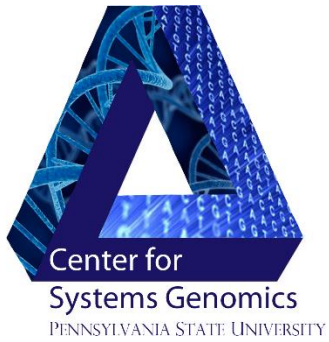
Retrieving information on genes and proteins from biological and genomic databases

Marylyn D Ritchie, PhD
Professor, Biochemistry and Molecular Biology
Director, Center for Systems Genomics
The Pennsylvania State University

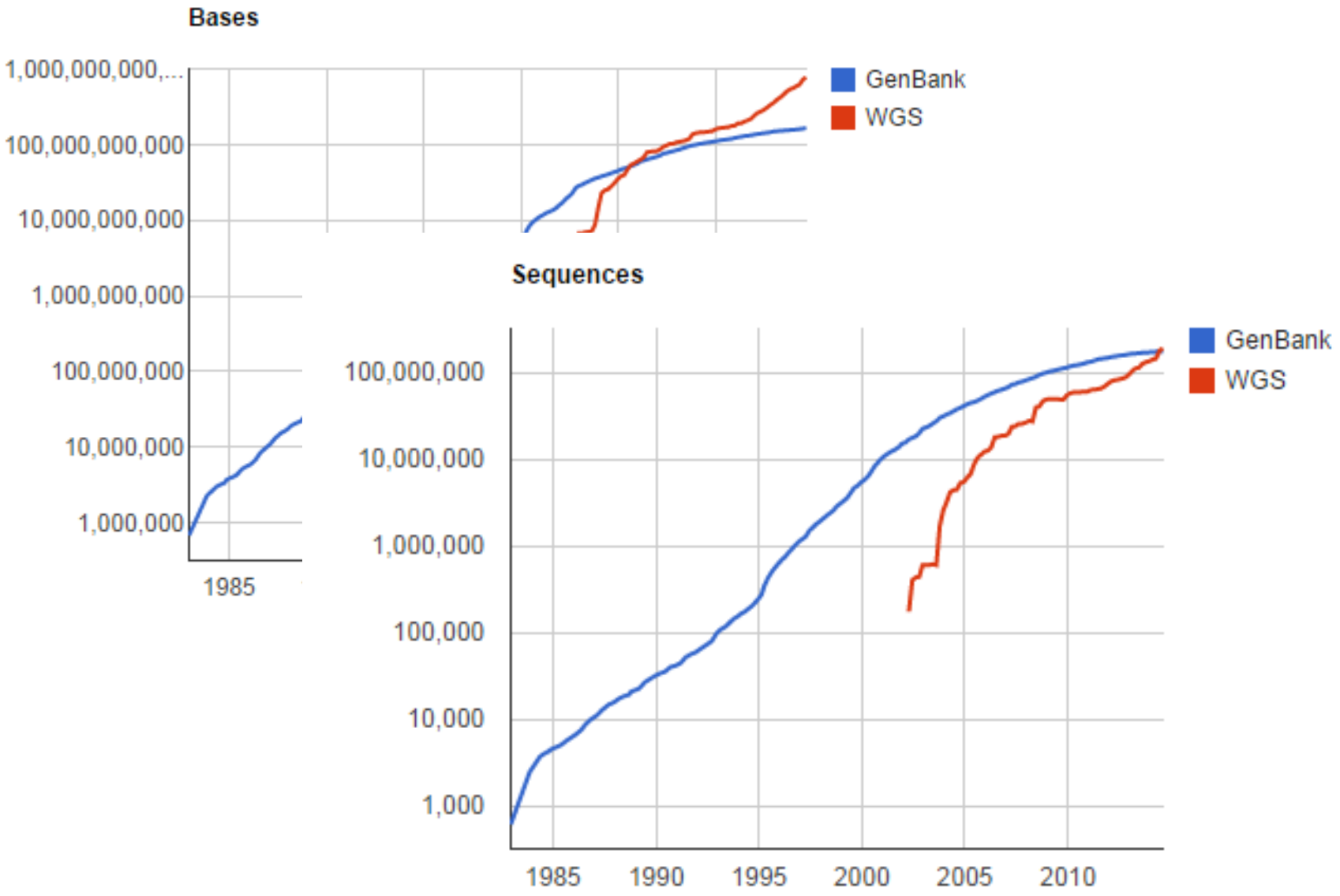


GenBank

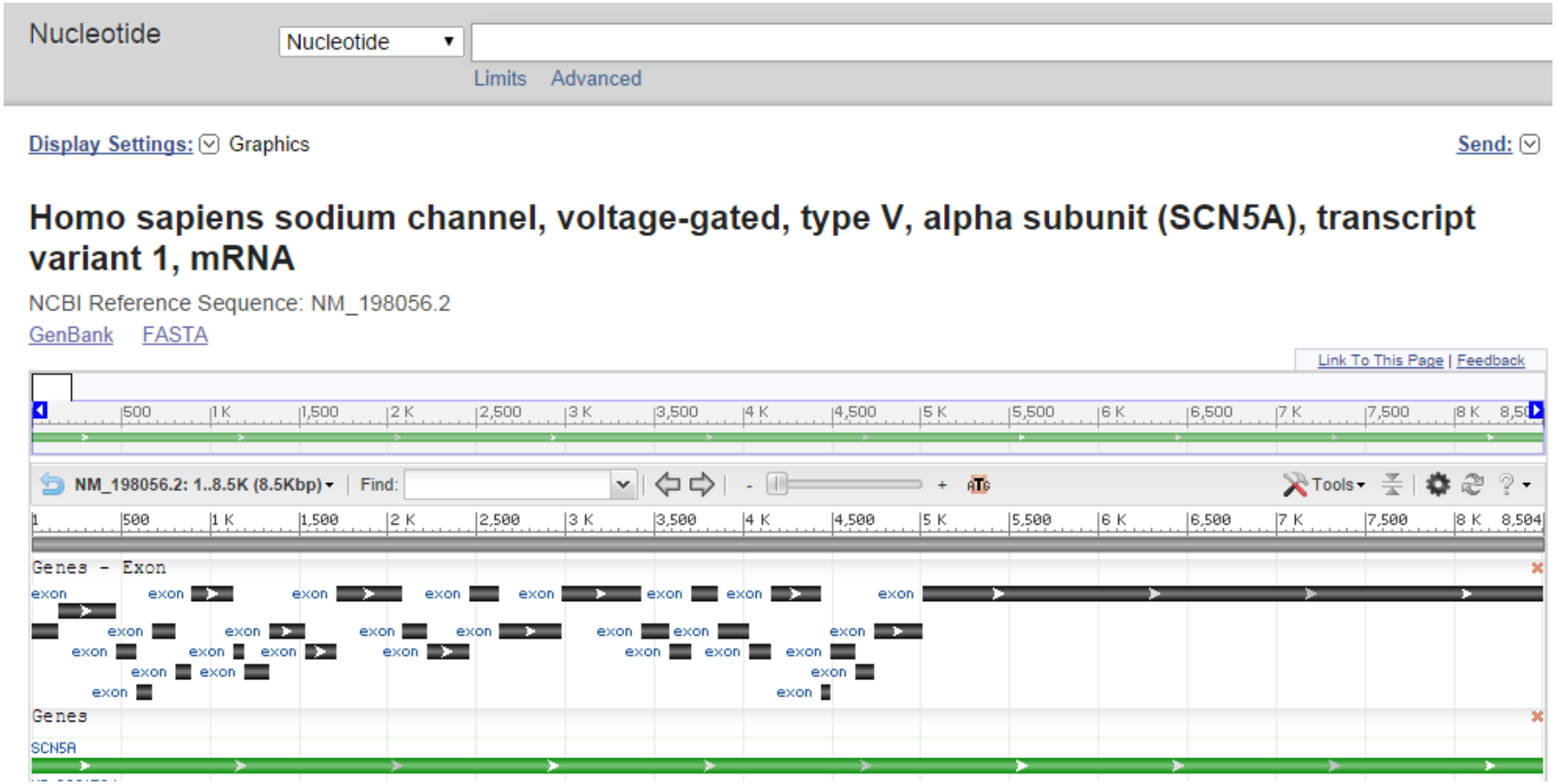
- Repository of nucleic acid sequences
- As of 2001, held 9.5 billion bases in 8.2 million entries

		GenBank		WGS	
Release	Date	Bases	Sequences	Bases	Sequences
3	Dec 1982	680338	606		
119	Aug 2000	9,545,724,824	8,214,339		
129	Apr 2002	19,072,679,701	16,769,983	692,266,338	172,768
203	Aug 2014	165,722,980,375	174,108,750	774,052,098,731	189,080,419

GenBank


Bases
1,000,000,000,...
100,000,000,000
10,000,000,000
1,000,000,000
100,000,000
10,000,000
1,000,000
1985
GenBank
WGS
Sequences
100,000,000
10,000,000
1,000,000
100,000
10,000
1,000
1985
1990
1995
2000
2005
2010
GenBank
WGS

GenBank



SCN5A – 32 exons

NCBI Resources How To Sign in to NCBI

Nucleotide Nucleotide Search

Limits Advanced Help

Display Settings: GenBank Send:

Homo sapiens sodium channel, voltage-gated, type V, alpha subunit (SCN5A), transcript variant 1, mRNA

NCBI Reference Sequence: NM_198056.2

FASTA Graphics

Go to:

```
LOCUS       NM_198056               8504 bp    mRNA    linear   PRI 03-MAY-2014
DEFINITION  Homo sapiens sodium channel, voltage-gated, type V, alpha subunit
            (SCN5A), transcript variant 1, mRNA.
ACCESSION   NM_198056 XM_940695
VERSION     NM_198056.2  GI:124518659
KEYWORDS    RefSeq.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
            Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 8504)
  AUTHORS   Beltran-Alvarez P, Espejo A, Schmauder R, Beltran C, Mrowka R,
            Linke T, Batlle M, Perez-Villa F, Perez GJ, Scornik FS, Benndorf K,
            Pagans S, Zimmer T and Brugada R.
  TITLE     Protein arginine methyl transferases-3 and -5 increase cell surface
            expression of cardiac sodium channel
  JOURNAL   FEBS Lett. 587 (19), 3159-3165 (2013)
   PUBMED   23912080
```

Change region shown

Customize view

Basic Features
- Default features
- Gene, RNA, and CDS features only

Features added by NCBI
- 801 SNPs
- 7 conserved domains

Display options
- Show sequence
- Show reverse complement

Update View

Analyze this sequence
- Run BLAST
- Pick Primers
- Highlight Sequence Features
- Find in this Sequence

http://www.ncbi.nlm.nih.gov/nuccore/NM_198056.2

Amino Acid Sequence

```
/translation="MANFLLPRGTSSFRRFTRESLAAIEKRMAEKQA
GLPEEEAPRPQLDLQASKKLPDLYGNPPQELIGEPLEDLDPFYSTQK
FRFSATNALYVLSPFHPIRRAAVKILVHSLFNMLIMCTILTNCVFMA
EYTFTAIYTFESLVKILARGFCLHAFTFLRDPWNWLDFSVIIMAYTT
RTFRVLRALKTISVISGLKTIVGALIQSVKKLADVMVLTVFCLSVFA
RHKCVRNFTALNGTNGSVEADGLVWESLDLYLSDPENYLLKNGTSDV
CPEGYRCLKAGENPDHGYTSFDSFAWAFLALFRLMTQDCWERLYQQT
FMLVIFLGSFYLVNLILAVVAMAYEEQNQATIAETEEKEKRFQEAME
RGVDTVSRSSLEMSPLAPVNSHERRSKRRKRMSSGTEECGEDRLPKS
LSLTRGLSRTSMKPRSSRGSIFTFRRRDLGSEADFADDENSTAGESE
LRRTSAQGQPSPGTSAPGHALHGKKNSTVDCNGVVSLLGAGDPEATS
EHPPDTTTPSEEPGGPQMLTSQAPCVDGFEEPGARQRALSAVSVLTS
CPPCWNRLAQRYLIWECCPLWMSIKQGVKLVVMDPFTDLTITMCIVL
MTSEFEEMLQVGNLVFTGIFTAEMTFKIIALDPYYYFQQGWNIFDSI
SRMSNLSVLRSFRLLRVFKLAKSWPTLNTLIKIIGNSVGALGNLTLV
GMQLFGKNYSELRDSDSGLLPRWHMMDFFHAFLIIFRILCGEWIETM
CLLVFLLVMVIGNLVVLNLFLALLLSSFSADNLTAPDEDREMNNLQL
VKRTTWDFCCGLLRQRPQKPAALAAQGQLPSCIATPYSPPPPETEKV
GEQPGQGTPGDPEPVCVPIAVAESDTDDQEEDEENSLGTEEESSKQQ
PPDSRTWSQVSATASSEAEASASQADWRQQWKAEPQAPGCGETPEDS
TAELLEQIPDLGQDVKDPEDCFTEGCVRRCPCCAVDTTQAPGKVWWR
SWFETFIIFMILLSSGALAFEDIYLEERKTIKVLLEYADKMFTYVFV
FKKYFTNAWCWLDFLIVDVSLVSLVANTLGFAEMGPIKSLRTLRALR
RVVVNALVGAIPSIMNVLLVCLIFWLIFSIMGVNLFAGKFGRCINQT
NNKSQCESLNLTGELYWTKVKVNFDNVGAGYLALLQVATFKGWMDIM
QPQWEYNLYMYIYFVIFIIFGSFFTLNLFIGVIIDNFNQQKKKLGGQ
YNAMKKLGSKKPQKPIPRPLNKYQGFIFDIVTKQAFDVTIMFLICLN
SPEKINILAKINLLFVAIFTGECIVKLAALRHYYFTNSWNIFDFVVV
IIQKYFFSPTLFRVIRLARIGRILRLIRGAKGIRTLLFALMMSLPAL
FIYSIFGMANFAYVKWEAGIDDMFNFQTFANSMLCLFQITTSAGWDG
YCDPTLPNSNGSRGDCGSPAVGILFFTTYIIISFLIVVNMYIAIILE
PLSEDDFDMFYEIWEKFDPEATQFIEYSVLSDFADALSEPLRIAKPN
VSGDRIHCMDILFAFTKRVLGESGEMDALKIQMEEKFMAANPSKISY
EEVSAMVIQRAFRRHLLQRSLKHASFLFRQQAGSGLSEEDAPEREGL
PLGPPSSSSISSTSFPPSYDSVTRATSDNLQVRGSDYSHSEDLADFP
"
```

Nucleotide Sequence

```
ORIGIN
        1 agacggcggc ggcgcccgta ggatgcaggg atcgctcccc cggggccgct gagcctgcgc
       61 ccagtgcccc gagccccgcg ccgagccgag tccgcgccaa gcagcagccg cccaccccgg
      121 ggcccggccg ggggaccagc agcttcccca caggcaacgt gaggagagcc tgtgcccaga
      181 agcaggatga gaagatggca aacttcctat tacctcgggg caccagcagc ttccgcaggt
      241 tcacacggga gtccctggca gccatcgaga agcgcatggc agagaagcaa gcccgcggct
      301 caaccacctt gcaggagagc cgagaggggc tgcccgagga ggaggctccc cggccccagc
      361 tggacctgca ggcctccaaa aagctgccag atctctatgg caatccaccc caagagctca
      421 tcggagagcc cctggaggac ctggacccct tctatagcac ccaaaagact ttcatcgtac
      481 tgaataaagg caagaccatc ttccggttca gtgccaccaa cgccttgtat gtcctcagtc
      541 ccttccaccc catccggaga gcggctgtga agattctggt tcactcgctc ttcaacatgc
      601 tcatcatgtg caccatcctc accaactgcg tgttcatggc ccagcacgac cctccaccct
      661 ggaccaagta tgtcgagtac accttcaccg ccatttacac ctttgagtct ctggtcaaga
      721 ttctggctcg aggcttctgc ctgcacgcgt tcactttcct tcgggaccca tggaactggc
      781 tggactttag tgtgattatc atggcataca caactgaatt tgtggacctg ggcaatgtct
      841 cagccttacg caccttccga gtcctccggg ccctgaaaac tatatcagtc atttcagggc
      901 tgaagaccat cgtgggggcc ctgatccagt ctgtgaagaa gctggctgat gtgatggtcc
      961 tcacagtctt ctgcctcagc gtctttgccc tcatcggcct gcagctcttc atgggcaacc
     1021 taaggcacaa gtgcgtgcgc aacttcacag cgctcaacgg caccaacggc tccgtggagg
     1081 ccgacggctt ggtctgggaa tccctggacc tttacctcag tgatccagaa aattacctgc
     1141 tcaagaacgg cacctctgat gtgttactgt gtgggaacag ctctgacgct gggacatgtc
     1201 cggagggcta ccggtgccta aaggcaggcg agaaccccga ccacggctac accagcttcg
     1261 attcctttgc ctgggccttt cttgcactct tccgcctgat gacgcaggac tgctgggagc
     1321 gcctctatca gcagaccctc aggtccgcag ggaagatcta catgatcttc ttcatgcttg
     1381 tcatcttcct ggggtccttc tacctggtga acctgatcct ggccgtggtc gcaatggcct
     1441 atgaggagca aaaccaagcc accatcgctg agaccgagga gaaggaaaag cgcttccagg
     1501 aggccatgga aatgctcaag aaagaacacg aggccctcac catcaggggt gtggataccg
     1561 tgtcccgtag ctccttggag atgtcccctt tggccccagt aaacagccat gagagaagaa
     1621 gcaagaggag aaaacggatg tcttcaggaa ctgaggagtg tggggaggac aggctcccca
     1681 agtctgactc agaagatggt cccagagcaa tgaatcatct cagcctcacc cgtggcctca
     1741 gcaggacttc tatgaagcca cgttccagcc gcgggagcat tttcaccttt cgcaggcgag
     1801 acctgggttc tgaagcagat tttgcagatg atgaaaacag cacagcgggg gagagcgaga
     1861 gccaccacac atcactgctg gtgccctggc ccctgcgccg gaccagtgcc cagggacagc
     1921 ccagtcccgg aacctcggct cctggccacg ccctccatgg caaaaagaac agcactgtgg
     1981 actgcaatgg ggtggtctca ttactggggg caggcgaccc agaggccaca tccccaggaa
     2041 gccacctcct ccgccctgtg atgctagagc acccgccaga cacgaccacg ccatcggagg
     2101 agccaggcgg gccccagatg ctgacctccc aggctccgtg tgtagatggc ttcgaggagc
     2161 caggagcacg gcagcgggcc ctcagcgcag tcagcgtcct caccagcgca ctggaagagt
     2221 tagaggagtc tcgccacaag tgtccaccat gctggaaccg tctcgcccag cgctacctga
```

Gene

Gene ▾ Search

Advanced

Display Settings: Full Report

Send to:

Hide s

SCN5A sodium channel, voltage-gated, type V, alpha subunit [*Homo sapiens* (human)]

Gene ID: 6331, updated on 29-Sep-2014

Summary

Official Symbol	SCN5A provided by HGNC
Official Full Name	sodium channel, voltage-gated, type V, alpha subunit provided by HGNC
Primary source	HGNC:HGNC:10593
See related	Ensembl:ENSG00000183873; HPRD:02543; MIM:600163; Vega:OTTHUMG00000156166
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	HB1; HB2; HH1; IVF; VF1; HBBD; ICCD; LQT3; SSS1; CDCD2; CMD1E; CMPD2; PFHB1; Nav1.5
Summary	The protein encoded by this gene is an integral membrane protein and tetrodotoxin-resistant voltage-gated sodium channel subunit. This protein is found primarily in cardiac muscle and is responsible for the initial upstroke of the action potential in an electrocardiogram. Defects in this gene are a cause of long QT syndrome type 3 (LQT3), an autosomal dominant cardiac disease. Alternative splicing results in several transcript variants encoding different isoforms. [provided by RefSeq, Jul 2008]

Genomic context

Location: 3p21
Exon count: 32

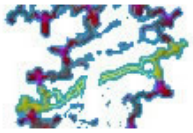
See SCN5A in Epigenomics, MapViewer

Table of contents

Related information

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:6331) Via Contig Annotation

The SNP GeneView page only reports human variation on GRCh38. A new Variation Viewer is available to view the gene SCN5A 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):				9		
mrna	transcript	protein	mrna orientation	Contig	Contig Label	List SNP
NM_198056.2	minus strand	NP_932173.1	reverse	NT_022517.19	GRCh38	<- currently shown
NM_001099404.1	minus strand	NP_001092874.1	reverse	NT_022517.19	GRCh38	View snp on GeneModel
NM_000335.4	minus strand	NP_000326.2	reverse	NT_022517.19	GRCh38	View snp on GeneModel
XM_006713284.1	minus strand	XP_006713347.1	reverse	NT_022517.19	GRCh38	View snp on GeneModel
XM_006713283.1	minus strand	XP_006713346.1	reverse	NT_022517.19	GRCh38	View snp on GeneModel
XM_006713282.1	minus strand	XP_006713345.1	reverse	NT_022517.19	GRCh38	View snp on GeneModel
NM_001160161.1	minus strand	NP_001153633.1	reverse	NT_022517.19	GRCh38	View snp on GeneModel
NM_001160160.1	minus strand	NP_001153632.1	reverse	NT_022517.19	GRCh38	View snp on GeneModel
NM_001099405.1	minus strand	NP_001092875.1	reverse	NT_022517.19	GRCh38	View snp on GeneModel

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_022517.19	NM_198056.2	NP_932173.1	reverse	minus strand	238, coding

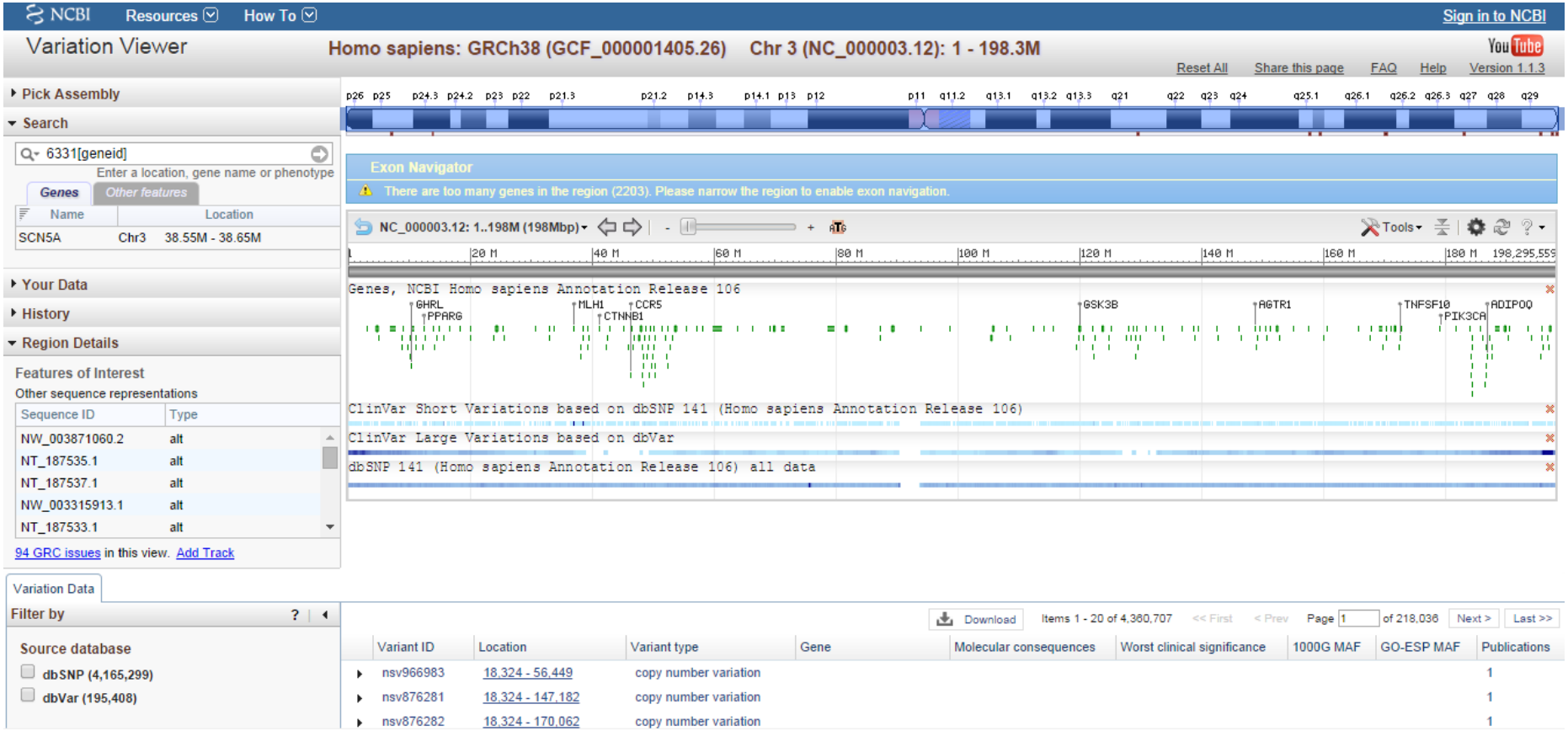
Region	Chr. position	mRNA pos	dbSNP rs# cluster id	Hetero-zygosity	Validation	MAF	Allele origin	3D	Linkout	Function	dbSNP allele	Protein residue	Codon pos	Amino acid pos	PubMed
	38550365	6198	rs376697724	N.D.						missense	A	Asn [N]	1	2002	
										contig reference	G	Asp [D]	1	2002	
	38550401	6162	rs371308670	N.D.						missense	T	Trp [W]	1	1990	
										contig reference	C	Arg [R]	1	1990	
	38550430	6133	rs76759236	0.500						missense	G	Ser [S]	2	1980	
										contig reference	C	Thr [T]	2	1980	
	38550522	6041	rs367778922	N.D.						synonymous	T	Tyr [Y]	3	1949	
										contig reference	C	Tyr [Y]	3	1949	
	38550523	6040	rs375614054	N.D.						missense	G	Cys [C]	2	1949	
										contig reference	A	Tyr [Y]	2	1949	
	38550528	6035	rs13324293	0.102		0.0542				synonymous	T	Ile [I]	3	1947	
										contig reference	C	Ile [I]	3	1947	
	38550530	6033	rs62241186	0.500						missense	G	Val [V]	1	1947	
										contig reference	A	Ile [I]	1	1947	
	38550564	5999	rs372582841	N.D.						synonymous	T	Leu [L]	3	1935	
										contig reference	C	Leu [L]	3	1935	
	38550570	5993	rs375254452	N.D.						synonymous	T	Ser [S]	3	1933	
										contig reference	C	Ser [S]	3	1933	
	38550576	5987	rs200594132	0.001		0.0005				synonymous	A	Ala [A]	3	1931	

Legend: Validation - Google Chrome

www.ncbi.nlm.nih.gov/SNP/snp_legend.cgi?lege

Validation status description

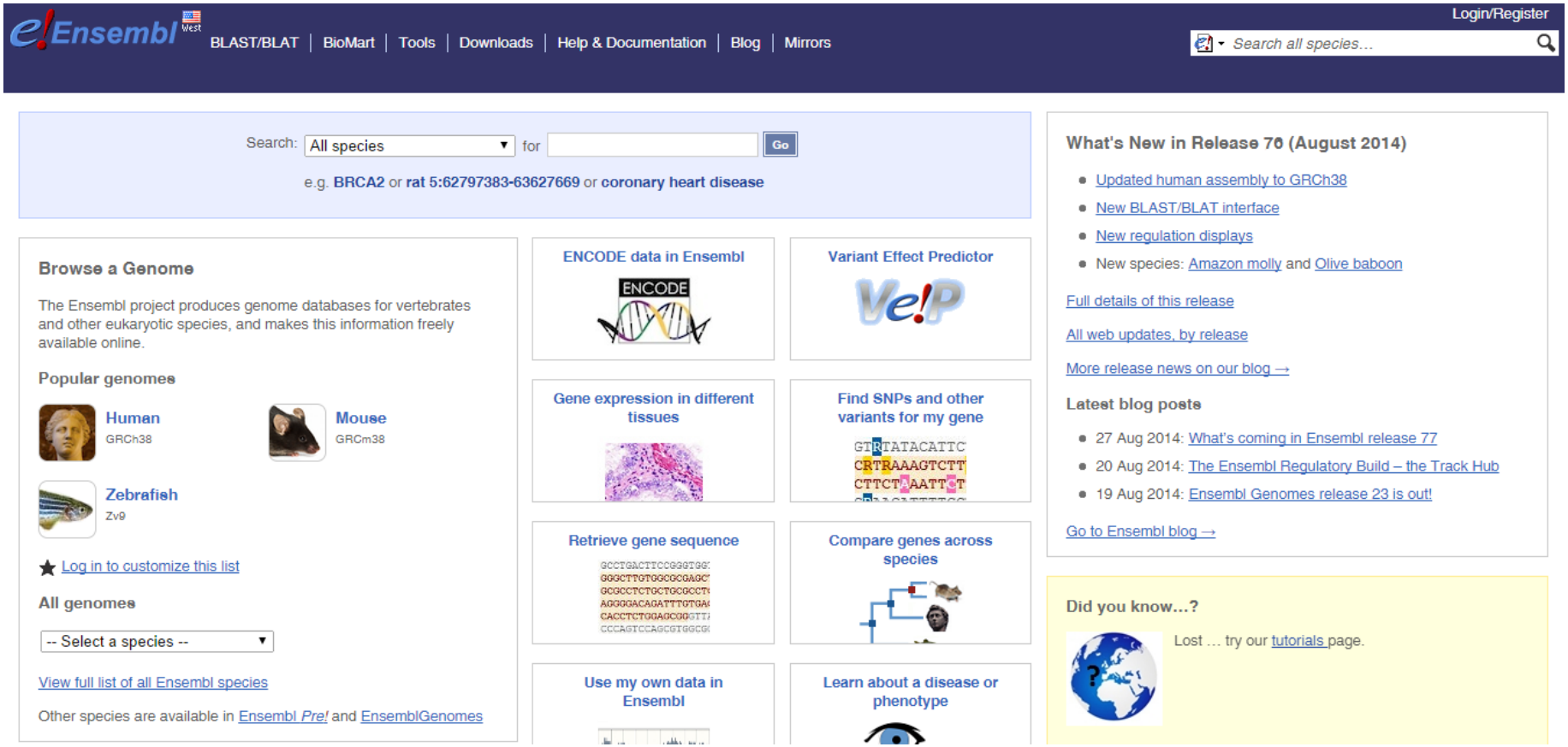
- Validated by multiple, independent submissions to the refSNP cluster
- Validated by frequency or genotype data: minor alleles observed in at least two chromosomes.
- Validated by submitter confirmation
- All alleles have been observed in at least two chromosomes apiece
- Genotyped by HapMap project
- SNP has been sequenced in 1000Genome project.
- Suspect SNPs: snp suspected from paralogous region (PMID: 21030649). Added to dbSNP on 01/21/2011.


NCBI
Resources
How To
Sign in to NCBI
Variation Viewer
Homo sapiens: GRCh38 (GCF_000001405.26) Chr 3 (NC_000003.12): 1 - 198.3M
YouTube
Reset All
Share this page
FAQ
Help
Version 1.1.3
Pick Assembly
Search
6331[geneid]
Enter a location, gene name or phenotype
Genes
Other features
Name
Location
SCN5A
Chr3
38.55M - 38.65M
Your Data
History
Region Details
Features of Interest
Other sequence representations
Sequence ID
Type
NW_003871060.2
alt
NT_187535.1
alt
NT_187537.1
alt
NW_003315913.1
alt
NT_187533.1
alt
94 GRC issues in this view.
Add Track
Exon Navigator
There are too many genes in the region (2203). Please narrow the region to enable exon navigation.
NC_000003.12: 1..198M (198Mbp)
Tools
Genes, NCBI Homo sapiens Annotation Release 106
GHRL
PPARG
MLH1
CTNNB1
CCR5
GSK3B
AGTR1
TNFSF10
PIK3CA
ADIPOQ
ClinVar Short Variations based on dbSNP 141 (Homo sapiens Annotation Release 106)
ClinVar Large Variations based on dbVar
dbSNP 141 (Homo sapiens Annotation Release 106) all data
Variation Data
Filter by
Source database
dbSNP (4,165,299)
dbVar (195,408)
Download
Items 1 - 20 of 4,360,707
Page 1 of 218,036
Next >
Last >>
Variant ID
Location
Variant type
Gene
Molecular consequences
Worst clinical significance
1000G MAF
GO-ESP MAF
Publications
nsv966983
18,324 - 56,449
copy number variation
1
nsv876281
18,324 - 147,182
copy number variation
1
nsv876282
18,324 - 170,062
copy number variation
1

Ensembl

- joint scientific project between the European Bioinformatics Institute and the Wellcome Trust Sanger Institute
- launched in 1999
- centralized resource for geneticists, molecular biologists and other researchers studying the genomes of our own species and other vertebrates and model organisms

http://www.ensembl.org



http://www.ensembl.org

Configure this page

Add your data

Gene: SCN5A ENSG00000183873

Description	sodium channel, voltage-gated, type V, alpha subunit [Source:HGNC Symbol;Acc:HGNC:10593]
Synonyms	CDCD2, CMD1E, CMPD2, HB1, HB2, HBBD, HH1, ICCD, IVF, LQT3, Nav1.5, PFHB1, SSS1
Location	Chromosome 3: 38,548,057-38,649,673 reverse strand.
INSDC coordinates	chromosome:GRCh38:CM000665.2:38548057:38649673:1
Transcripts	This gene has 15 transcripts (splice variants) Show transcript table

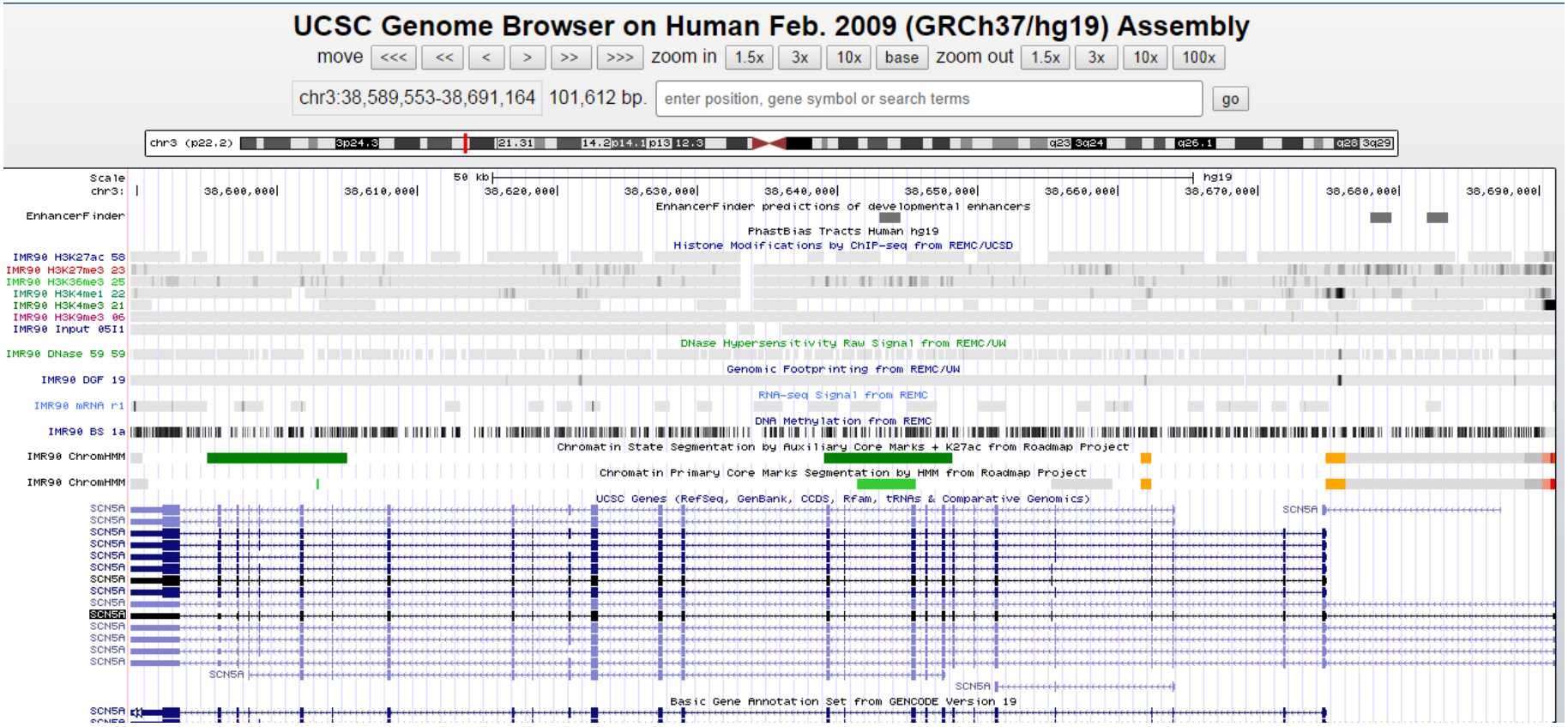
Summary

Name	SCN5A (HGNC Symbol)
CCDS	This gene is a member of the Human CCDS set: CCDS46796, CCDS46797, CCDS46798, CCDS46799, CCDS54569, CCDS54570
UniprotKB	This gene has proteins that correspond to the following Uniprot identifiers: Q14524
RefSeq	Overlapping RefSeq Gene ID 6331 matches and has similar biotype of protein_coding
LRG	LRG_289 provides a stable genomic reference framework for describing sequence variations for this gene
Ensembl version	ENSG00000183873.12
GRCh37 assembly	This gene maps to 38,589,548-38,691,164 in GRCh37 coordinates. View this locus in the GRCh37 archive: ENSG00000183873
Gene type	Known protein coding
Prediction Method	Annotation for this gene includes both automatic annotation from Ensembl and Havana manual curation, see article.
Alternative genes	**This gene corresponds to the following database identifiers:** **Havana gene:** OTTHUMG00000156166

UCSC Genome Browser

- on-line genome browser hosted by the University of California, Santa Cruz (UCSC)
- interactive website offering access to genome sequence data from a variety of vertebrate and invertebrate species and major model organisms
- integrated with a large collection of aligned annotations
- graphical viewer optimized to support fast interactive performance and is an open-source, web-based tool suite built on top of a MySQL database for rapid visualization, examination, and querying of the data at many levels

http://genome.ucsc.edu/


UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly
move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x 100x
chr3:38,589,553-38,691,164 101,612 bp.
enter position, gene symbol or search terms
go
chr3 (p22.2)
Scale
chr3:
50 kb
hg19
38,600,000
38,610,000
38,620,000
38,630,000
38,640,000
38,650,000
38,660,000
38,670,000
38,680,000
38,690,000
EnhancerFinder
EnhancerFinder predictions of developmental enhancers
PhastBias Tracts Human hg19
Histone Modifications by ChIP-seq from REMC/UCSD
IMR90 H3K27ac 58
IMR90 H3K27me3 23
IMR90 H3K36me3 25
IMR90 H3K4me1 22
IMR90 H3K4me3 21
IMR90 H3K9me3 06
IMR90 Input 05I1
DNase Hypersensitivity Raw Signal from REMC/UW
IMR90 DNase 59 59
Genomic Footprinting from REMC/UW
IMR90 DGF 19
RNA-seq Signal from REMC
IMR90 mRNA r1
DNA Methylation from REMC
IMR90 BS 1a
Chromatin State Segmentation by Auxiliary Core Marks + K27ac from Roadmap Project
IMR90 ChromHMM
Chromatin Primary Core Marks Segmentation by HMM from Roadmap Project
IMR90 ChromHMM
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)
SCN5A
Basic Gene Annotation Set from GENCODE Version 19

GeneCards

- searchable, integrated database of human genes
- provides comprehensive, updated, and user-friendly information
- all known and predicted human genes
- extracts and integrates gene-related data:
  - Genomic
  - Transcriptomic
  - Proteomic
  - Genetic
  - Clinical
  - functional information
- Automatically mined from >100 carefully selected web sources
- Allowing one-stop access to a very broad information base

http://www.genecards.org/

GeneCards


www.genecards.org/cgi-bin/carddisp.pl?gene=SCN5A
Ritchie Lab Blogs
Ritchie Lab Wiki
Ritchie Lab Forum
PASS Explorer
PSU-Websites
Interesting-Websites
GeneCards®
The Human Gene Compendium
מכון ויצמן למדע
WEIZMANN INSTITUTE OF SCIENCE
LifeMap SCIENCES
Guest
Free for academic non-profit institutions. ALL other users need a Commercial license from LifeMap Sciences, Inc.
Home
GeneCards Guide
Suite
Terms of Use
About Us
User Feedback
GeneCards Plus
Set Analyses:
GeneALaCart
GeneAnalytics
VarElect
keyword(s)
Search
Advanced Search
SCN5A Gene
protein-coding GIFtS: 67
GCID: GC03M038589
Sodium Channel, Voltage-Gated, Type V, Alpha Subunit
(Previous names: sodium channel, voltage-gated, type V, alpha (long QT syndrome...)
(Previous symbol: CMD1E)
Proteins & Enzymes
Antibodies
Assays & Kits
Interaction networks and pathway maps
QIAGEN
See SCN5A-related diseases at MalaCards
ORIGENE
Proteins
Antibodies
Assays/ Genes/ shRNA/ Primers
GenScript
Make Research Easy
Genes
Peptides
Proteins
Jump to Section...
Aliases
for SCN5A gene
(According to 1HGNC, 2Entrez Gene, 3UniProtKB/Swiss-Prot, 4UniProtKB/TrEMBL, 5OMIM, 6GeneLoc, 7Ensembl, 8DME, 9miRBase, 10fRNAdb, 12H-InvDB, 13NCBI, 14NONCODE, and/or 15RNAdb)
About This Section
Aliases
Sodium Channel, Voltage-Gated, Type V, Alpha Subunit1 2
CMD1E1 2 5
Sodium Channel Protein Cardiac Muscle Subunit Alpha2 3
Voltage-Gated Sodium Channel Subunit Alpha Nav1.52 3
HH12 3
CDCD22 5
HB12 5
LQT32 5
Sodium Channel, Voltage-Gated, Type V, Alpha (Long QT Syndrome 3)1
CMPD22
HB22
HBBD2
ICCD2
IVF2
Nav1.52
PFHB12
GeneAnalytics ™
POWERFUL GENE SET ANALYSIS
Cells, Diseases, Pathways, Functions and Compounds relevant to your gene set powered by GeneCards
SIGN UP FREE >

If the focus is primarily SNPs….

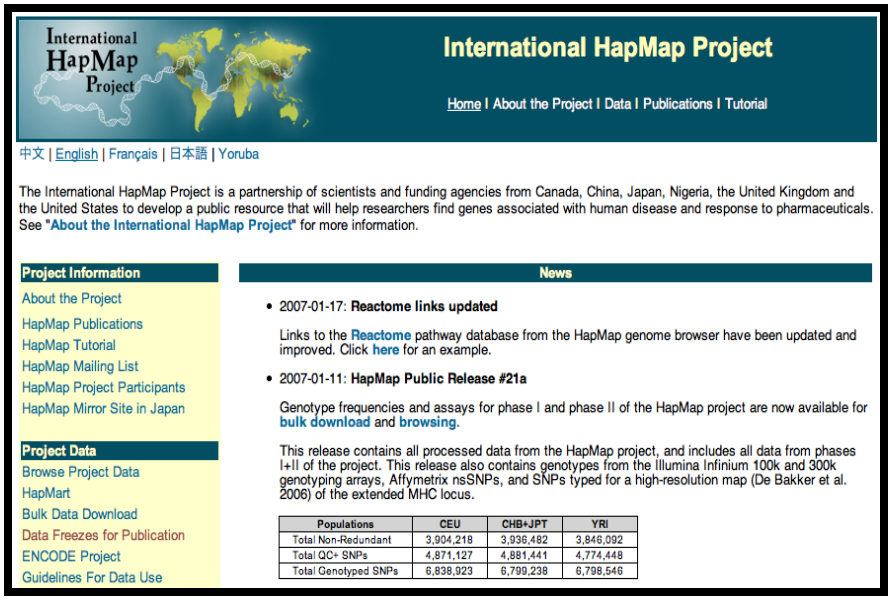
HapMap Project: Create a genome-wide SNP map

Genotype SNPs in four populations:

- CEPH (CEU) (Europe - n = 90, trios)
- Yoruban (YRI) (Africa - n = 90, trios)
- Japanese (JPT) (Asian - n = 45)
- Chinese (HCB) (Asian - n =45)

To produce a genome-wide map of common variation

Common Variant/Common Disease



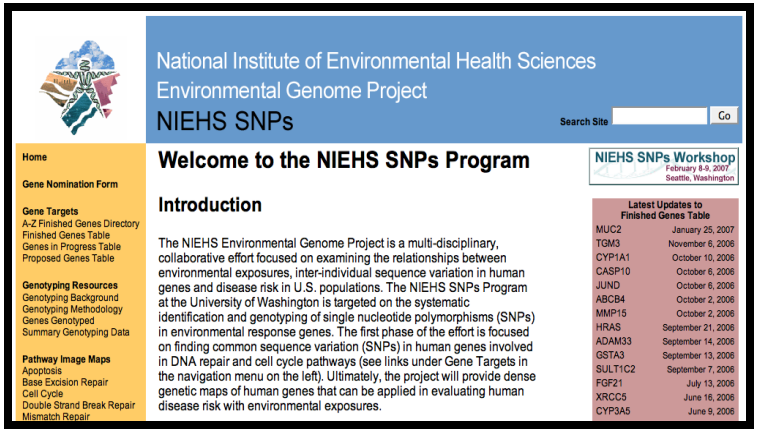
Low density - genome-wide

Phase I - 1M SNPs

Phase II - 4M SNPs

Density ~ 1 SNP/kb

High density - candidate gene



643 genes - 15 Mbp

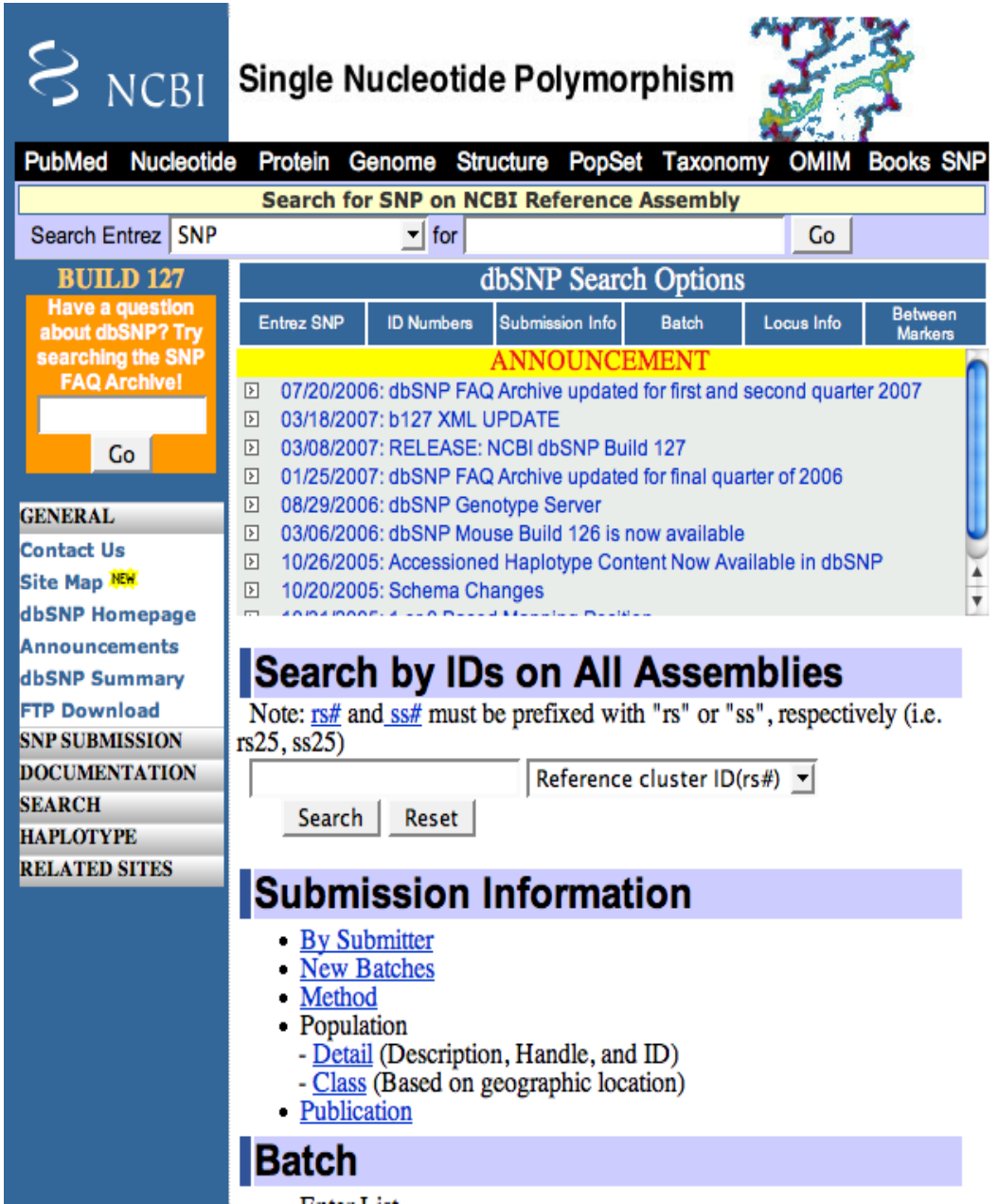
92,300 SNPs – 1 SNP/166 bp

322 genes - 7 Mbp

37,450 SNPs – 1 SNP/186 bp

SNP Discovery: dbSNP database

dbSNP
-NCBI SNP database



SNP data submitted to dbSNP: Clustering

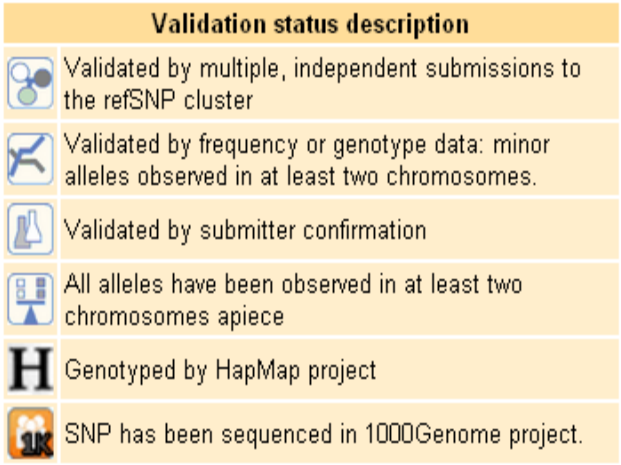
dbSNP processing of SNPs

SNPs submitted
by research community
(submitted SNPs = ss#)

Unique mapping
to a genome location
(reference SNP = rs#)

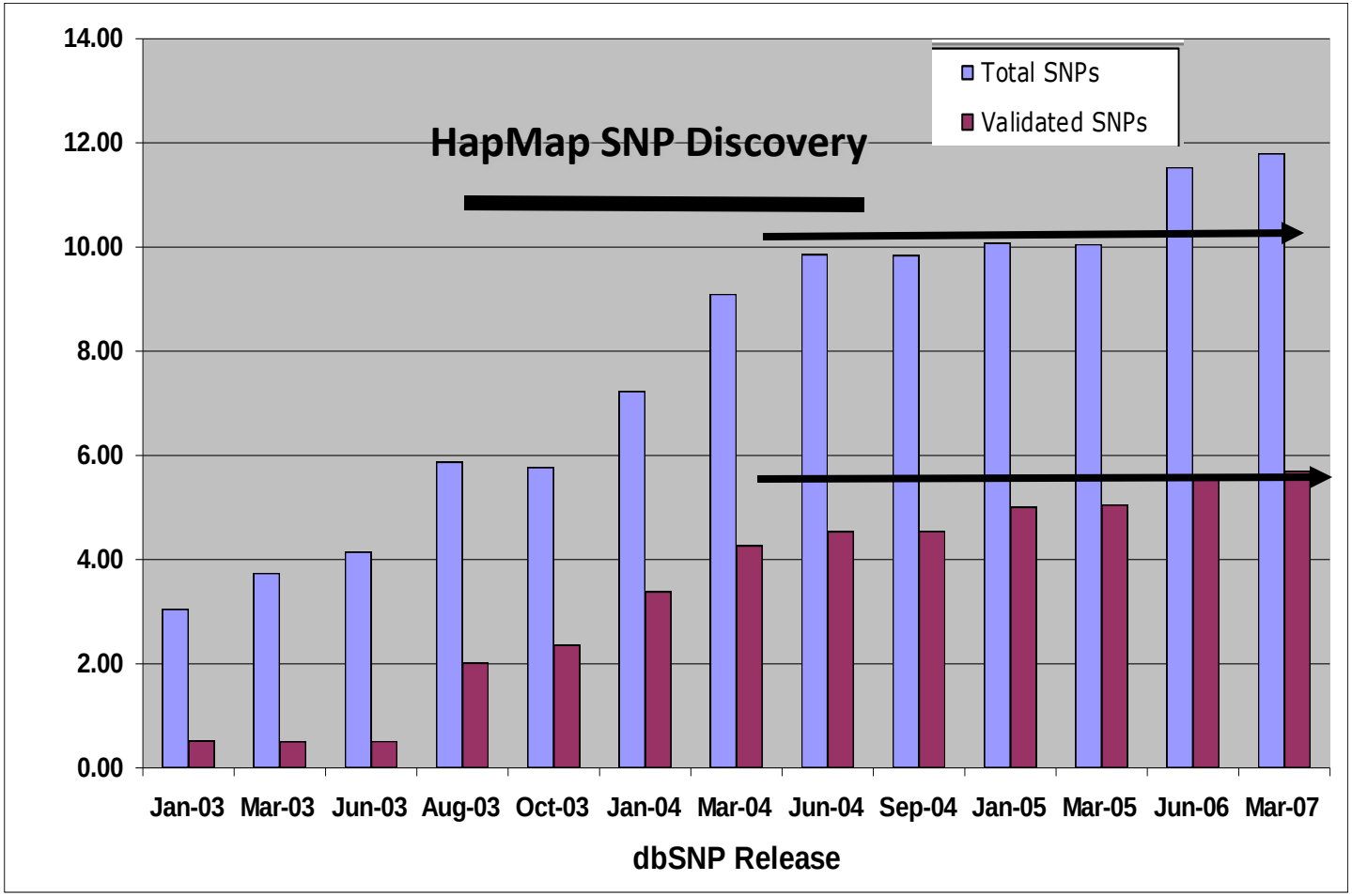
Validated

Unvalidated



Validation status description
Validated by multiple, independent submissions to the refSNP cluster
Validated by frequency or genotype data: minor alleles observed in at least two chromosomes.
Validated by submitter confirmation
All alleles have been observed in at least two chromosomes apiece
Genotyped by HapMap project
SNP has been sequenced in 1000Genome project.

HapMap Discovery Increased SNP Density and Validated SNPs



14+ million rs SNPs

6.5 million validated rs SNPs

rs #'s are THE nomenclature for SNPs

Table 1 Association between SNPs in the chromosome 20 locus and AGA in the German sample

				MAF[a]		Genotypes[b]			
SNP (position)[c]	Sample	Cases[d]	Controls[e]	Cases[d]	Controls[e]	Cases[d]	Controls[e]	*P*	OR (95% CI)[f]
rs6137444	GWAS	296	347	0.264 (C)	0.383 (C)	14/128/154	49/168/130	3.11×10^{-6}	1.74 (1.37–2.21)
(21,733,639 bp)	Replication	319	234	0.277 (C)	0.404 (C)	21/135/163	45/99/90	1.57×10^{-5}	1.76 (1.37–2.27)
	Combined[g]	605	579	0.269 (C)	0.391 (C)	35/255/315	93/267/219	2.20×10^{-10}	
rs2180439	GWAS	296	347	0.292 (C)	0.429 (C)	21/131/144	66/166/115	3.85×10^{-7}	1.82 (1.45–2.30)
(21,801,100 bp)	Replication	319	234	0.303 (C)	0.485 (C)	23/147/149	62/103/69	1.37×10^{-9}	2.17 (1.70–2.78)
	Combined[g]	605	579	0.293 (C)	0.452 (C)	43/268/294	127/269/183	2.67×10^{-15}	
rs1998076	GWAS	296	347	0.282 (A)	0.427 (A)	20/120/144	65/163/115	1.30×10^{-7}	1.90 (1.50–2.41)
(21,828,045 bp)	Replication	319	234	0.301 (A)	0.479 (A)	23/146/150	61/102/71	3.69×10^{-9}	2.13 (1.66–2.73)
	Combined[g]	605	579	0.292 (A)	0.448 (A)	43/267/295	126/267/186	7.73×10^{-15}	
rs201571	GWAS	296	347	0.289 (C)	0.411 (C)	17/137/142	61/163/123	4.31×10^{-6}	1.72 (1.36–2.17)
(21,961,514 bp)	Replication	319	234	0.314 (C)	0.483 (C)	30/140/149	58/110/66	2.21×10^{-8}	2.05 (1.60–2.62)
	Combined[g]	605	579	0.298 (C)	0.44 (C)	46/269/290	119/272/188	1.21×10^{-12}	
rs6113491	GWAS	296	347	0.359 (C)	0.483 (C)	29/154/112	88/159/100	8.63×10^{-6}	1.66 (1.33–2.08)
(22,005,415 bp)	Replication	319	234	0.364 (C)	0.447 (A)	38/156/125	77/105/52	8.13×10^{-10}	2.17 (1.70–2.77)
	Combined[g]	605	579	0.359 (C)	0.488 (A)	66/302/237	165/263/151	1.13×10^{-13}	

Hillmer et al (2008) *Nat Genet* 40:1279-81

Increasing SNP Density: HapMap ENCODE Project

ENCODE = **ENC**yclopedia **O**f **D**NA **E**lements

Catalog all functional elements in 1% of the genome (30 Mb)

10 Regions x 500 kb/region (Pilot Project)

David Altschuler (Broad), Richard Gibbs (Baylor)

16 CEU, 16 YRI, 8 HCB, 8 JPT

Comprehensive PCR based resequencing across these regions

ENCODE Project Information

Resequencing Project

Genotyping Project

Perlegen Genotyping Component

ENCODE Links

ENCODE genotype data dumps

About the ENCODE Project

ENCODE Regions Genotype Information

Region name	Chromosome band	Genomic interval (NCBI)	Available SNPs		CEU		
			dbSNP	New SNPs	rs#	no rs#	o rs#
ENr112	2p16.3	Chr2:51633239..52133238	1,624	1,720	,064	93	922
ENr131	2q37.1	Chr2:234778639..235278638	1,787	1,233	1,179	71	704
ENr113	4q26	Chr4:118705475..119205474	1,516	1,819	1,017	1,61	,597
ENm010	7p15.2	Chr7:26699793..27199792	1,274	1,857	757	45	456
ENm013	7q21.13	Chr7:89395718..89895717	1,545	1,713	927	1,38	,391
ENm014	7q31.33	Chr7:126135436..126632577	1,354	1,562	963	1,42	,419
ENr321	8q24.11	Chr8:118769628..119269627	1,468	1,682	936	90	903
ENr232	9q34.11	Chr9:127061347..127561346	1,494	1,646	694	70	689
ENr123	12q12	Chr12:38626477..39126476	1,904	1,551	859		0
ENr213	18q12.1	Chr18:23717221..24217220	1,391	1,465	809	82	819
		Total	**15,357**	**16,248**	**9,205**	**8,97**	**,900**

15,357 dbSNP

16,248 New SNPs

50% of SNPs in dbSNP

5 Mb/31,500 SNPs = 1/160 bp

Population descriptors:

CEU: CEPH (Utah residents with ancestry from northern and western Europe)

HCB: Han Chinese in Beijing, China

JPT: Japanese in Tokyo, Japan

YRI: Yoruba in Ibadan, Nigeria

National Institute of Environmental Health Sciences
Environmental Genome Project
NIEHS SNPs

Search Site Go

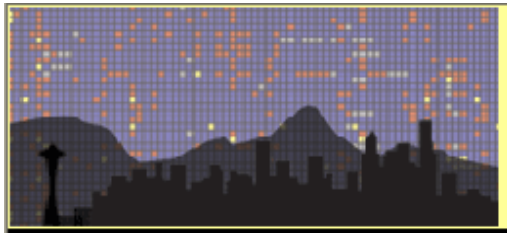
Goal:
Comprehensively identify all common sequence variation in candidate genes

Initial biological focus:
Candidate environmental response genes involved in DNA repair, cell cycle, apoptosis, metabolism, cell signaling, and oxidative stress.

Approach:
Direct resequencing of genes

Samples:
PDR-90 ethnically diverse individuals representative of U.S. population (397 genes)
EGP95-95 samples from four ethnic groups (227 genes)
(24 HapMap Asians, 22 HapMap Europeans, 12 HapMap Yorubans, 15 African Americans, 22 Hispanics

Website:
egp.gs.washingon.edu



SeattleSNPs

Variation Discovery Resource

Goal:
Comprehensively identify all common sequence variation in candidate genes

Initial biological focus:
Candidate environmental response genes involved in lipid metabolism, inflammation, and blood pressure regulation.

Approach:
Direct resequencing of genes

Samples:
P1: 23 CEPHs and 24 African-American (overlaps with Perlegen)
P2: 23 CEPHs and 24 Yorubans (overlaps with HapMap)

Website:
pga.gs.washington.edu

Summary of SeattleSNPs and NIEHS SNPs genotypes in dbSNP

Table 1. Summary of genotype data contained in dbSNP

Data set	Genotypes	SNPs	Populations	Individuals	Average SNP density	Reference
HAPMAP	159,862,776	954,302	4	270	3149	(International HapMap Consortium 2003)
PERLEGEN	110,385,051	1,576,578	3	71	1938	(Hinds et al. 2005)
Affymetrix	6,189,466	125,778	6	116	24,029	(Kennedy et al. 2003)
TSC	4,932,382	19,048	17	1963	312,754	(International SNP Map Working Group 2001)
EGP	3,184,170	37,737	1	90	72,443	(Livingston et al. 2004)
PGA/UW	573,194	15,981	2	47	153,861	(Crawford et al. 2004)
IIPGA	176,162	3801	3	47	430,361	(Innate Immunity PGA, http://innateimmunity.net/)
NIHPDR	159,549	1982	1[a]	448	1,419,125	(Collins et al. 1998)
WICVAR	33,240	1462	1	130	2,011,277	
HG_BONN	24,522	320	1	143	5,284,550	(Freudenberg-Hua et al. 2003)

[a]The NIHPDR data contains a single mixed population.

643 genes sequenced (NIEHS SNPs)

15 Mb scanned

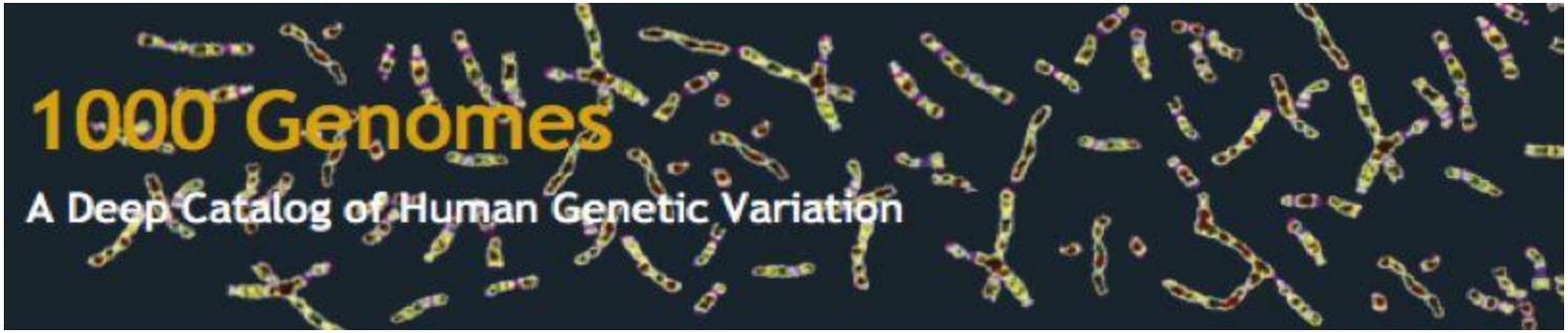
> 92,000 genotyped SNPs identified

> 8 million genotypes deposited in dbSNP

Nov 2005 - Zaitlen et al. Genome Research 15:1594-1600

Summary: The Current State of SNP Resources

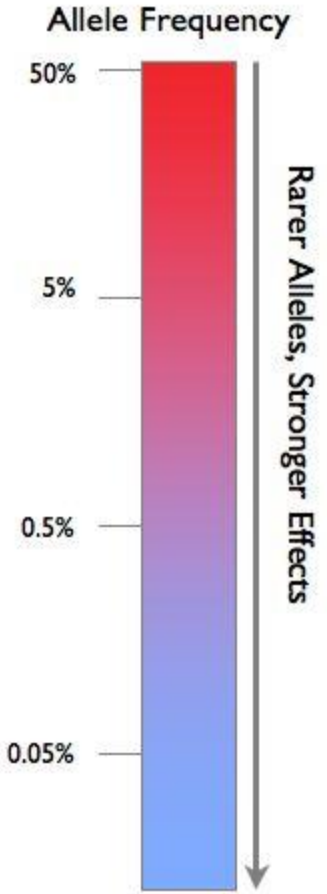
- Approximately 10 million common SNPs exist in the human genome (1/300 bp).
- Random SNP discovery processes generate many SNPs (HapMap)
- Random approaches to SNP discovery have reached limits of discovery and validation (1/600 bp; 50% SNP validation)
- Most validated SNPs (6+ million) have been genotyped by the HapMap (3 pops)
- Resequencing approaches continue to catalog important variants (rare and common not captured by the HapMap)



1000 genomes project: motivation

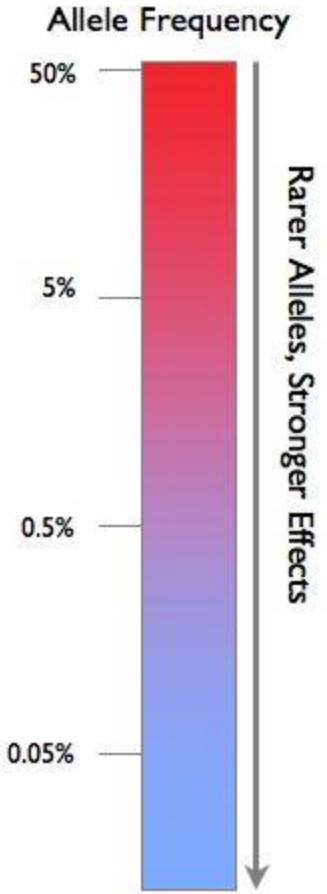
- GWAS shows that systematic association studies can be used to map disease genes
- The first generation of GWAS was well powered only for SNPs with > 5% MAF
- Next generation sequencing now makes it possible to create a complete catalogue of human polymorphism for SNPs and CNVs

Exploring the full range of genetic variants



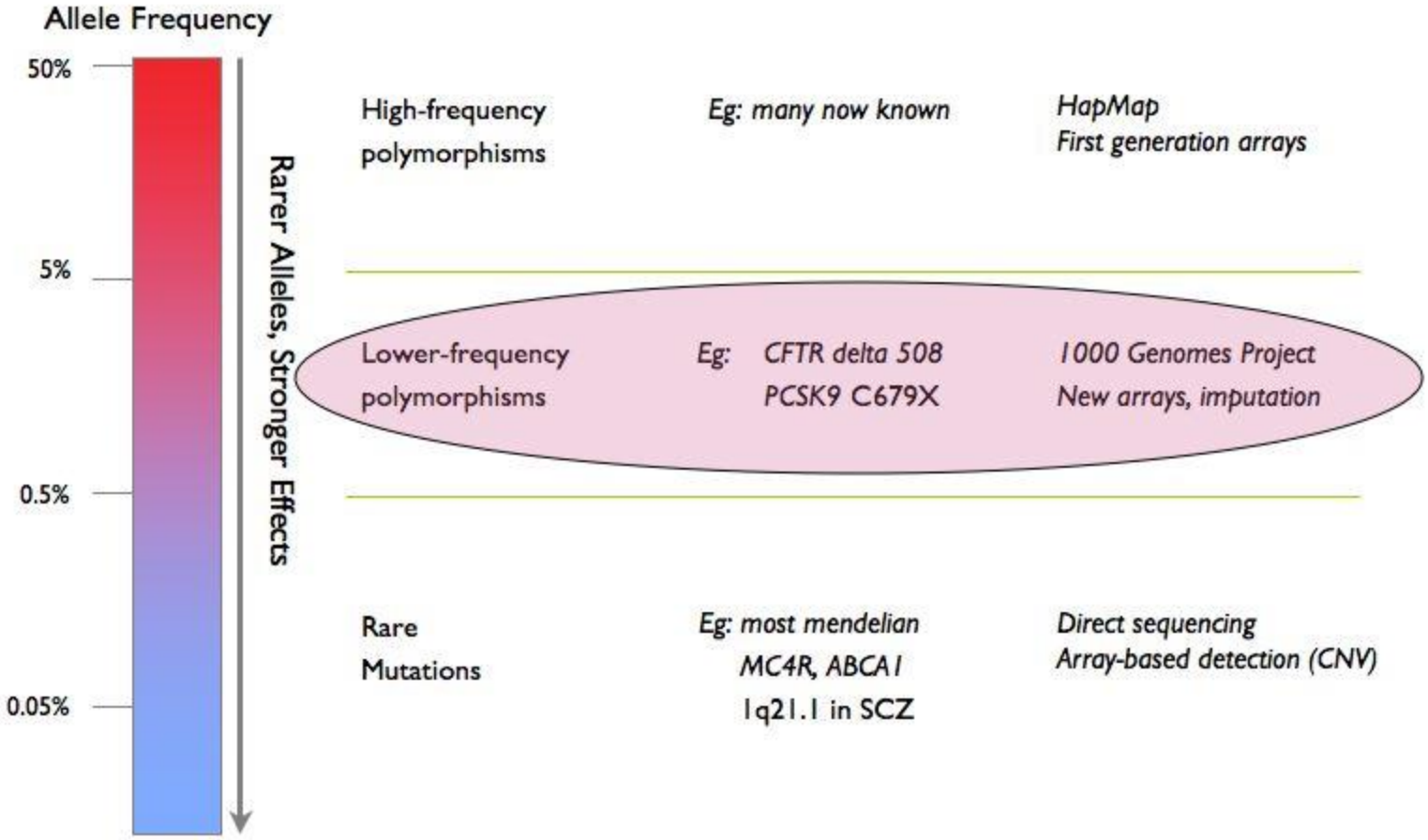
High-frequency polymorphisms	*Eg: many now known*	*HapMap* *First generation arrays*
Rare Mutations	*Eg: most mendelian* *MC4R, ABCA1* 1q21.1 in SCZ	*Direct sequencing* *Array-based detection (CNV)*

Exploring the full range of genetic variants



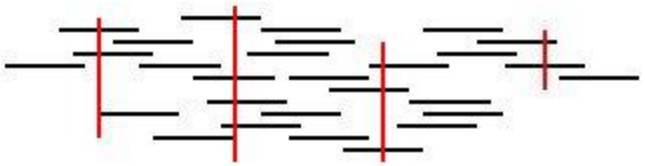
High-frequency polymorphisms	*Eg: many now known*	*HapMap* *First generation arrays*
Lower-frequency polymorphisms	*Eg: CFTR delta 508* *PCSK9 C679X*	*1000 Genomes Project* *New arrays, imputation*
Rare Mutations	*Eg: most mendelian* *MC4R, ABCA1* 1q21.1 in SCZ	*Direct sequencing* *Array-based detection (CNV)*

Exploring the full range of genetic variants



1000 Genomes Project

Random Coverage
0.2 to 0.4 X depth



Produce a catalog of variants across the genome

in multiple populations with allele frequencies ≥ 1%

Samples and ELSI Group

Leena Peltonen (co-chair) Sanger Institute
Bartha Knoppers (co-chair) University of Montreal
Aravinda Chakravarti (co-chair) Johns Hopkins
Gonçalo Abecasis University of Michigan
Richard Gibbs Baylor College of Medicine
Lynn Jorde University of Utah
Eric Juengst Case Western Reserve University
Jane Kaye Oxford University
Alastair Kent Genetic Interest Group
Rick Kittles University of Chicago
Jim Mullikin National Human Genome Research Institute
Mike Province Washington University in St. Louis
Charles Rotimi Howard University
Yeyang Su Beijing Genomics Institute
Chris Tyler-Smith Sanger Institute
Ling Yang Beijing Genomics Institute

Data Flow Group (being formed)

Paul Flicek (co-chair) European Bioinform[illegible]
Stephen Sherry (co-chair) National Center [illegible]
Ewan Birney European Bioinformatics Instit[illegible]
Clive Brown Sanger Institute
David Dooli[illegible] Washington University in St[illegible]
Richard Gibb[illegible] [illegible]
Sol Katzman [illegible]
Hoda Khouri [illegible] Information
Martin Shumway National Center for Biotechnology Information
Jun Wang Beijing Genomics Institute
George Weinstock Baylor College of Medicine
(Broad representative)

Steering Committee

Richard Durbin (co-chair) Sanger Institute
David Altshuler (co-chair) Broad / MGH / Harvard
Gonçalo Abecasis University of Michigan
Aravinda Chakravarti Johns Hopkins
Andrew Clark Cornell University
Francis Collins National Human Genome Research Institute
Peter Donnelly Oxford University
Paul Flicek European Bioinformatics Institute
Stacey Gabriel Broad Institute
Richard Gibbs Baylor College of Medicine
Bartha Knoppers University of Montreal
Eric Lander Broad Institute
Elaine Mardis Washington University in St. Louis
Gil McVean Oxford University
Debbie Nickerson University of Washington
Leena Peltonen Sanger Institute
Stephen Sherry National Center for Biotechnology Information
Rick Wilson Washington University in St. Louis
Huanming (Henry) Yang Beijing Genomics Institute

Production Group

Elaine Mardis (co-chair) Washington University in St. Louis
Stacey Gabriel (co-chair) Broad Institute
Richard Durbin Sanger Institute
Richard Gibbs Baylor College of Medicine
David Jaffe Broad Institute
Ruiqiang Li Beijing Genomics Institute
Donna Muzny Baylor College of Medicine
Chad Nusbaum Broad Institute
Aarno Palotie Sanger Institute
Dan Turner Sanger Institute
Jun W[illegible] [illegible]
Wei Wa[illegible] [illegible]
Rick Wi[illegible] [illegible]

Funders

Alan Schafer Wellcome Trust
Francis Collins National Human Genome Research Institute
Lisa Brooks National Human Genome Research Institute
Audrey Duncanson Wellcome Trust
Adam Felsenfeld National Human Genome Research Institute
Mark Guyer National Human Genome Research Institute
Ruth Jamieson Wellcome Trust
[illegible]
[illegible] **Li** [illegible]
[illegible]
Jane Peterson National Human Genome Research Institute
Anne Pierson National Human Genome Research Institute
Zhiwu Ren National Planning and Development Committee
Jian Wang Beijing Genomics Institute

Analysis Group

Gil McVean (co-chair) Oxford University
Gonçalo Abecasis (co-chair) University of Michigan
David Altshuler Broad / MGH / Harvard
Paul de Bakker Broad / BWH / Harvard
Brian Browning University of Auckland
Sharon Browning University of Auckland
Carlos Bustamante Cornell University
David Carter Sanger Institute
Aravinda Chakravarti Johns Hopkins
Andrew Clark Cornell University
Don Conrad Sanger Institute
Mark Daly Broad / MGH / Harvard
Manolis Dermitzakis Sanger Institute
Peter Donnelly Oxford University
Richard Durbin Sanger Institute
Evan Eichler University of Washington
Paul Flicek European Bioinformatics Institute
Bryan Howie Oxford University
Matt Hurles Sanger Institute
David Jaffe Broad Institute
Lynn Jorde University of Utah
Hoda Khouri National Center for Biotechnology Information
Eric Lander Broad Institute
Charles Lee Brigham and Women's Hospital
Guoqing Li Beijing Genomics Institute
Heng Li Sanger Institute
Ruiqiang Li Beijing Genomics Institute
Yingrui Li Beijing Genomics Institute
Yun Li University of Michigan
Jonathan Marchini Oxford University
Gabor Marth Boston College
Steve McCarroll Broad Institute
Jim Mullikin National Human Genome Research Institute
Simon Myers Oxford University
Rasmus Nielsen University of California, Berkeley
Alkes Price Broad / Harvard
Jonathan Pritchard University of Chicago
Mike Province Washington University in St Louis
Molly Przeworski University of Chicago
Shaun Purcell Broad / MGH / Harvard
Noah Rosenberg University of Michigan
Pardis Sabeti Broad / Harvard
[illegible] University of Michigan
Stev[illegible] Sch[illegible]ner Broad Institute
[illegible]**at** Cold Spring Harbor Laboratory
[illegible]**y** National Center for Biotechnology Information
Matt[illegible] S[illegible]hens University of Chicago
Simo[illegible] University of Southern California
Chris Tyler-Smith Sanger Institute
Jun Wang Beijing Genomics Institute
David Wheeler Baylor College of Medicine
Hongkun Zheng Beijing Genomics Institute

www.1000genomes.org

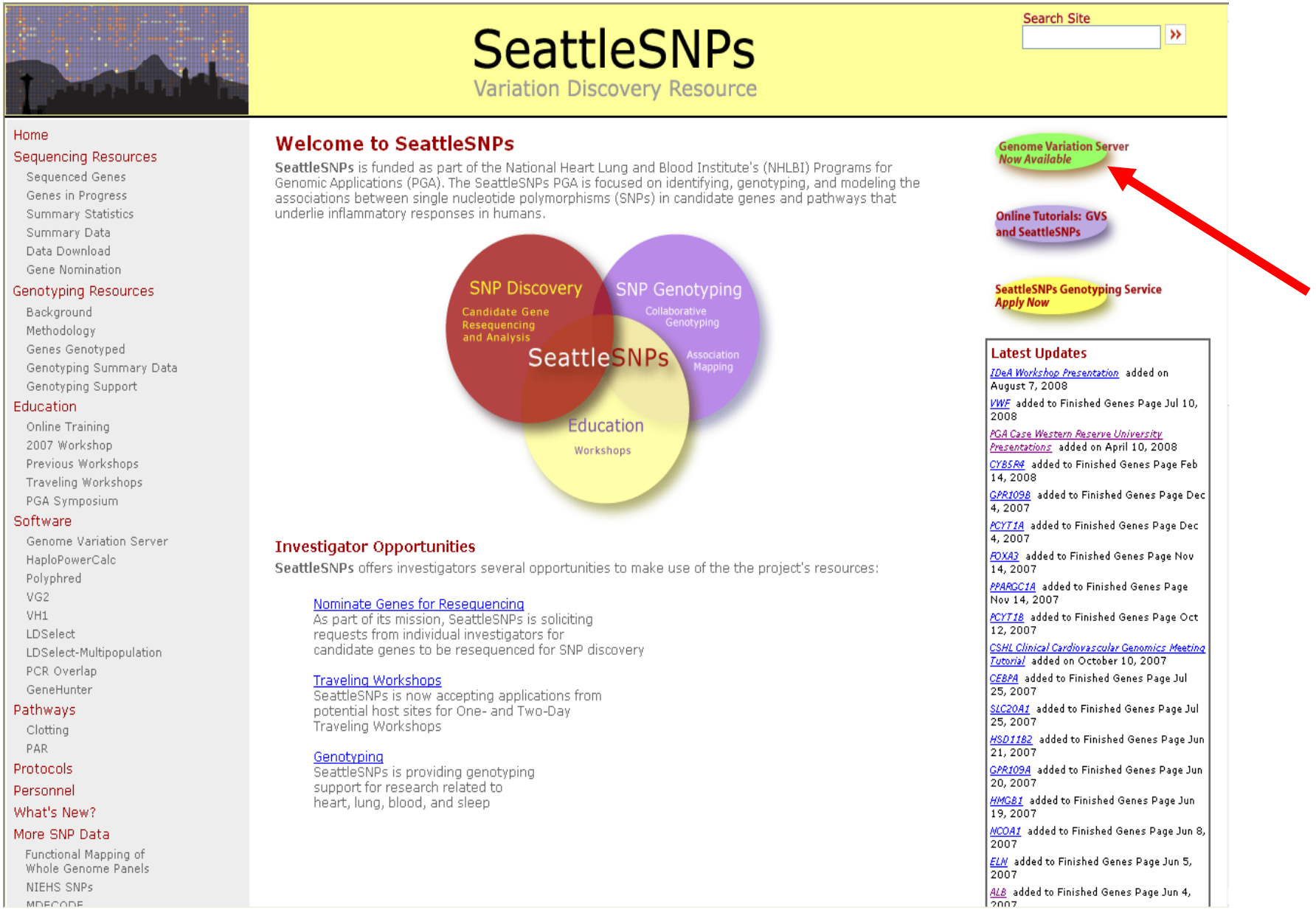
Where to find SNPs and Linkage Disequlibrium Data

For your gene or region of interest, search

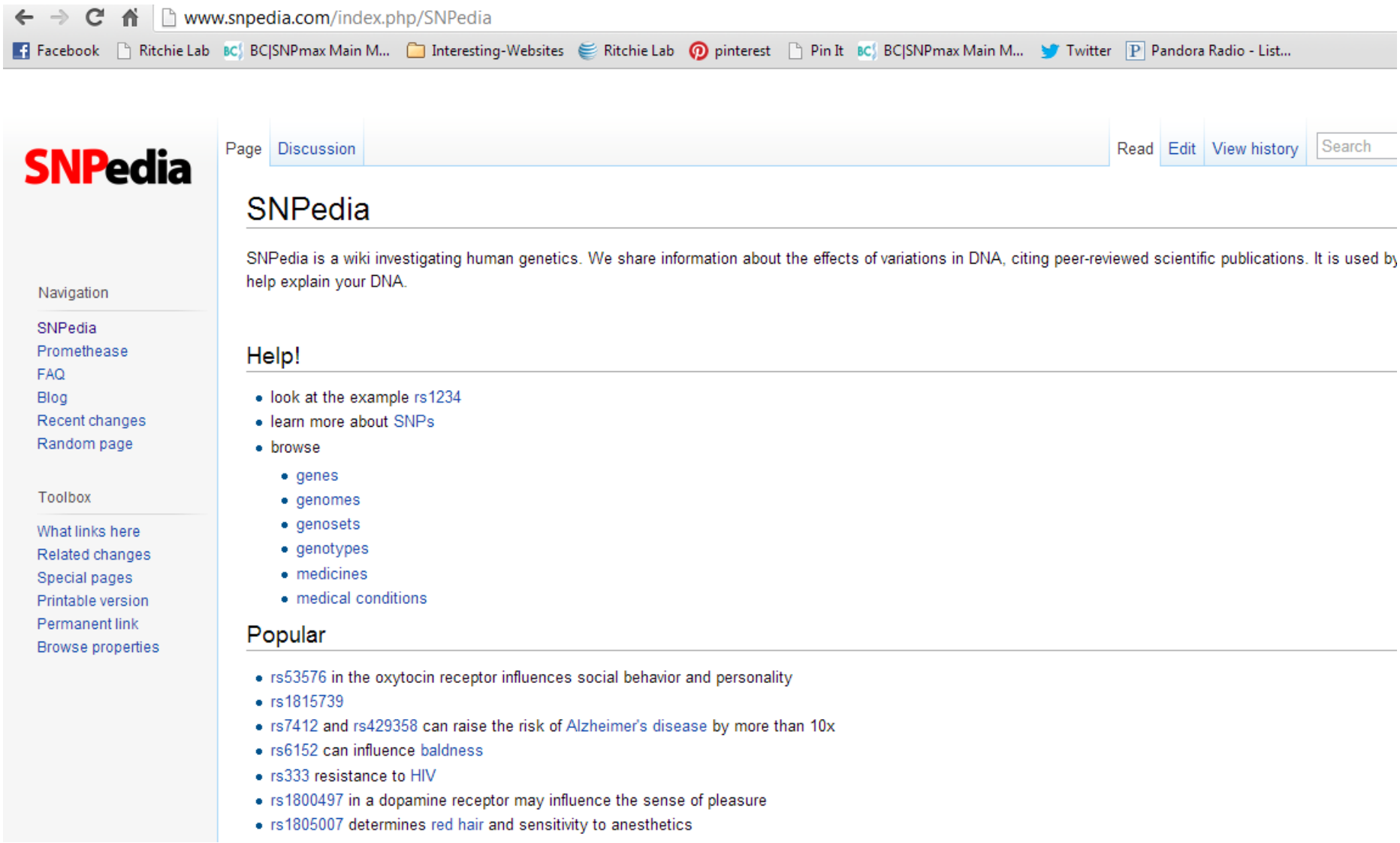
Genome Variation Server

- HapMap
  www.hapmap.org
- NIEHS SNPs
  egp.gs.washington.edu
- SeattleSNPs PGA
  pga.gs.washington.edu

Visualizing Pair-wise LD



Catalog of SNP effects



SNP-related Websites

- dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP/)
- SeattleSNPs (pga.gs.washington.edu)
- NIEHS SNPs (egp.gs.washington.edu)
- Genome Variation Server (http://gvs.gs.washington.edu/GVS/
- HapMap (www.hapmap.org)
- SNPedia (www.snpedia.com)

Assignment

- Search your favorite gene in the databases discussed today
- If you do not have a favorite gene, try one of mine:
  - *SCN5A*
  - *RYR1*
  - *CETP*
  - *PCSK9*
  - *FTO*
  - *CDKN2B*
  - *PTPN22*

Questions???