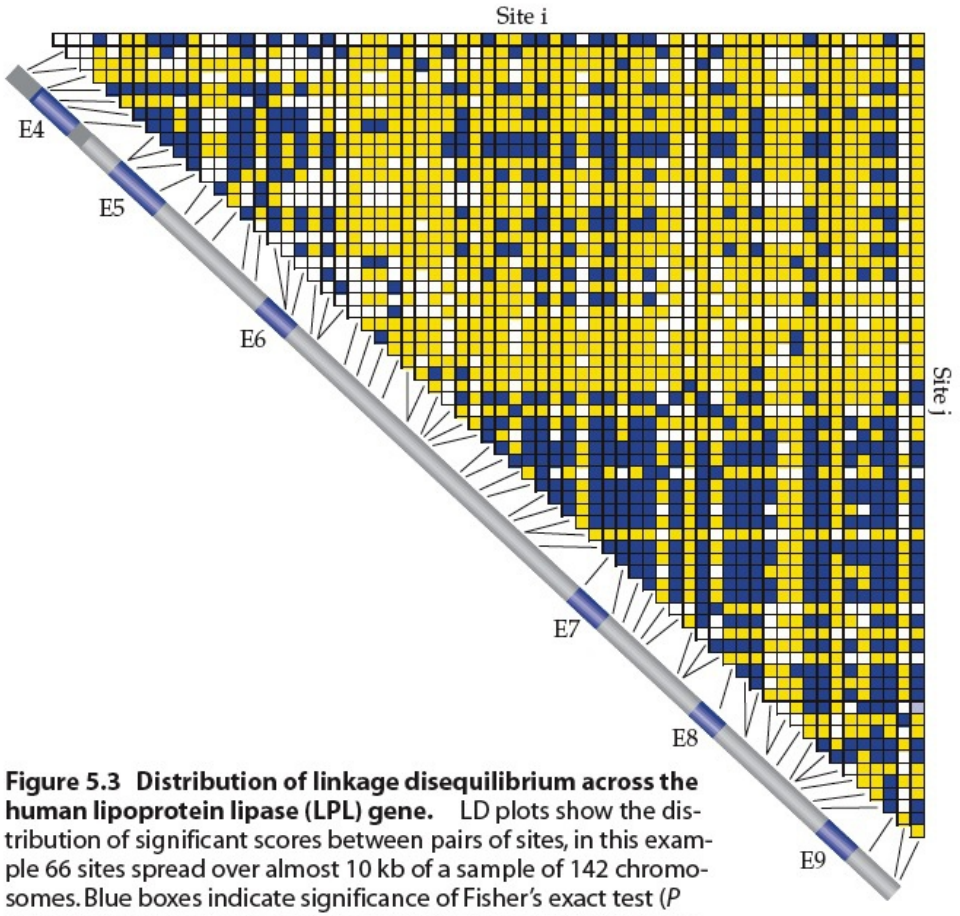
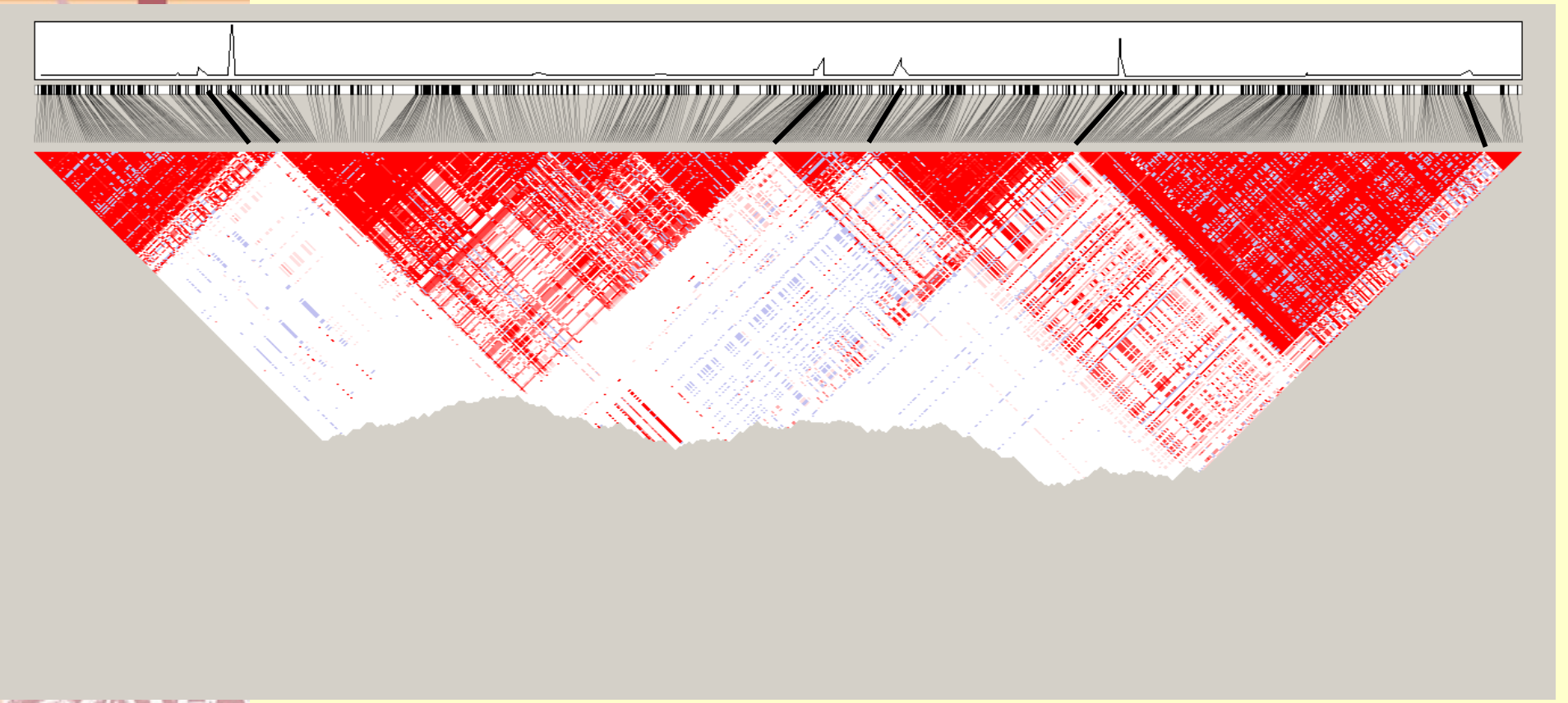


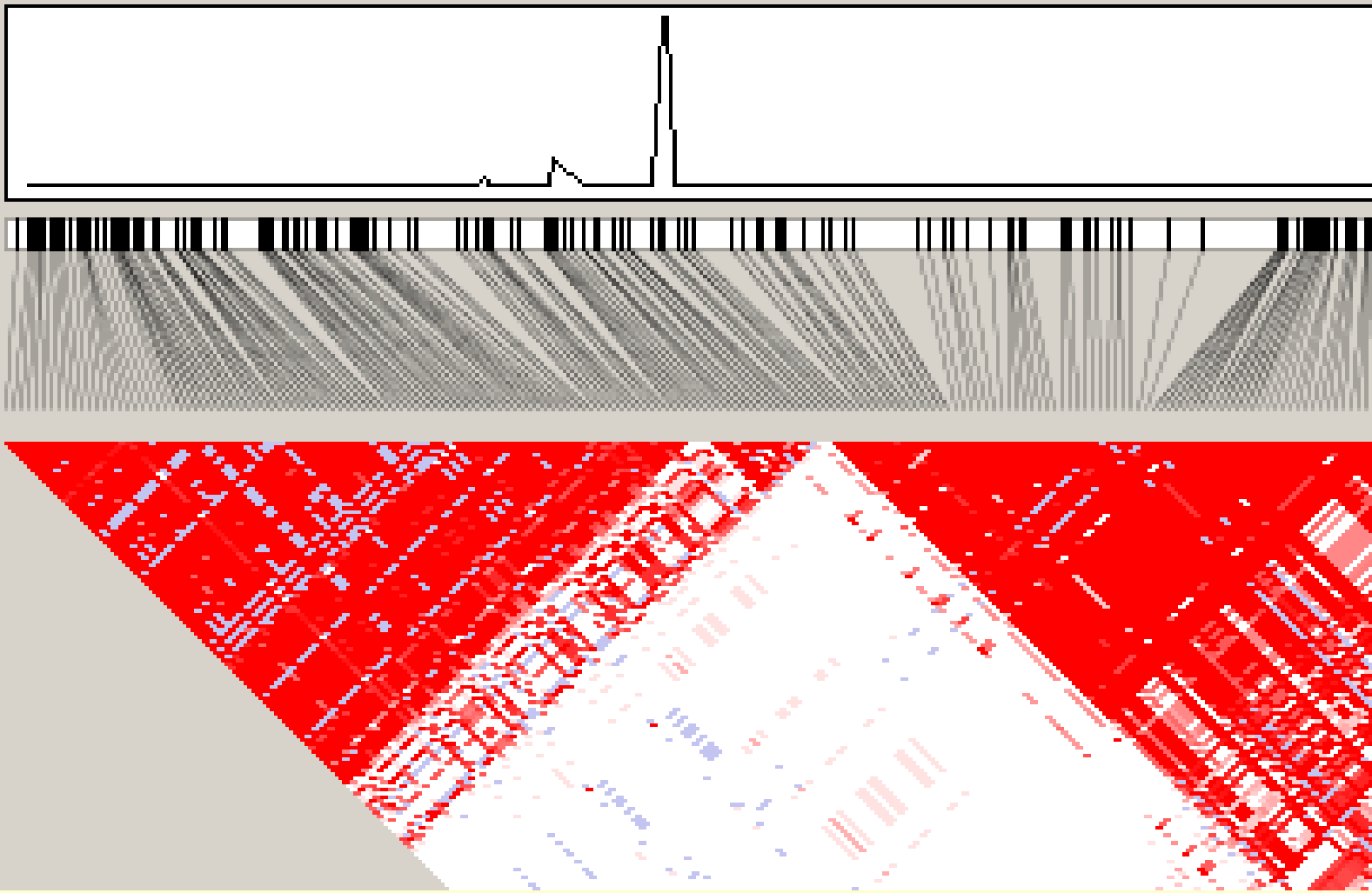
Figure 5.3 Distribution of linkage disequilibrium across the human lipoprotein lipase (LPL) gene. LD plots show the distribution of significant scores between pairs of sites, in this example 66 sites spread over almost 10 kb of a sample of 142 chromosomes. Blue boxes indicate significance of Fisher's 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 linkage disequilibrium structure



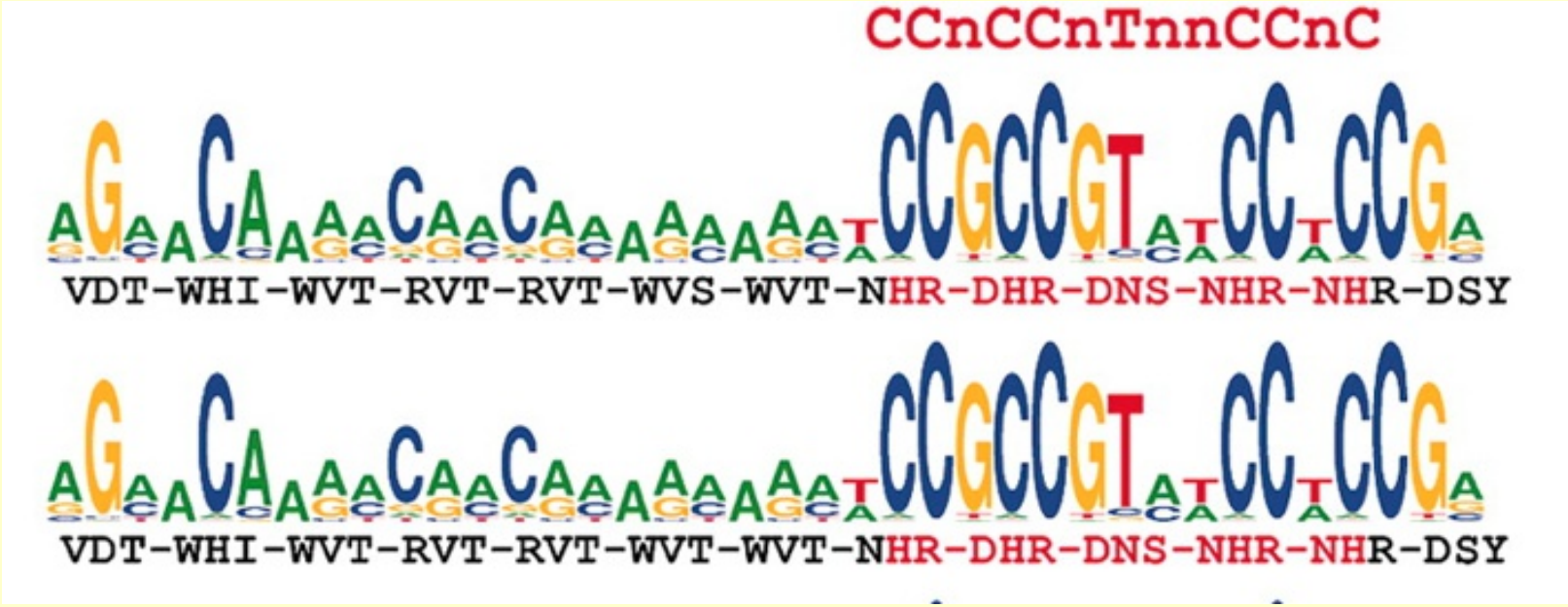
7q21

Recombination hotspots are widespread and account for linkage disequilibrium structure



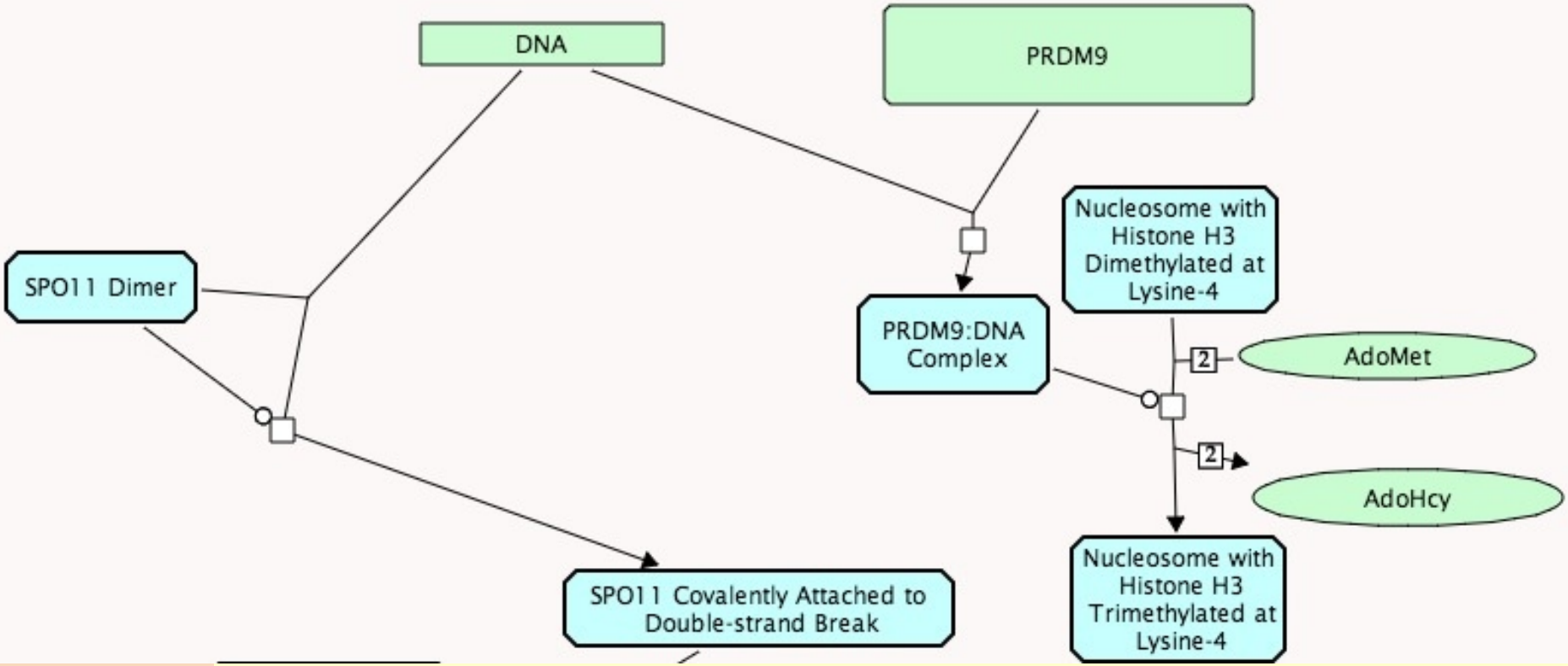
7q21

Consensus binding site for PRDM9



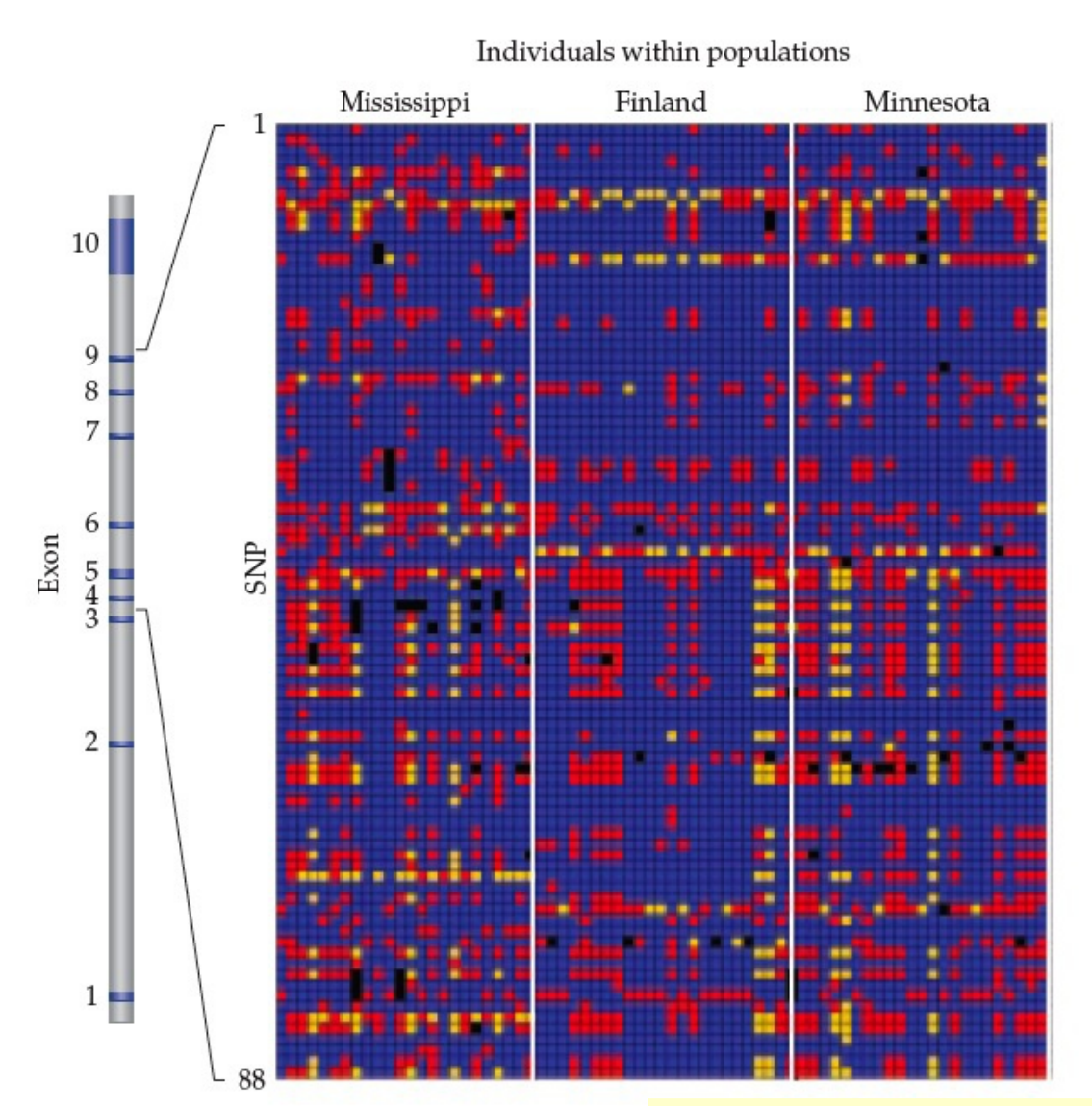
Baudet et al, PRDM9 and Meiotic Recombination Science 237. 836.

Initiation of Meiotic Recombination by PRDM9



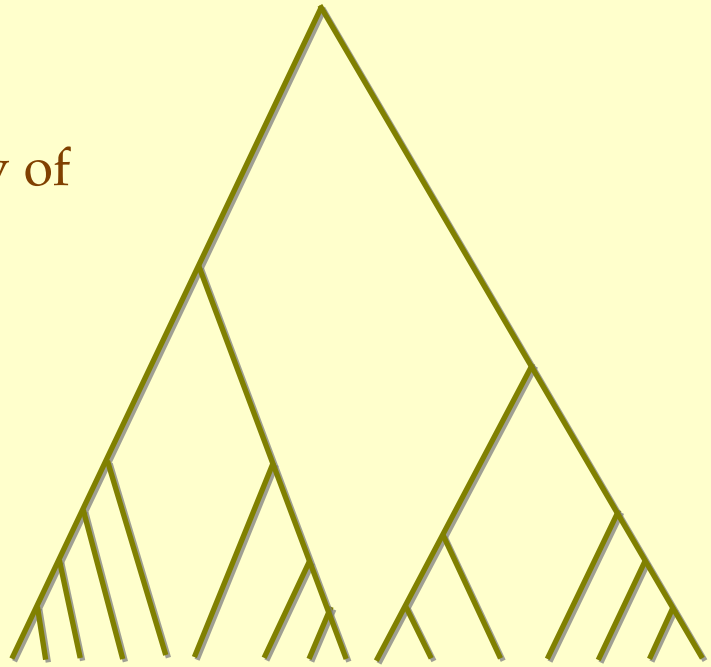
Observation of Haplotypes

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



A Haplotype Map of the Human Genome

http://www.nature.com/nature/journal/v437/n7063/full/nature04226.html

Francis Collins, 2008

National Human Genome Research Institute

http://www.genome.gov/

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

Google™ Search SEARCH

Research Funding | Research at NHGRI | Health | Education | Issues in Genetics | Newsroom | Careers & Training | About | For You

Highlights RSS

International gathering highlights opportunities in genomic medicine

At the Global Leaders in Genomic Medicine meeting, held Jan. 8-9 at the National Academy of Sciences in Washington, D.C., attendees learned how both large and small countries are creating innovative programs and plans for implementing genomic medicine. The two-day meeting was sponsored by NHGRI. **Read more**

Genomics in Medicine: *Genome and Transcriptome Dynamics in Cancer Cells*

Join us Friday, February 7th at 8:00 a.m. for our next Genomics in Medicine lecture, part of the 2013-2014 series, featuring Thomas Ried, M.D., senior investigator and chief of the Cancer Genomics Section, National Cancer Institute, NIH. His talk, *Genome and Transcriptome Dynamics in Cancer Cells*, will focus on malignant cells that carry two specific and concurrent alterations of the cellular transcriptome. **Read more**

Apply for NHGRI-ASHG's new education fellowship for genetics professionals

To help cultivate an educated citizenry, the American Society of Human Genetics (ASHG) and NHGRI have teamed up to sponsor the new Genetics and Education Fellowship. Every year, one genetics professional will receive comprehensive training and experience to help prepare him or her for a career in genetics and genomics education. **Read more**

For You

- Students
- Educators
- Patients
- Health Professionals
- Grant Applicants
- News Media

Newsroom

TCGA bladder cancer study reveals potential drug targets, similarities to several cancers
January 29, 2014

NIH study links family structure to high blood pressure in African-American men
December 12, 2013

With new study, aquatic comb jelly floats into new evolutionary position
December 12, 2013

NIH deposits first batch of genomic data for Alzheimer's disease
December 2, 2013

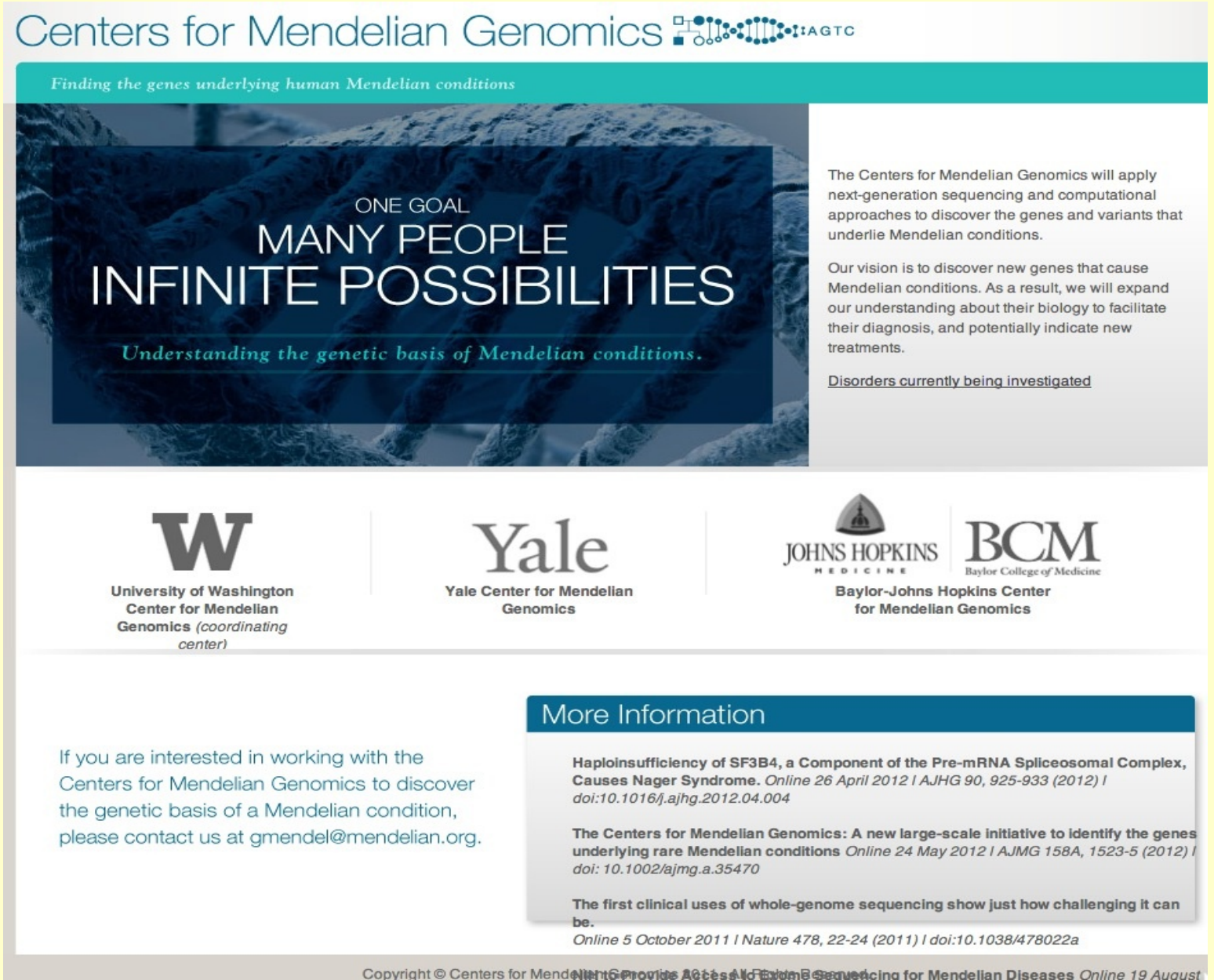
GenomeTV View All

Quick Links

Connect & Collaborate

Centers for Mendelian Genomics

http://mendelian.org/



Centers for Mendelian Genomics

http://mendelian.org/

INVITED COMMENT

AMERICAN JOURNAL OF medical genetics

The Centers for Mendelian Genomics: A New Large-Scale Initiative to Identify the Genes Underlying Rare Mendelian Conditions

Michael J. Bamshad,[1,2,3]* Jay A. Shendure,[2] David Valle,[4] Ada Hamosh,[4] James R. Lupski,[5,6,7,8] Richard A. Gibbs,[5,8] Eric Boerwinkle,[8,9] Richard P. Lifton,[10] Mark Gerstein,[11] Murat Gunel,[10,12] Shrikant Mane,[10] and Deborah A. Nickerson[2]

on behalf of the Centers for Mendelian Genomics

[1]Department of Pediatrics, University of Washington, Seattle, Washington

[2]Department of Genome Sciences, University of Washington, Seattle, Washington

[3]Seattle Children's Hospital, Seattle, Washington

[4]McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland

[5]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas

[6]Department of Pediatrics, Baylor College of Medicine, Houston, Texas

[7]Texas Children's Hospital, Houston, Texas

[8]Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas

[9]Human Genetics Center, University of Texas Health Sciences Center at Houston, Houston, Texas

[10]Department of Genetics, Yale University School of Medicine, New Haven, Connecticut

[11]Department of Biophysics and Biochemistry, Yale University School of Medicine, New Haven, Connecticut

[12]Department of Neurosurgery, Yale University School of Medicine, New Haven, Connecticut

Manuscript Received: 2 April 2012; Manuscript Accepted: 19 April 2012

Portrait of a Glitch

- Revere La Noue, MFA, Stanford, 2005
- What is this film about?
- What classes of glitches are mentioned?
- What do these glitches cause?
- Why did I show this film?