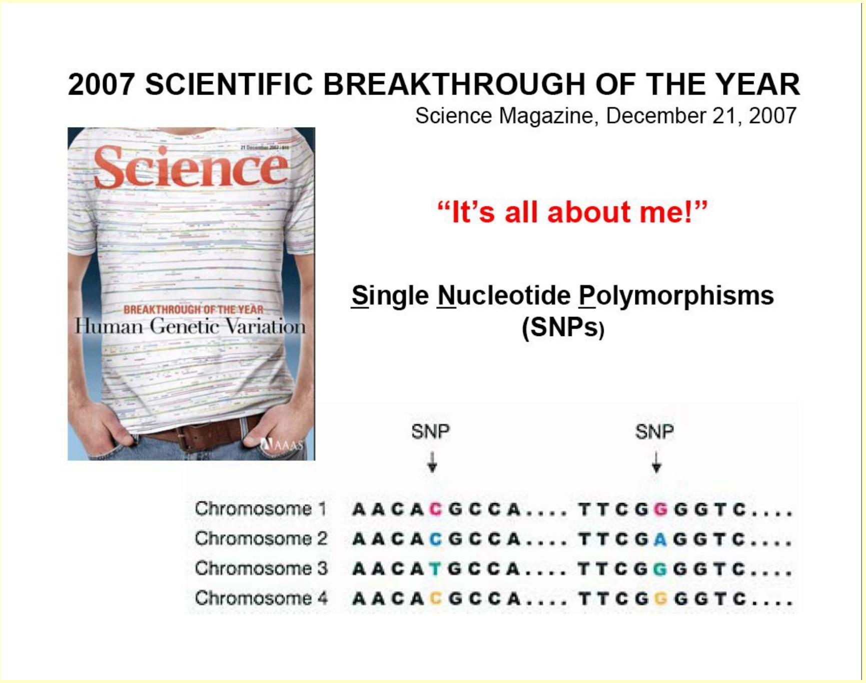
Your Genes and Your Health

http://bio84.stanford.edu/

Stanford Continuing Studies

http://continuingstudies.stanford.edu/

Single Nucleotide Polymorphisms (SNPs)



2007 SCIENTIFIC BREAKTHROUGH OF THE YEAR

Science Magazine, December 21, 2007

“It's all about me!”

Single Nucleotide Polymorphisms (SNPs)

Chromosome 1 AACACGCCA.... TTCGGGGTC....
Chromosome 2 AACACGCCA.... TTCGAGGTC....
Chromosome 3 AACATGCCA.... TTCGGGGTC....
Chromosome 4 AACACGCCA.... TTCGGGGTC....

Doug Brutlag, Professor Emeritus of
Biochemistry & Medicine (by courtesy)
Stanford University School of Medicine

NIH Precision Medicine Initiative

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

National Institutes of Health
Turning Discovery Into Health

Search NIH

NIH Employee Intranet | Staff Directory | En Español

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

Home » Research & Training

PRECISION MEDICINE INITIATIVE COHORT PROGRAM

Precision Medicine Initiative

- Scale and Scope
- Participation
- Funding Opportunities
- FAQ
- Advisory Groups
- Events
- Announcements
- PMI in the News
- Multimedia

Read blogs on precision medicine by the NIH Director.

Stay up to date on funding opportunities.

About the Precision Medicine Initiative Cohort Program

Far too many diseases do not have a proven means of prevention or effective treatments. We must gain better insights into the biological, environmental, and behavioral influences on 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

THE PRECISION MEDICINE INITIATIVE®

WHAT IS IT?

Precision medicine is an emerging approach for disease prevention and treatment that takes into account people's individual variations in genes, environment, and lifestyle.

The Precision Medicine Initiative® will generate the scientific evidence needed to **move the concept of precision medicine into clinical practice.**

Email Updates

Sign up to receive email updates about the Precision Medicine Initiative.

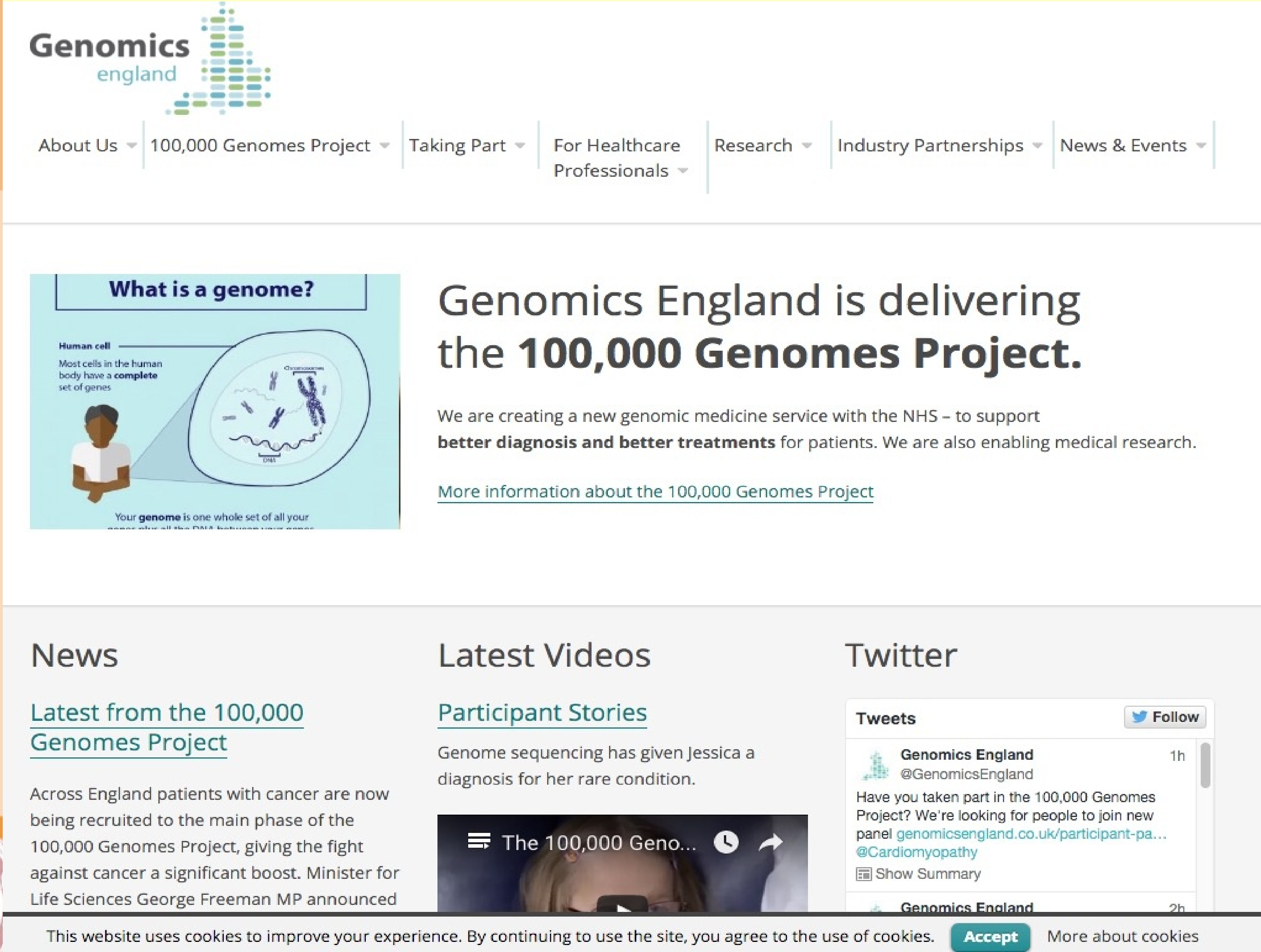
Sign up for updates

Related Links

- PMI Working Group Final Report pdf
- 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
- Precision Medicine Initiative YouTube Channel

Genomics England

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



Genomics England is delivering the **100,000 Genomes Project.**

We are creating a new genomic medicine service with the NHS – to support **better diagnosis and better treatments** for patients. We are also enabling medical research.

More information about the 100,000 Genomes Project

News

Latest from the 100,000 Genomes Project

Across England patients with cancer are now being recruited to the main phase of the 100,000 Genomes Project, giving the fight against cancer a significant boost. Minister for Life Sciences George Freeman MP announced

Latest Videos

Participant Stories

Genome sequencing has given Jessica a diagnosis for her rare condition.

Twitter

Genomics England @GenomicsEngland 1h

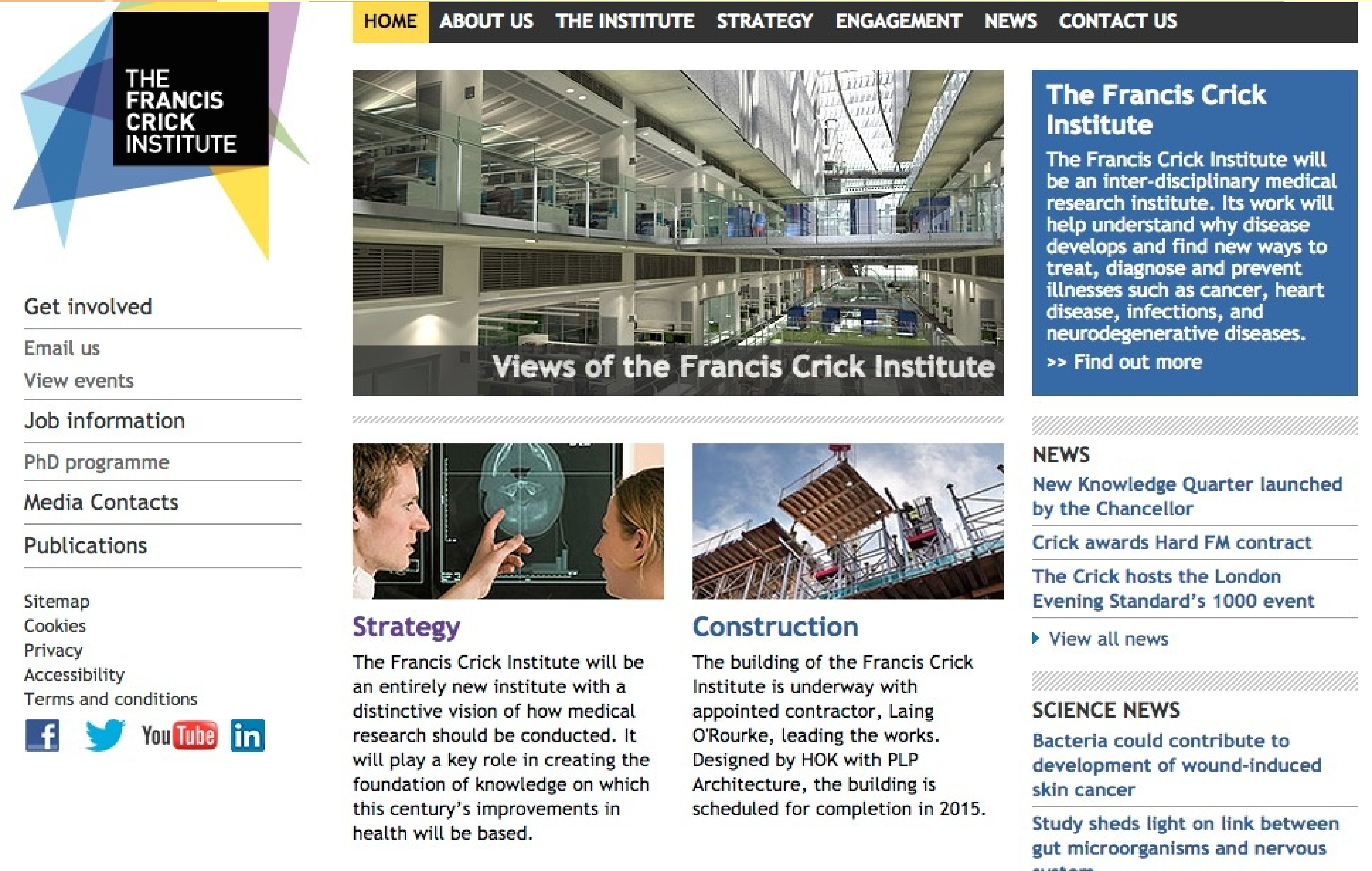
Have you taken part in the 100,000 Genomes Project? We're looking for people to join new panel genomicsengland.co.uk/participant-pa... @Cardiomyopathy

Genomics England 2h

This website uses cookies to improve your experience. By continuing to use the site, you agree to the use of cookies.

Francis Crick Institute

http://www.crick.ac.uk/



Views of the Francis Crick Institute

The Francis Crick Institute

The Francis Crick Institute will be an inter-disciplinary medical research institute. Its work will help understand why disease develops and find new ways to treat, diagnose and prevent illnesses such as cancer, heart disease, infections, and neurodegenerative diseases.

>> Find out more

- Get involved
- Email us
- View events
- Job information
- PhD programme
- Media Contacts
- Publications

Sitemap
Cookies
Privacy
Accessibility
Terms and conditions

Strategy

The Francis Crick Institute will be an entirely new institute with a distinctive vision of how medical research should be conducted. It will play a key role in creating the foundation of knowledge on which this century's improvements in health will be based.

Construction

The building of the Francis Crick Institute is underway with appointed contractor, Laing O'Rourke, leading the works. Designed by HOK with PLP Architecture, the building is scheduled for completion in 2015.

NEWS

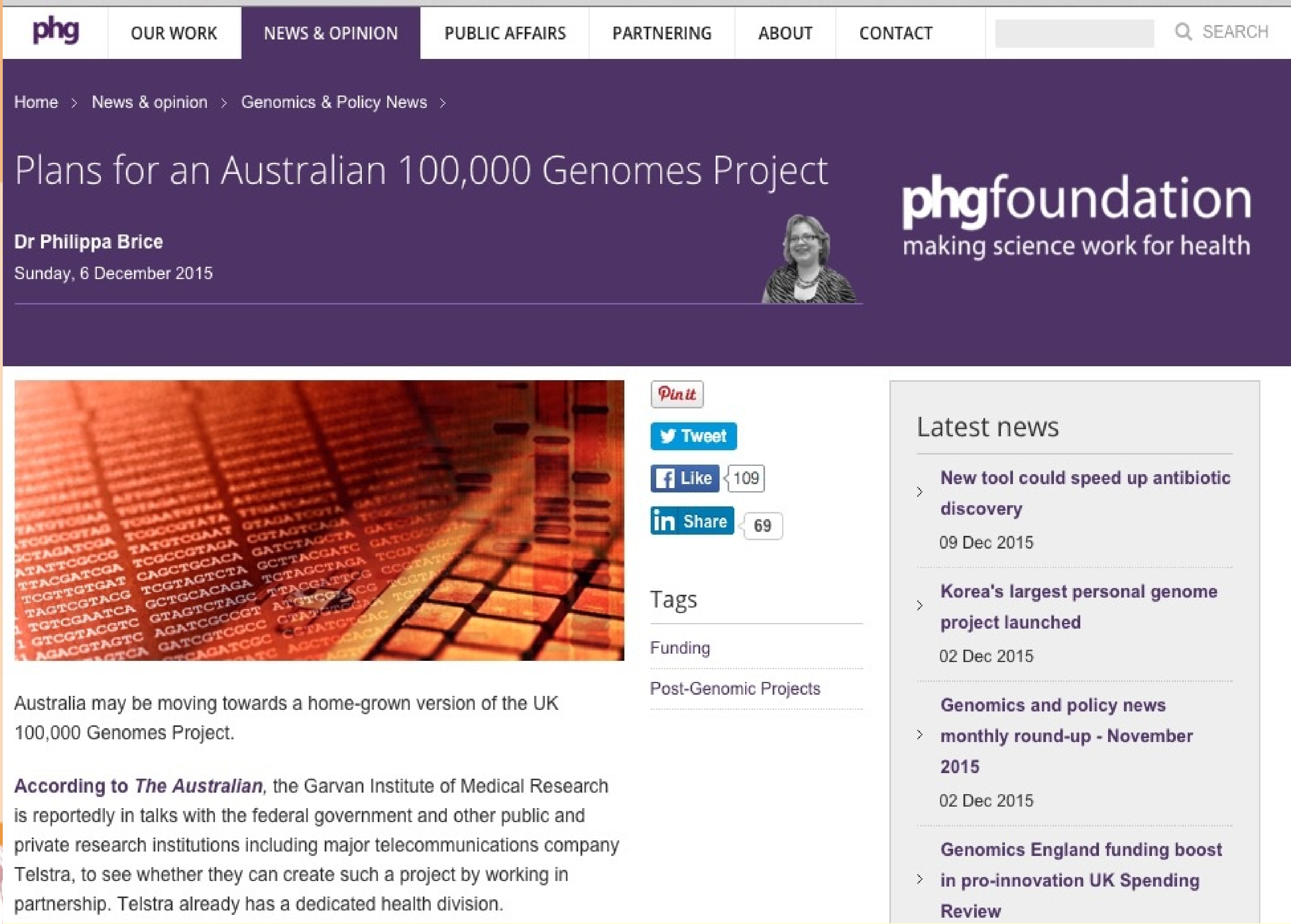
- New Knowledge Quarter launched by the Chancellor
- Crick awards Hard FM contract
- The Crick hosts the London Evening Standard's 1000 event
- View all news

SCIENCE NEWS

- Bacteria could contribute to development of wound-induced skin cancer
- Study sheds light on link between gut microorganisms and nervous

Australian 100,000 Genomes Project

http://www.phgfoundation.org/news/17075/



Plans for an Australian 100,000 Genomes Project

Dr Philippa Brice

Sunday, 6 December 2015

Australia may be moving towards a home-grown version of the UK 100,000 Genomes Project.

According to *The Australian*, the Garvan Institute of Medical Research is reportedly in talks with the federal government and other public and private research institutions including major telecommunications company Telstra, to see whether they can create such a project by working in partnership. Telstra already has a dedicated health division.

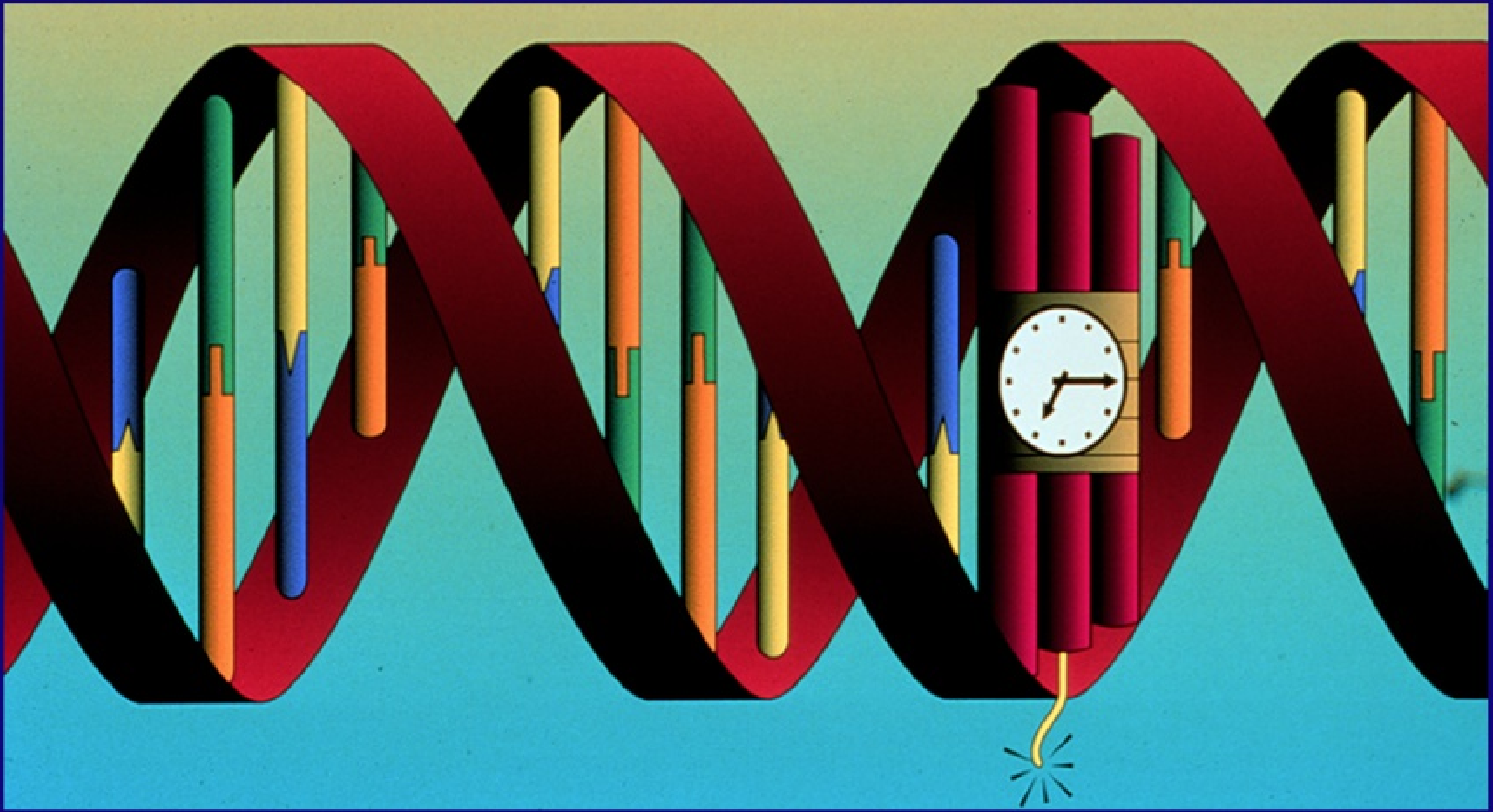
Tags

- Funding
- Post-Genomic Projects

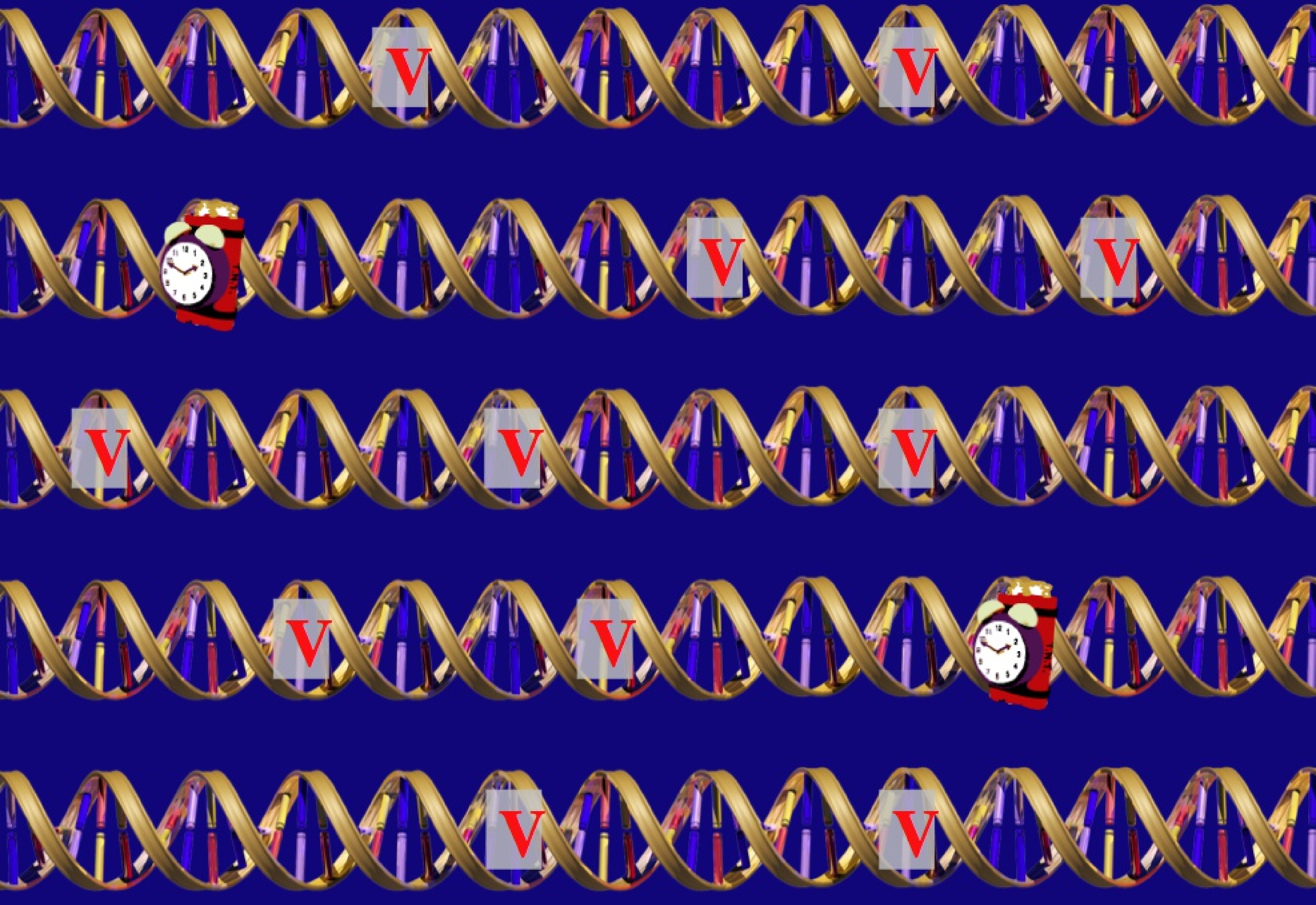
Latest news

- **New tool could speed up antibiotic discovery**
  09 Dec 2015
- **Korea's largest personal genome project launched**
  02 Dec 2015
- **Genomics and policy news monthly round-up - November 2015**
  02 Dec 2015
- **Genomics England funding boost in pro-innovation UK Spending Review**

All humans are ~99.9% identical at the DNA sequence level, and yet…

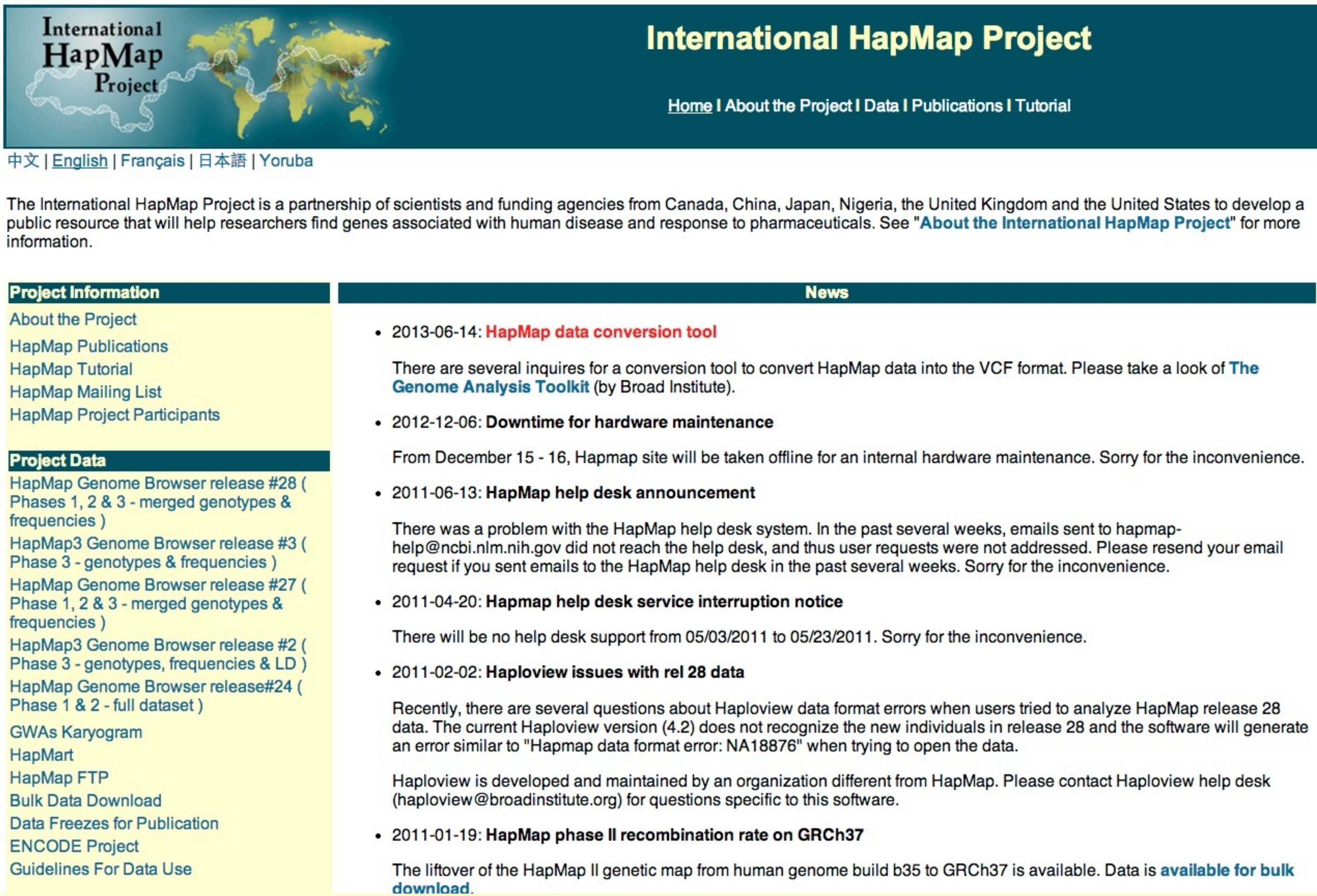


all of us carry a significant number of ‘glitches’ in our genomes.


V
V
V
V
V
V
V
V
V
V
V
V

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

Project Data

- HapMap Genome Browser release #28 (Phases 1, 2 & 3 - merged genotypes & frequencies)
- HapMap3 Genome Browser release #3 (Phase 3 - genotypes & frequencies)
- HapMap Genome Browser release #27 (Phase 1, 2 & 3 - merged genotypes & frequencies)
- HapMap3 Genome Browser release #2 (Phase 3 - genotypes, frequencies & LD)
- HapMap Genome Browser release#24 (Phase 1 & 2 - full dataset)
- GWAs Karyogram
- HapMart
- HapMap FTP
- Bulk Data Download
- Data Freezes for Publication
- ENCODE Project
- Guidelines For Data Use

News

- 2013-06-14: **HapMap data conversion tool**

  There are several inquires for a conversion tool to convert HapMap data into the VCF format. Please take a look of **The Genome Analysis Toolkit** (by Broad Institute).

- 2012-12-06: **Downtime for hardware maintenance**

  From December 15 - 16, Hapmap site will be taken offline for an internal hardware maintenance. Sorry for the inconvenience.

- 2011-06-13: **HapMap help desk announcement**

  There was a problem with the HapMap help desk system. In the past several weeks, emails sent to hapmap-help@ncbi.nlm.nih.gov did not reach the help desk, and thus user requests were not addressed. Please resend your email request if you sent emails to the HapMap help desk in the past several weeks. Sorry for the inconvenience.

- 2011-04-20: **Hapmap help desk service interruption notice**

  There will be no help desk support from 05/03/2011 to 05/23/2011. Sorry for the inconvenience.

- 2011-02-02: **Haploview issues with rel 28 data**

  Recently, there are several questions about Haploview data format errors when users tried to analyze HapMap release 28 data. The current Haploview version (4.2) does not recognize the new individuals in release 28 and the software will generate an error similar to "Hapmap data format error: NA18876" when trying to open the data.

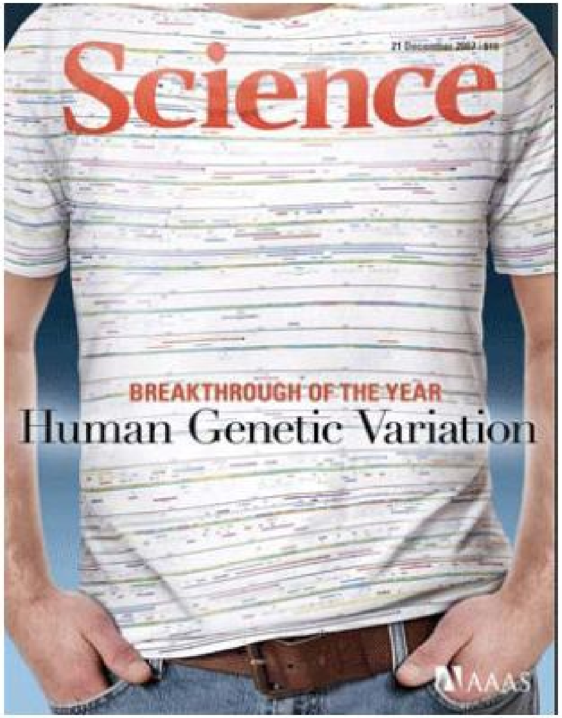
  Haploview is developed and maintained by an organization different from HapMap. Please contact Haploview help desk (haploview@broadinstitute.org) for questions specific to this software.

- 2011-01-19: **HapMap phase II recombination rate on GRCh37**

  The liftover of the HapMap II genetic map from human genome build b35 to GRCh37 is available. Data is **available for bulk download**.

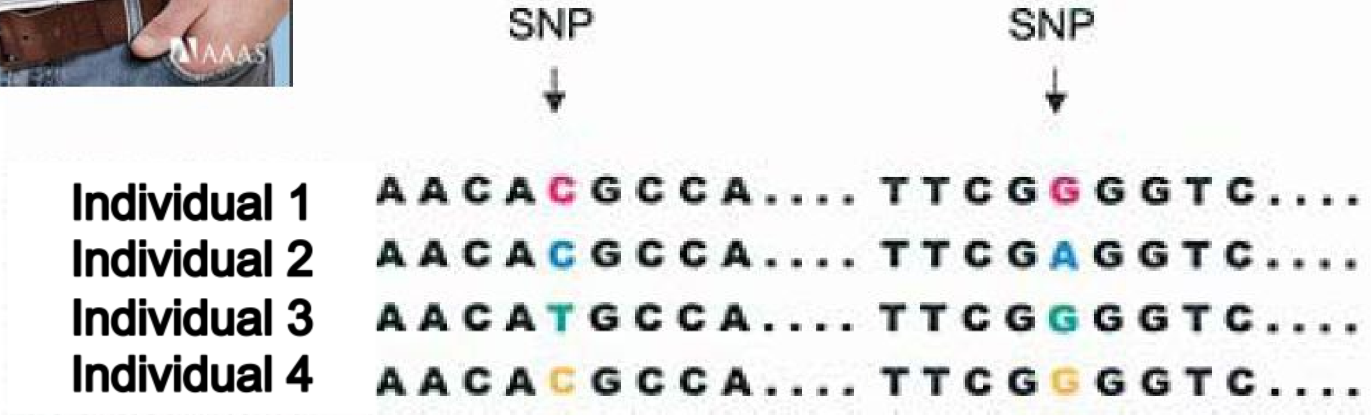
2007 SCIENTIFIC BREAKTHROUGH OF THE YEAR

Science Magazine, December 21, 2007



“It’s all about me!”

Single Nucleotide Polymorphisms (SNPs)

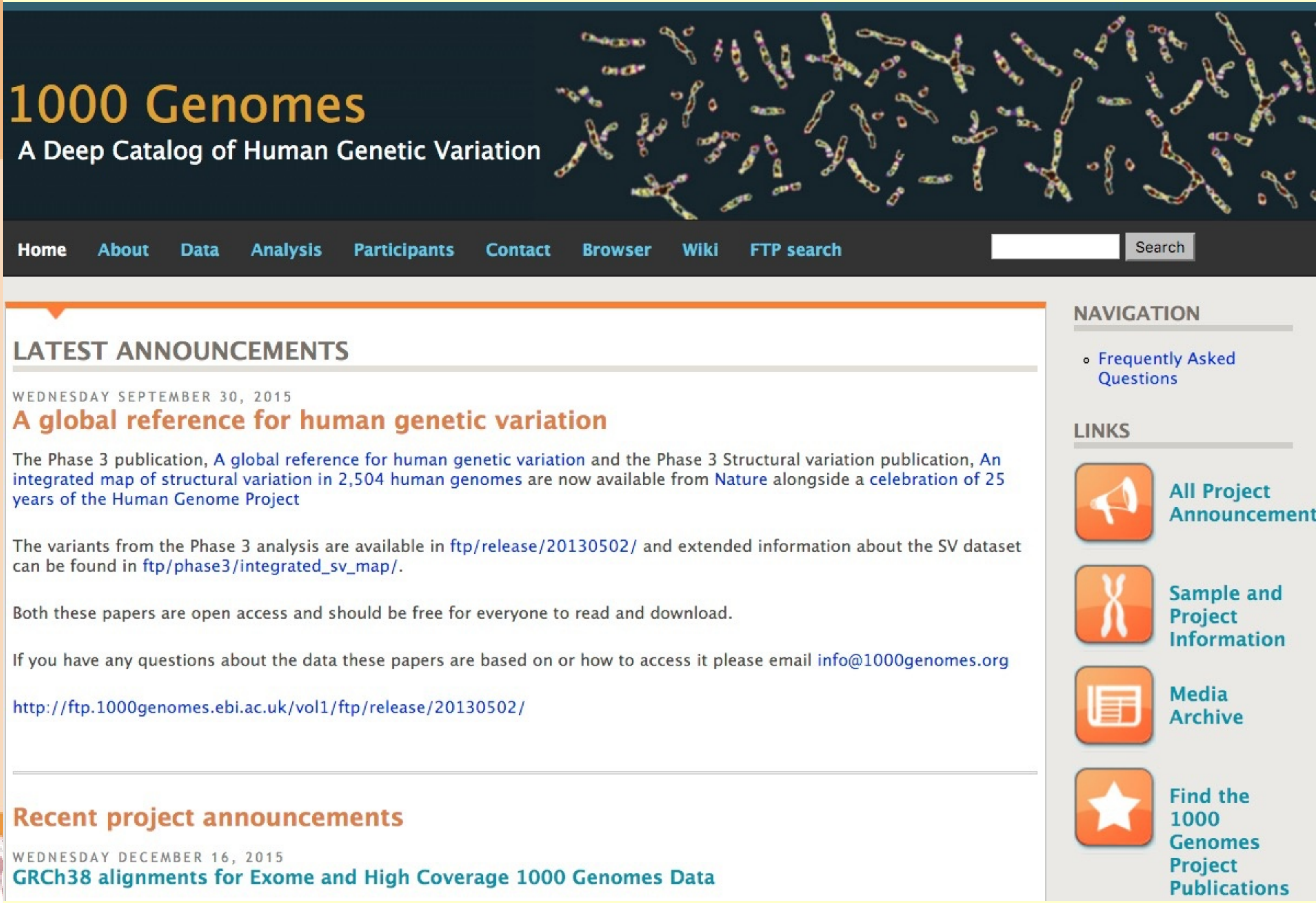


Individual 1	AACACGCCA.... TTCGGGGTC....
Individual 2	AACACGCCA.... TTCGAGGTC....
Individual 3	AACATGCCA.... TTCGGGGTC....
Individual 4	AACACGCCA.... TTCGGGGTC....

1000 Genomes
A Deep Catalog of Human Genetic Variation

Thousand Genomes Project

http://www.1000genomes.org/



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

WEDNESDAY DECEMBER 16, 2015

GRCh38 alignments for Exome and High Coverage 1000 Genomes Data

NAVIGATION

- Frequently Asked Questions

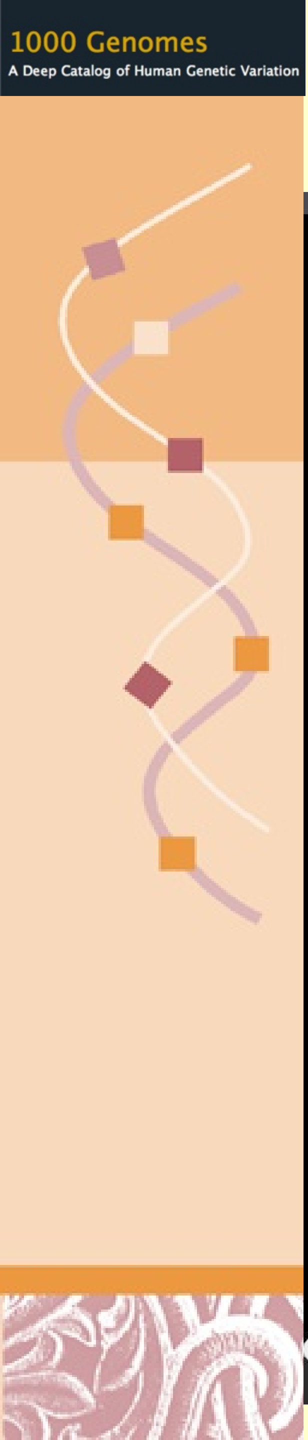
LINKS

All Project Announcement

Sample and Project Information

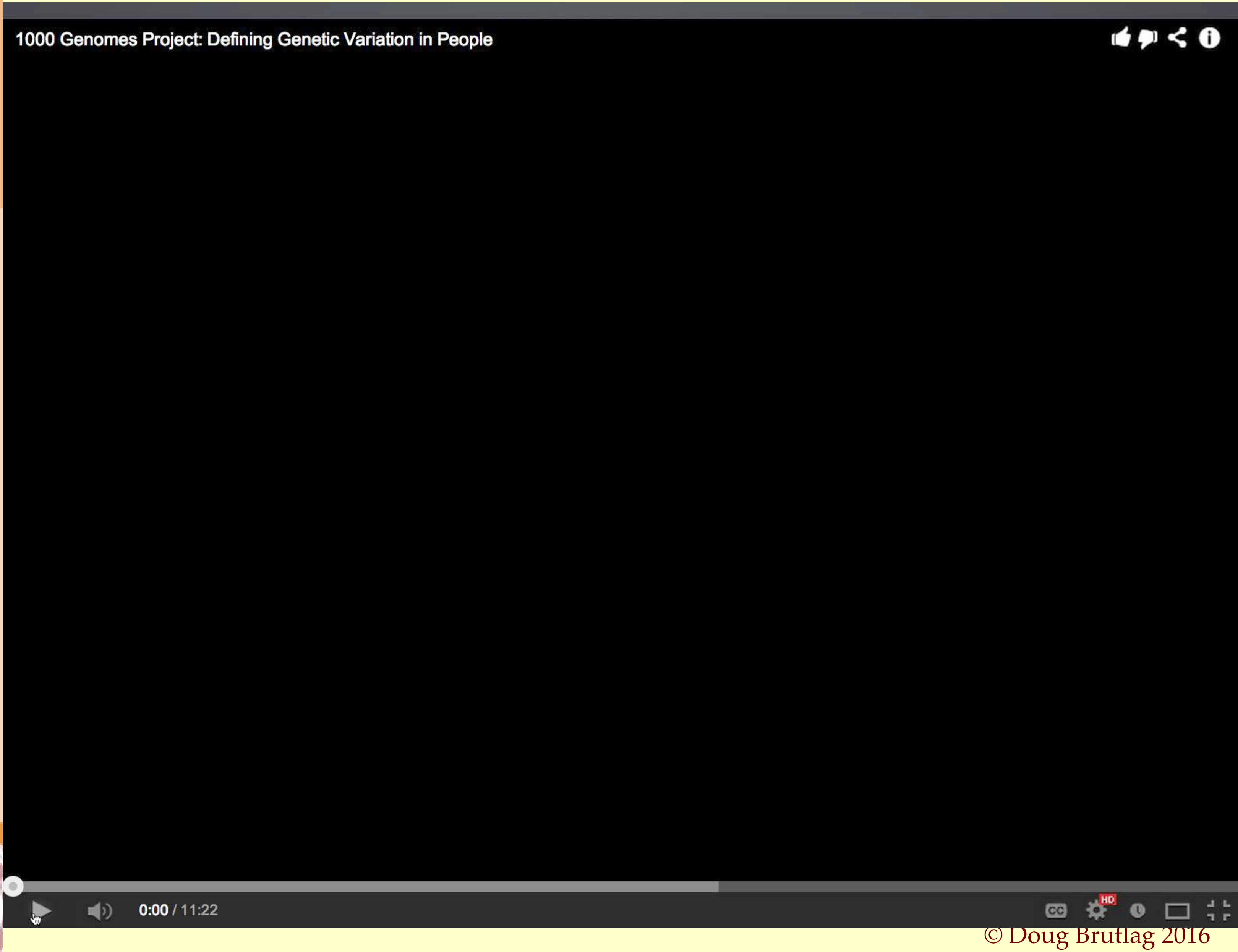
Media Archive

Find the 1000 Genomes Project Publications



Thousand Genomes Project

http://www.1000genomes.org/

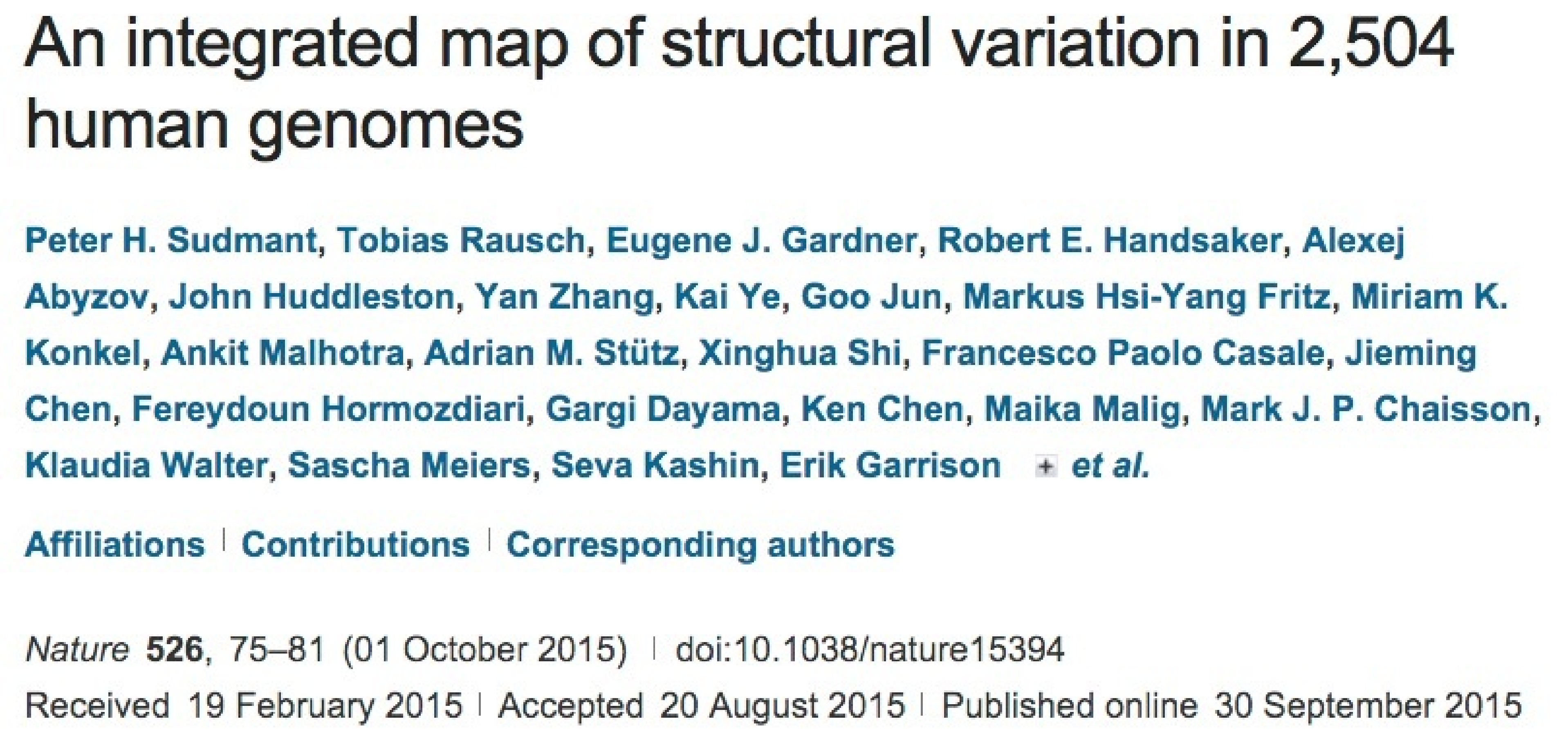


© Doug Brutlag 2016

1000 Genomes
A Deep Catalog of Human Genetic Variation

Thousand Genomes Project

http://www.1000genomes.org/



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

Peter H. Sudmant, Tobias Rausch, Eugene J. Gardner, Robert E. Handsaker, Alexej Abyzov, John Huddleston, Yan Zhang, Kai Ye, Goo Jun, Markus Hsi-Yang Fritz, Miriam K. Konkel, Ankit Malhotra, Adrian M. Stütz, Xinghua Shi, Francesco Paolo Casale, Jieming Chen, Fereydoun Hormozdiari, Gargi Dayama, Ken Chen, Maika Malig, Mark J. P. Chaisson, Klaudia Walter, Sascha Meiers, Seva Kashin, Erik Garrison + *et al.*

Affiliations | Contributions | Corresponding authors

Nature **526**, 75–81 (01 October 2015) | doi:10.1038/nature15394
Received 19 February 2015 | Accepted 20 August 2015 | Published online 30 September 2015

Nature 526, 75-81. (30 September 2015)

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 ethnic group has its own distribution of SNPs
- About 3 million sites where any individual varies from the consensus human genome.
- Each person differs from each other in 3 million places (about 0.1% of the genome)
- SNP sequence variations are common, unlike disease causing mutations which are rare.

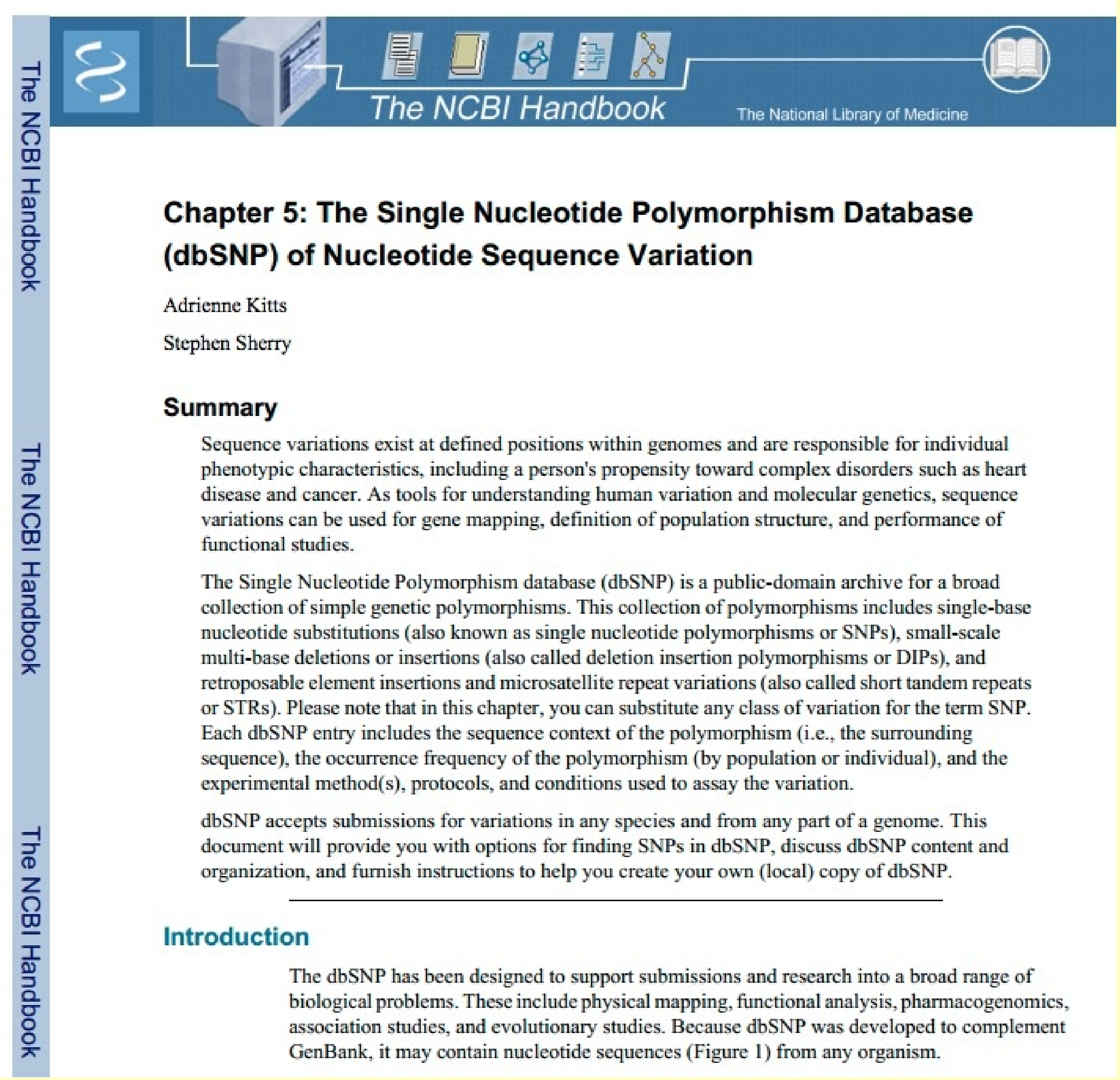
Single (Simple) Nucleotide Polymorphisms (SNPs)

GCTGTATGACTAGAAGATCGAT
GCTGTATGACGAGAAGATCGAT

- SNPs can be used for identifying individuals and forensics
- SNPs are used for mapping & genome-wide association studies of complex disease
- SNPs are used for ancestry tracking & family relationships
- SNPs are used for estimating predisposition to disease
- SNPs are used to predict risk of common genetic diseases
- While SNPs are linked with disease, they do not cause disease
- SNPs are used for personalized medicine and genomics
- SNPs are used for classifying patients in clinical trials
- In short, SNPs are used as genetic markers to map human diseases and traits and migrations.

The dbSNP Database

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



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.

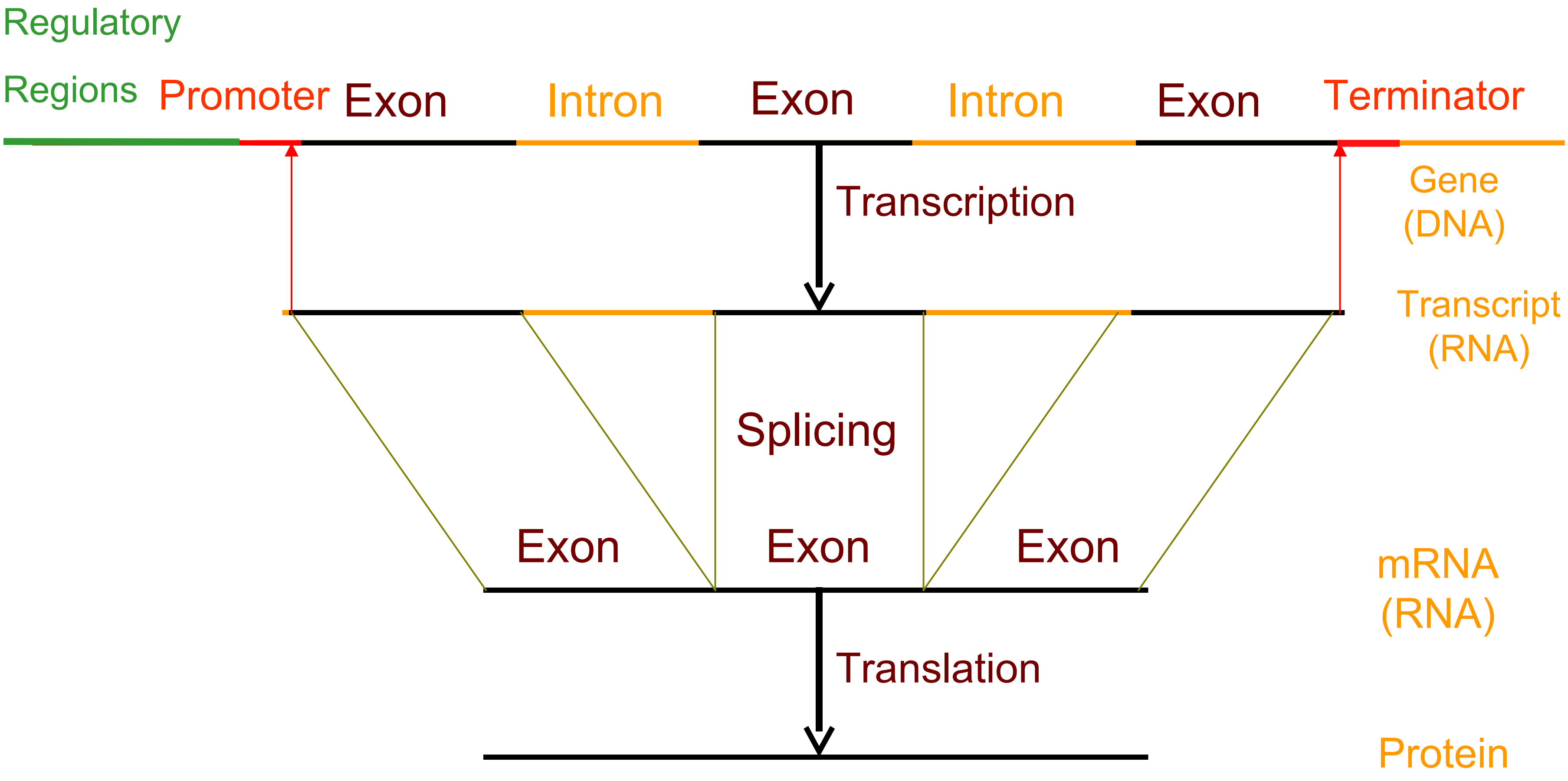
The Single Nucleotide Polymorphism database (dbSNP) is a public-domain archive for a broad collection of simple genetic polymorphisms. This collection of polymorphisms includes single-base nucleotide substitutions (also known as single nucleotide polymorphisms or SNPs), small-scale multi-base deletions 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.

Gene Transcription into RNA and Translation into Protein



Regulatory Regions

Promoter

Exon

Intron

Exon

Intron

Exon

Terminator

Gene (DNA)

Transcription

Transcript (RNA)

Splicing

Exon

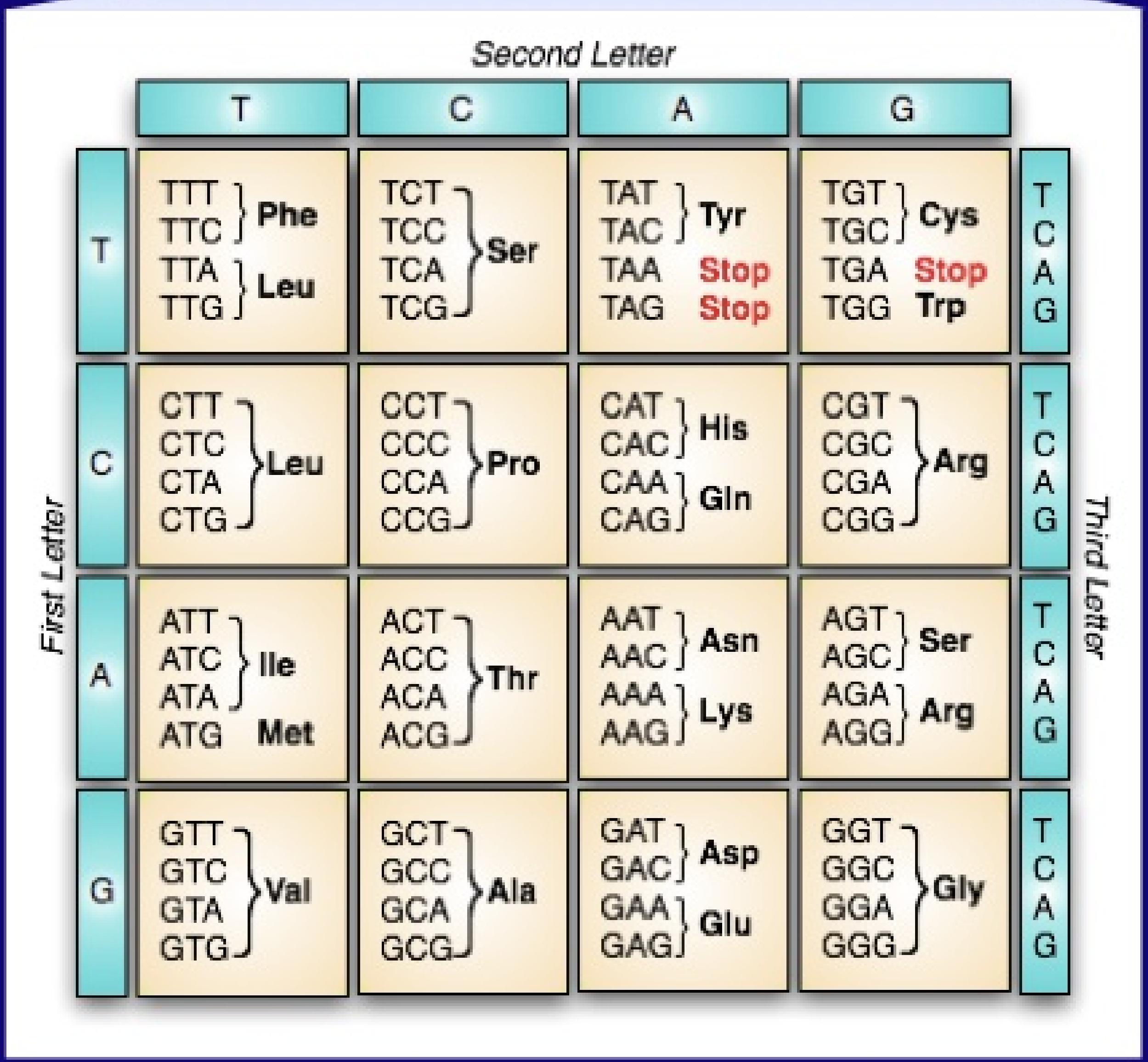
Exon

Exon

mRNA (RNA)

Translation

Protein

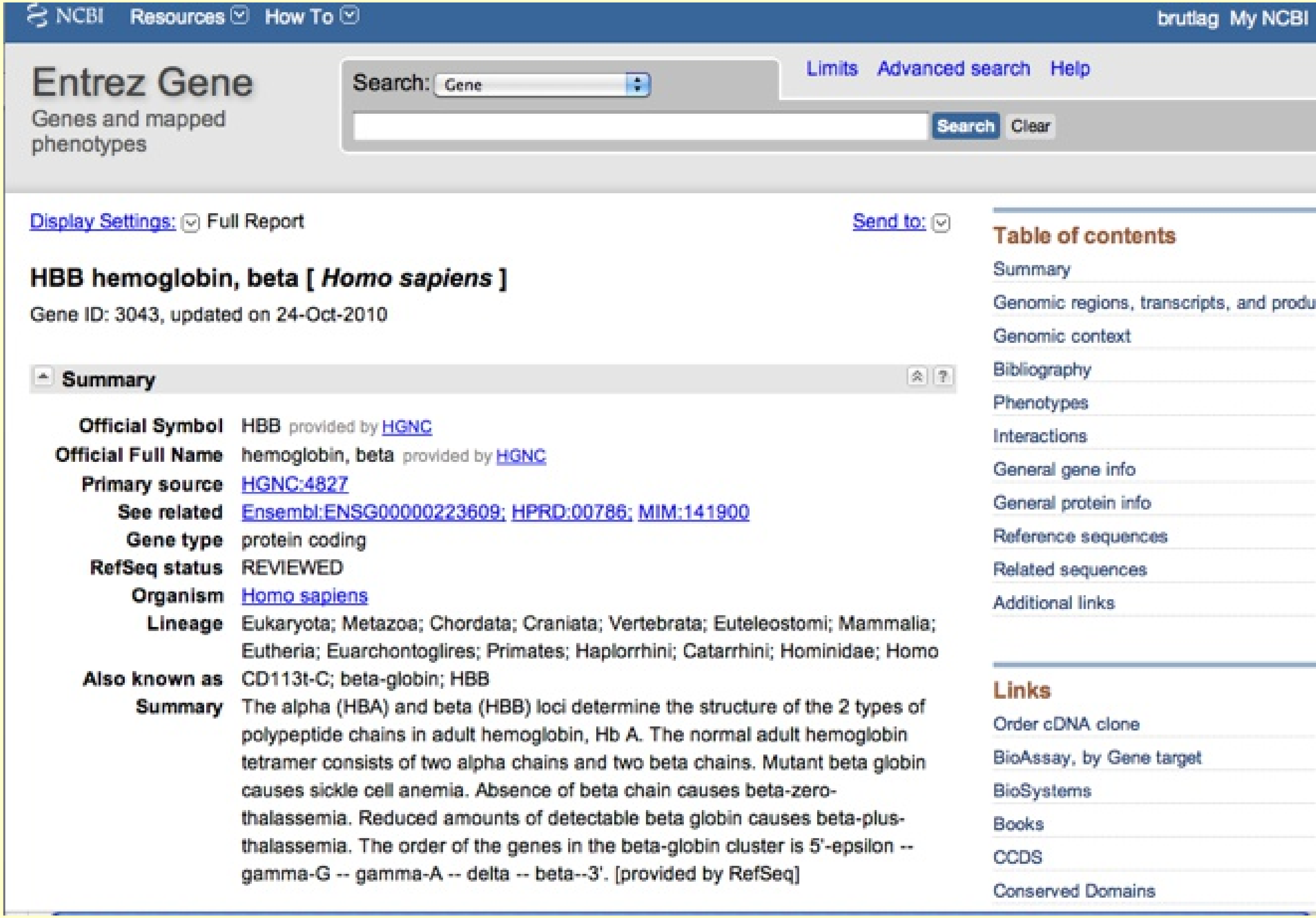


	Second Letter				
First Letter	T	C	A	G	Third Letter
T	TTT TTC } Phe TTA TTG } Leu	TCT TCC TCA TCG } Ser	TAT TAC } Tyr TAA Stop TAG Stop	TGT TGC } Cys TGA Stop TGG Trp	T C A G
C	CTT CTC CTA CTG } Leu	CCT CCC CCA CCG } Pro	CAT CAC } His CAA CAG } Gln	CGT CGC CGA CGG } Arg	T C A G
A	ATT ATC ATA } Ile ATG Met	ACT ACC ACA ACG } Thr	AAT AAC } Asn AAA AAG } Lys	AGT AGC } Ser AGA AGG } Arg	T C A G
G	GTT GTC GTA GTG } Val	GCT GCC GCA GCG } Ala	GAT GAC } Asp GAA GAG } Glu	GGT GGC GGA GGG } Gly	T C A G

The Genetic Code

Human β-Hemoglobin Gene

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



Entrez Gene

Genes and mapped phenotypes

Display Settings: Full Report

HBB hemoglobin, beta [*Homo sapiens*]

Gene ID: 3043, updated on 24-Oct-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	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo
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]

Table of contents

- Summary
- Genomic regions, transcripts, and produ
- Genomic context
- Bibliography
- Phenotypes
- Interactions
- General gene info
- General protein info
- Reference sequences
- Related sequences
- Additional links

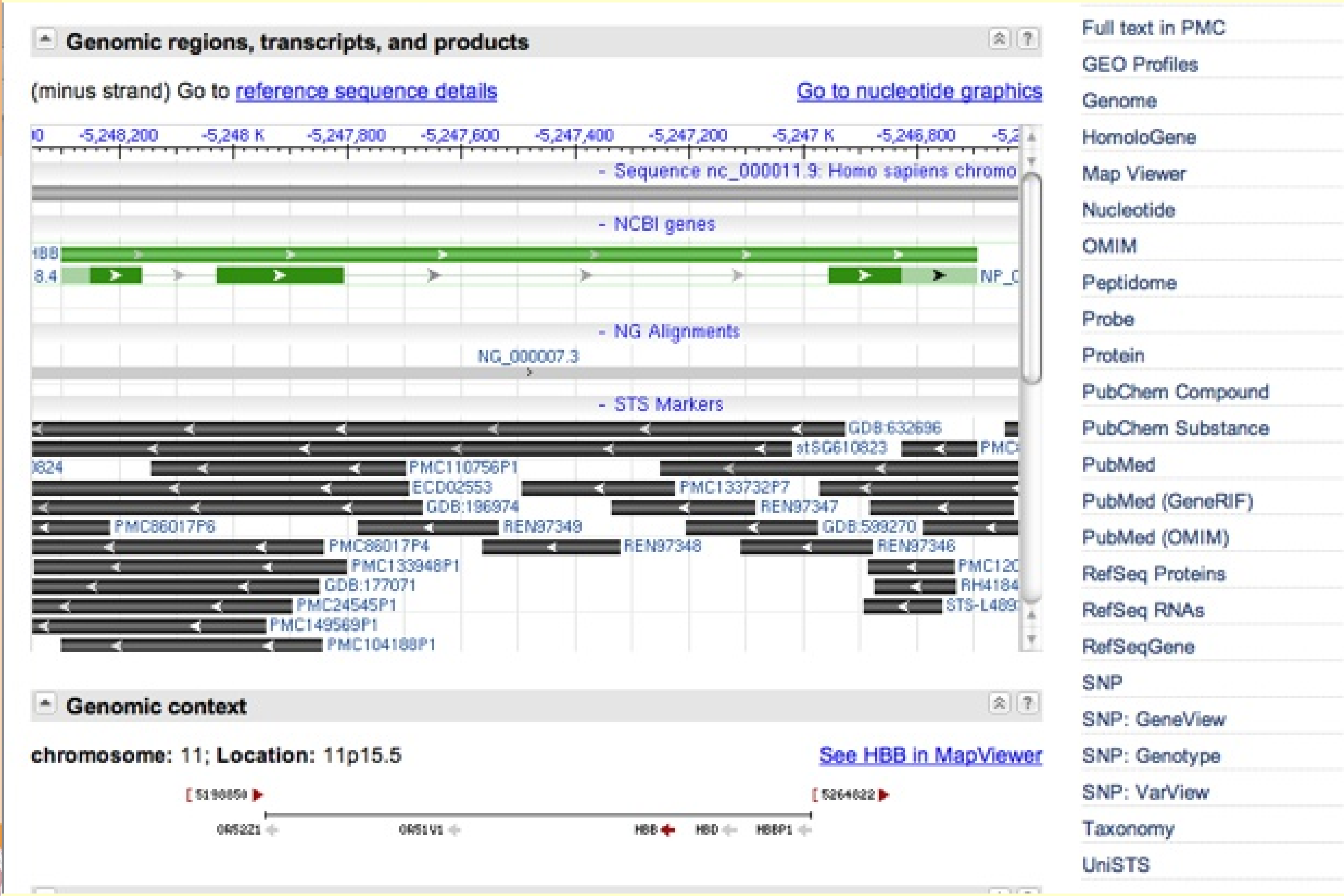
Links

- Order cDNA clone
- BioAssay, by Gene target
- BioSystems
- Books
- CCDS
- Conserved Domains

© Doug Brutlag 2016

Human β-Hemoglobin Gene

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



Genomic regions, transcripts, and products

(minus strand) Go to reference sequence details

Go to nucleotide graphics

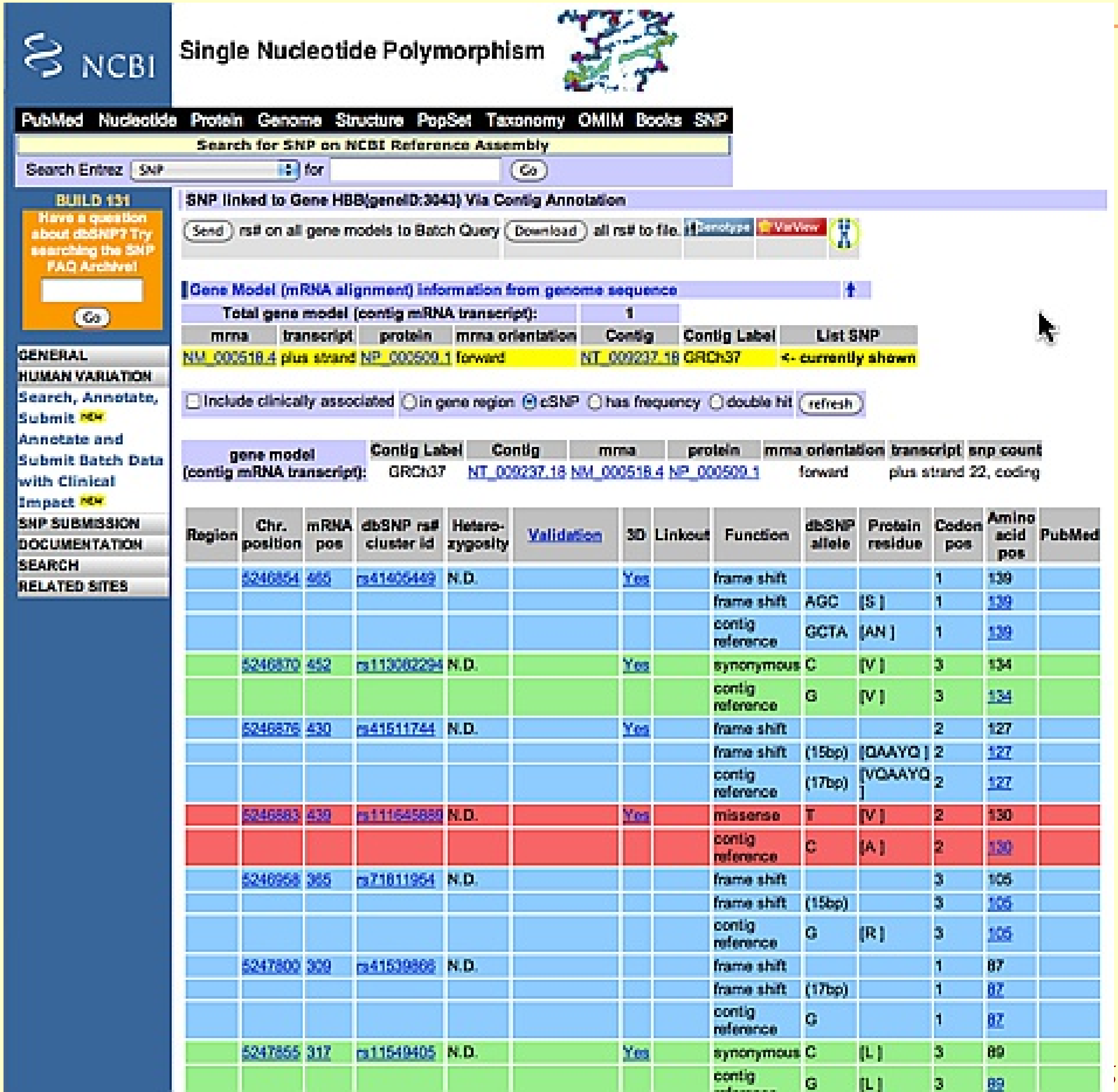
Genomic context

chromosome: 11; **Location:** 11p15.5

See HBB in MapViewer

Human β-Hemoglobin Gene SNPs

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



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

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	plus strand	NP_000509.1	forward	NT_009237.18	GRCh37	<- currently shown

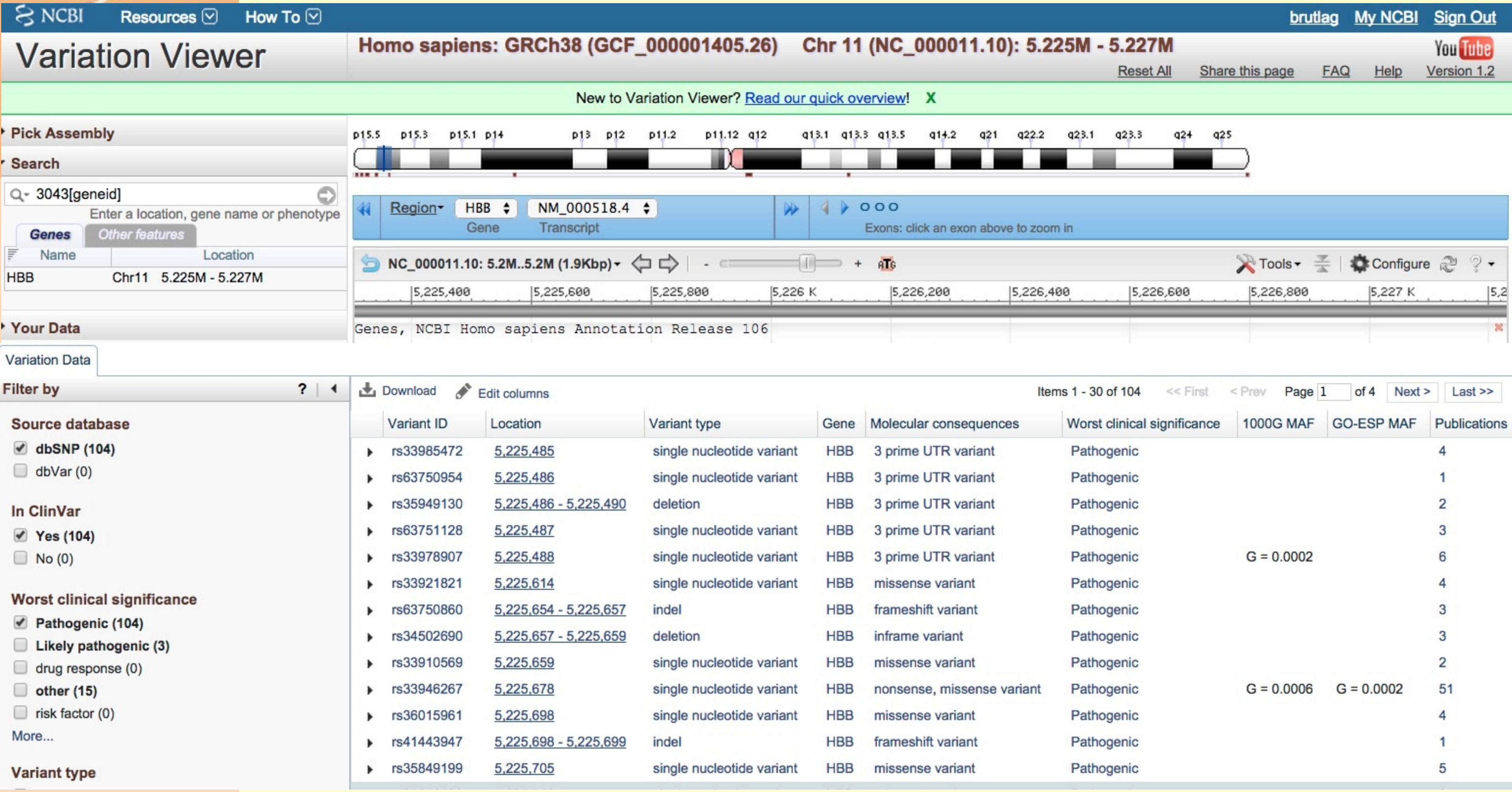
☐ Include clinically associated ○ in gene region ● cSNP ○ has frequency ○ double hit

gene model (contig mRNA transcript):	Contig Label	Contig	mrna	protein	mrna orientation	transcript	snp count
	GRCh37	NT_009237.18	NM_000518.4	NP_000509.1	forward	plus strand	22, coding

Region	Chr. position	mRNA pos	dbSNP rs# cluster id	Hetero-zygosity	Validation	3D	Linkout	Function	dbSNP allele	Protein residue	Codon pos	Amino acid pos	PubMed
	5246854	455	rs41405449	N.D.		Yes		frame shift			1	139	
								frame shift	AGC	[S]	1	139	
								contig reference	GCTA	[AN]	1	139	
	5246870	452	rs113082294	N.D.		Yes		synonymous	C	[V]	3	134	
								contig reference	G	[V]	3	134	
	5246876	430	rs41511744	N.D.		Yes		frame shift			2	127	
								frame shift	(15bp)	[QAAYQ]	2	127	
								contig reference	(17bp)	[VQAAYQ]	2	127	
	5246883	439	rs111645889	N.D.		Yes		missense	T	[V]	2	130	
								contig reference	C	[A]	2	130	
	5246958	365	rs71811954	N.D.				frame shift			3	105	
								frame shift	(15bp)		3	105	
								contig reference	G	[R]	3	105	
	5247800	309	rs41539868	N.D.				frame shift			1	87	
								frame shift	(17bp)		1	87	
								contig reference	G		1	87	
	5247855	317	rs11549405	N.D.		Yes		synonymous	C	[L]	3	89	
								contig	G	[L]	3	89	

β-Hemoglobin SNP Variation Viewer

http://www.ncbi.nlm.nih.gov/sites/varvu?gene=3043



Homo sapiens: GRCh38 (GCF_000001405.26) Chr 11 (NC_000011.10): 5.225M - 5.227M

New to Variation Viewer? Read our quick overview! X

3043[geneid]

Name	Location
HBB	Chr11 5.225M - 5.227M

HBB NM_000518.4

NC_000011.10: 5.2M..5.2M (1.9Kbp)

Genes, NCBI Homo sapiens Annotation Release 106

Source database
- ☑ dbSNP (104)
- ☐ dbVar (0)

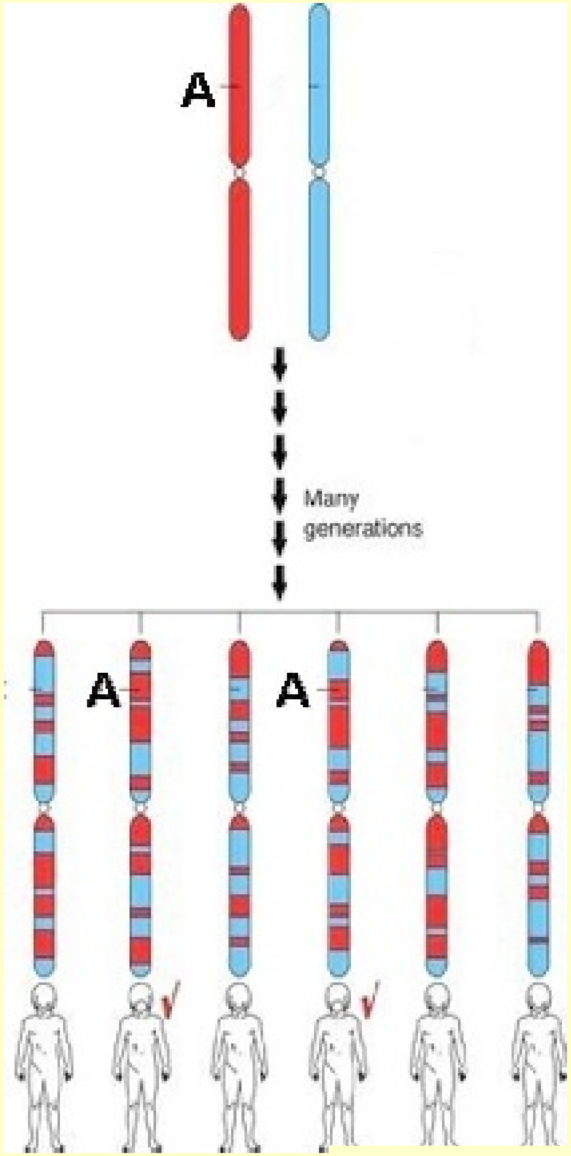
In ClinVar
- ☑ Yes (104)
- ☐ No (0)

Worst clinical significance
- ☑ Pathogenic (104)
- ☐ Likely pathogenic (3)
- ☐ drug response (0)
- ☐ other (15)
- ☐ risk factor (0)

Items 1 - 30 of 104 Page 1 of 4

Variant ID	Location	Variant type	Gene	Molecular consequences	Worst clinical significance	1000G MAF	GO-ESP MAF	Publications
rs33985472	5,225,485	single nucleotide variant	HBB	3 prime UTR variant	Pathogenic			4
rs63750954	5,225,486	single nucleotide variant	HBB	3 prime UTR variant	Pathogenic			1
rs35949130	5,225,486 - 5,225,490	deletion	HBB	3 prime UTR variant	Pathogenic			2
rs63751128	5,225,487	single nucleotide variant	HBB	3 prime UTR variant	Pathogenic			3
rs33978907	5,225,488	single nucleotide variant	HBB	3 prime UTR variant	Pathogenic	G = 0.0002		6
rs33921821	5,225,614	single nucleotide variant	HBB	missense variant	Pathogenic			4
rs63750860	5,225,654 - 5,225,657	indel	HBB	frameshift variant	Pathogenic			3
rs34502690	5,225,657 - 5,225,659	deletion	HBB	inframe variant	Pathogenic			3
rs33910569	5,225,659	single nucleotide variant	HBB	missense variant	Pathogenic			2
rs33946267	5,225,678	single nucleotide variant	HBB	nonsense, missense variant	Pathogenic	G = 0.0006	G = 0.0002	51
rs36015961	5,225,698	single nucleotide variant	HBB	missense variant	Pathogenic			4
rs41443947	5,225,698 - 5,225,699	indel	HBB	frameshift variant	Pathogenic			1
rs35849199	5,225,705	single nucleotide variant	HBB	missense variant	Pathogenic			5

Origin of Haplotypes



© Gibson & Muse, A Primer of Genome Science

SNP Linkage in the Human LPL Gene

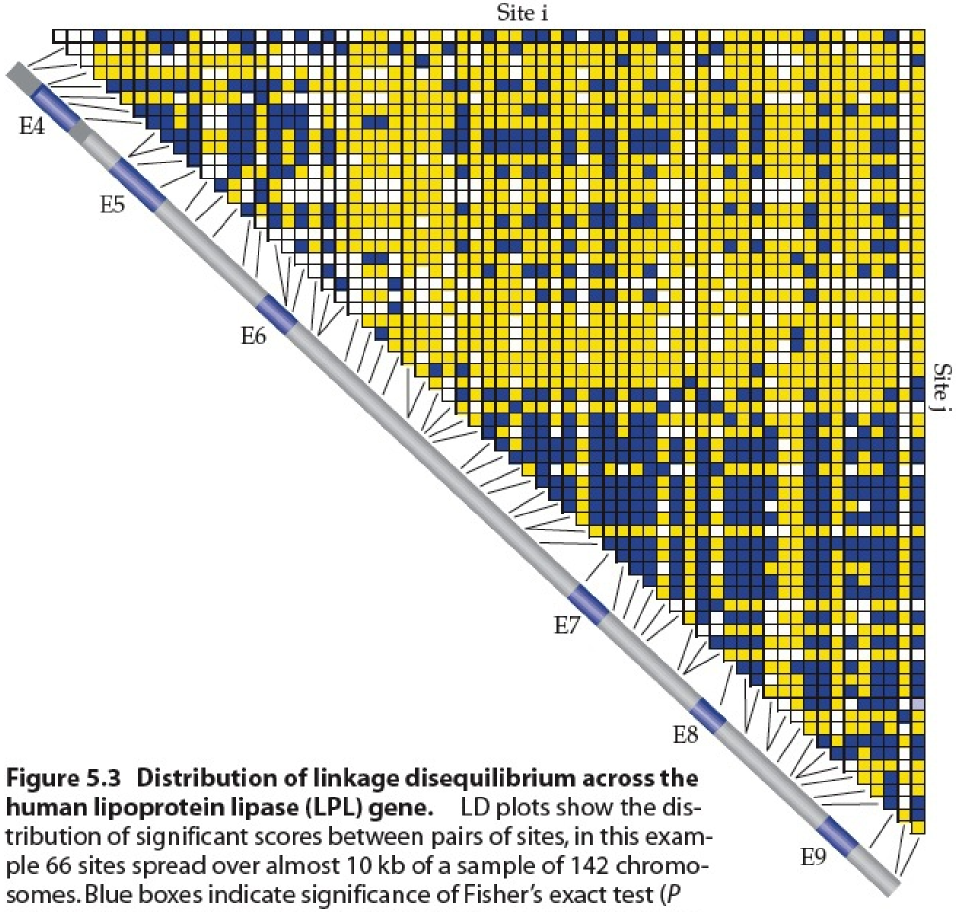
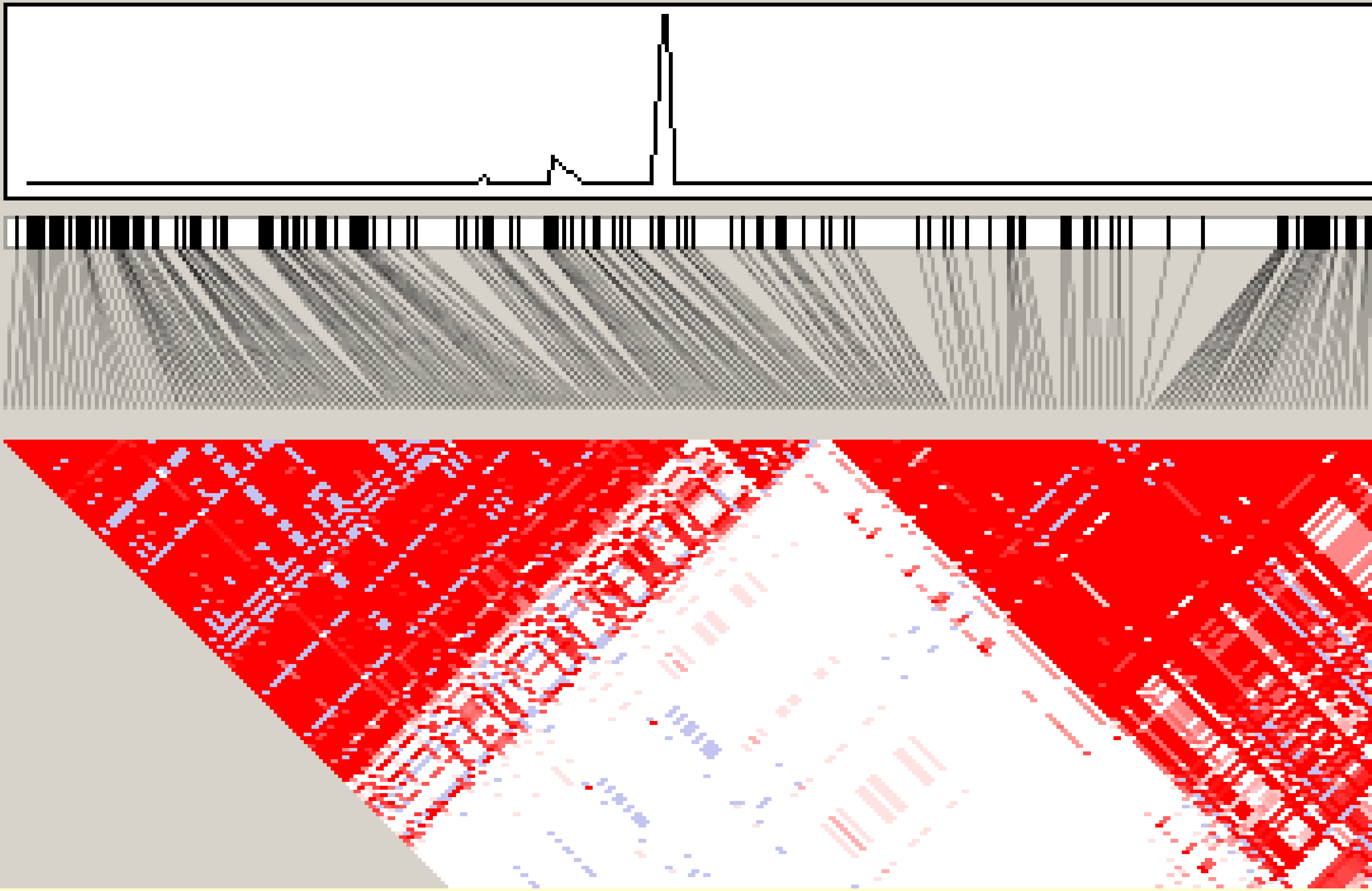


Figure 5.3 Distribution of linkage disequilibrium across the human lipoprotein lipase (LPL) gene. LD plots show the distribution of significant scores between pairs of sites, in this example 66 sites spread over almost 10 kb of a sample of 142 chromosomes. Blue boxes indicate significance of Fisher's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

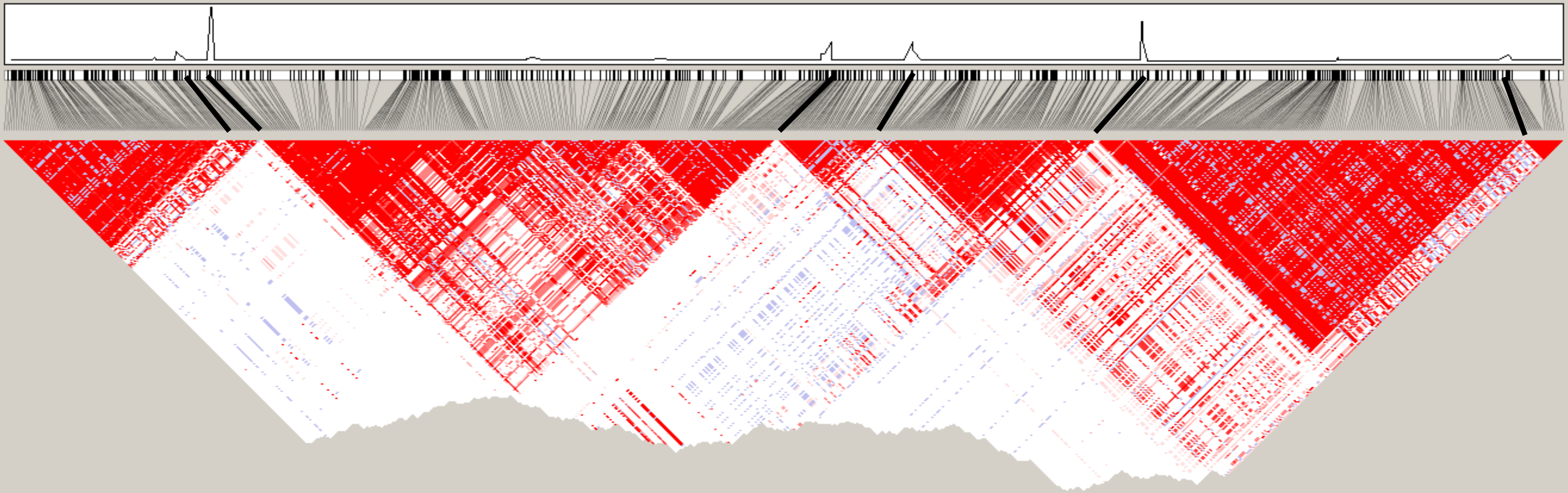
Recombination hotspots are widespread and account for SNP linkage structure



7q21

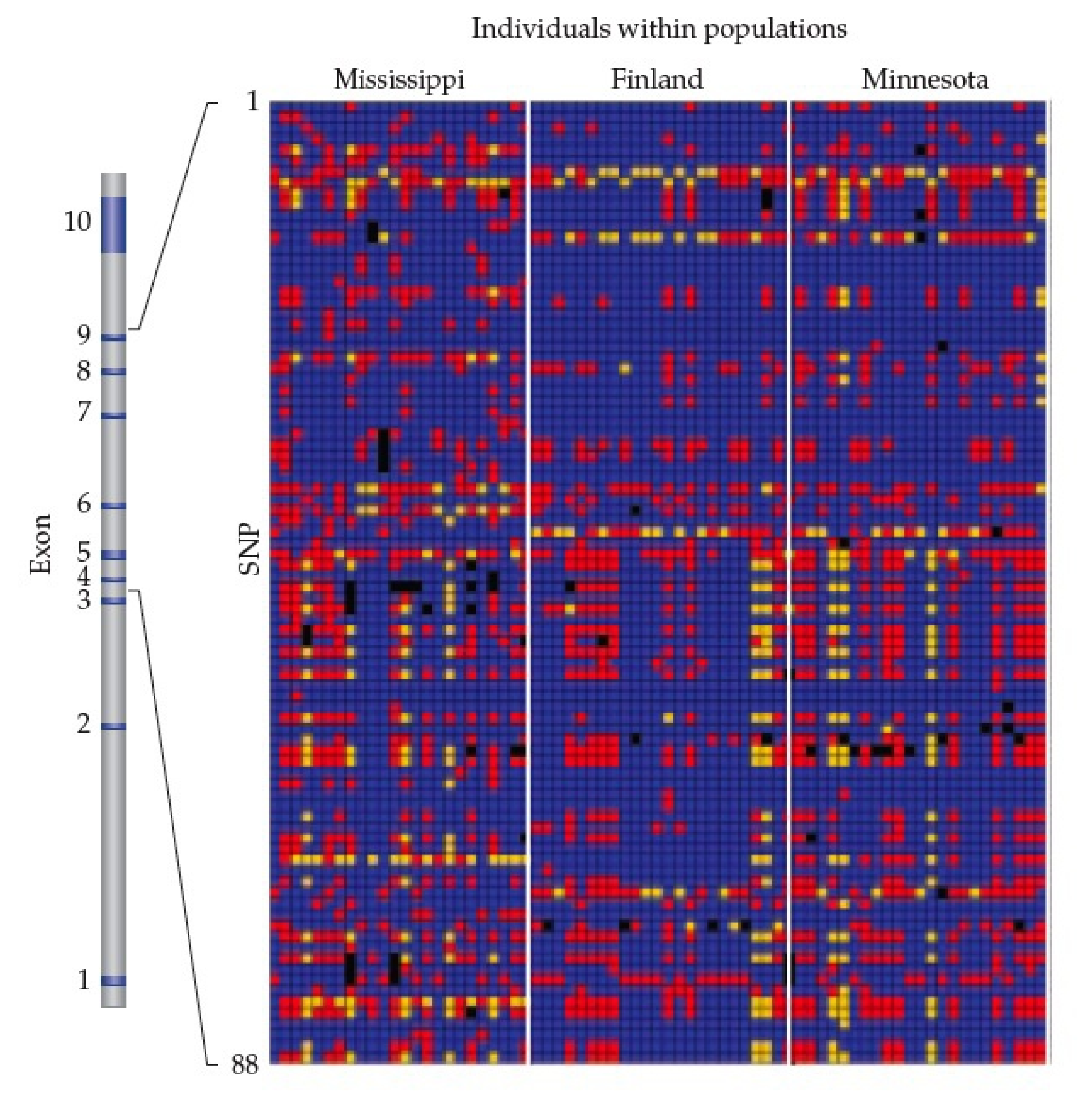
© Gibson & Muse, A Primer of Genome Science

Recombination hotspots are widespread and account for SNP linkage structure



7q21

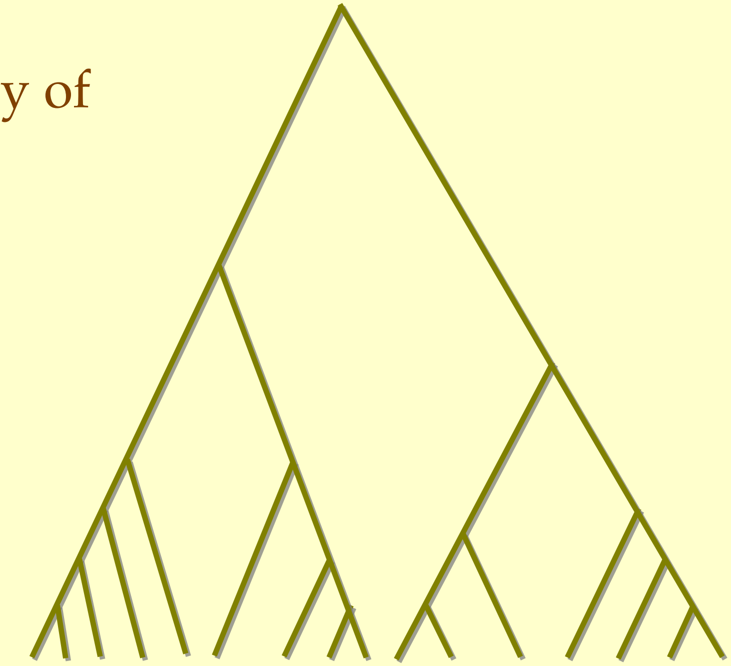
SNPs in Populations

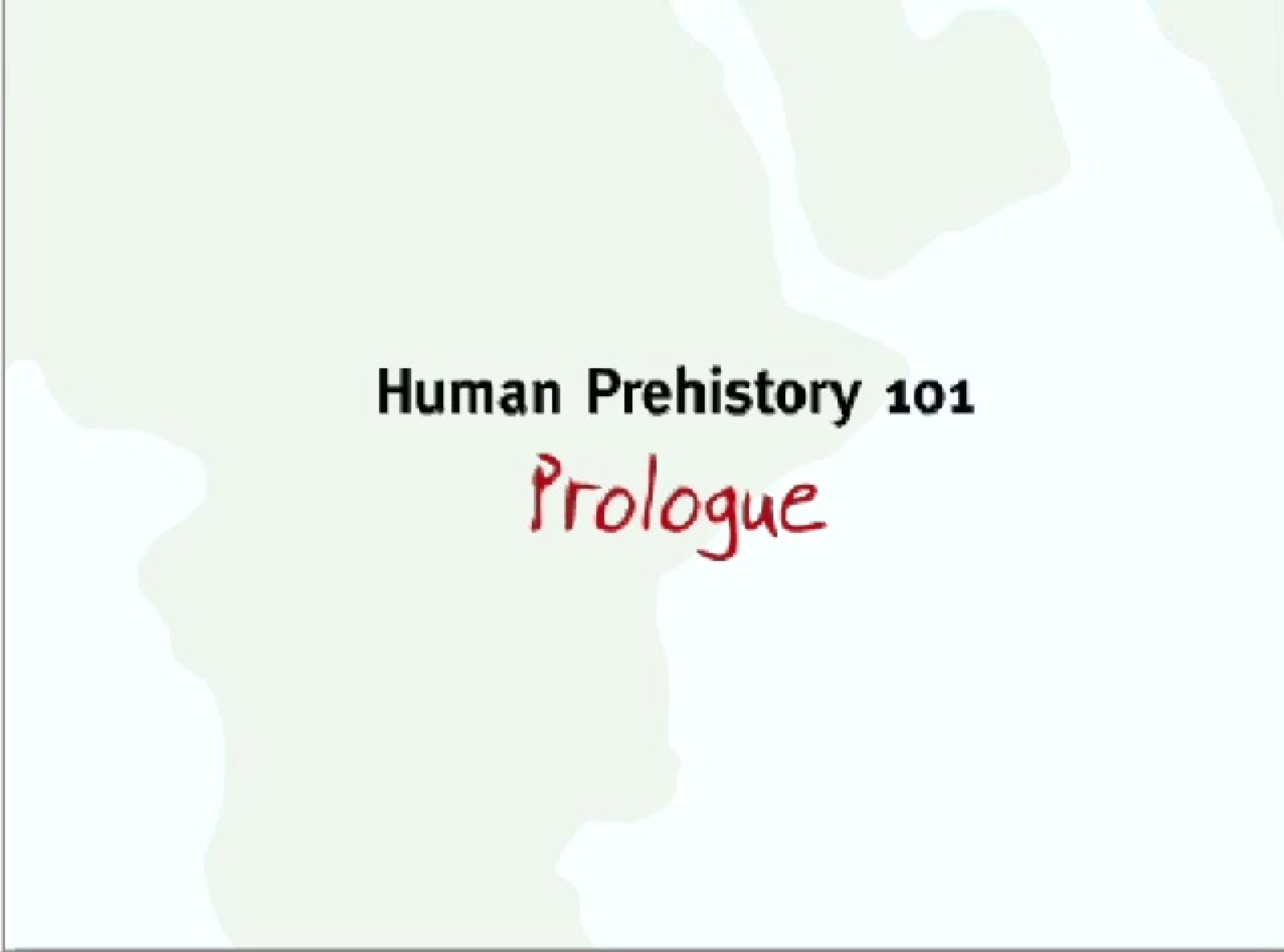


© Gibson & Muse, A Primer of Genome Science

Sequence and Distance-Based Evolutionary Trees

- Sequence-Based Methods
  - Assigns mutations to branches
  - Minimize number of changes
  - Topology maximizes similarity of neighboring leaves





Human Prehistory 101

Prologue

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?