

# 1000 Genomes Browser Orientation

<http://browser.1000genomes.org>

Based on Project Phase I Data

European Bioinformatics Institute

11 May, 2011

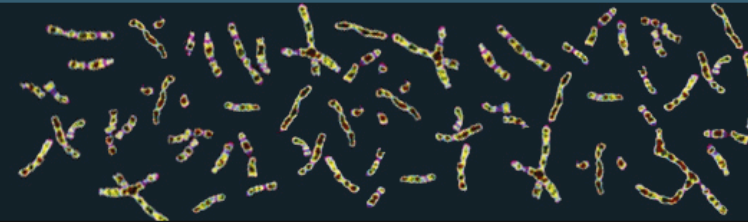
# Overview

- Based on version 62 of Ensembl code and NCBI37 version of the human genome assembly
- Contains all of the gene information normally present in Ensembl
  - Gene and transcript annotation, external references, sequence data
- Incorporates essentially all of the 1000 Genomes Phase I data
- 1000 Genomes pilot project data are available on a separate browser –  
<http://pilotbrowser.1000genomes.org>
- Please send questions to [info@1000genomes.org](mailto:info@1000genomes.org)

# 1000 Genomes Browser Guide Page

## 1000 Genomes

A Deep Catalog of Human Genetic Variation



[Home](#) [About](#) [Data](#) [Analysis](#) [Participants](#) [Contact](#) [Browser](#) [Wiki](#) [FTP search](#)

[Home](#) >

### THE ENSEMBL BROWSER

The 1000 genomes project uses a project specific version of the [Ensembl Browser](#) to visualise its variants. This browser can be found at <http://browser.1000genomes.org>.

The main browser currently displays the snps and indels from the [20100804](#) release. This runs Ensembl version 62.

There is also a version of the browser which holds the pilot data from [A map of human genome variation from population-scale sequencing](#), *Nature* 467,1061.1073 at <http://pilotbrowser.1000genomes.org>. This uses Ensembl version 60.

You can find instructions for how to use our browser [here](#) in doc format. There is also a tutorial for the pilot browser [here](#)

There are also more tips for ensembl both on the [main ensembl site](#) and their [blog](#)


### NAVIGATION

- [Frequently Asked Questions](#)

# 1000 Genomes Browser Home Page

## 1000 Genomes

A Deep Catalog of Human Genetic Variation




Tools | Help

### Search 1000 Genomes

e.g. gene BRCA2 or Chromosome 6:133098746-133108745

### Start Browsing 1000 Genomes data

[Browse Human](#) →  
GRCh37

[Protein variations](#) →  
View the consequences of sequence variation at the level of each protein in the genome.

[Individual genotypes](#) →  
Show different individual's genotype, for a variant.

### Browser update May 2011

based on the **Main project data released in November 2010 for 629 individuals. It can be found on [the ftp site](#).**

Please see [www.1000genomes.org](http://www.1000genomes.org) for more information about the data presented here and instructions for downloading the complete data set.


- [View sample data](#)


### The 1000 Genomes Browser

**Ensembl-based browser provides early access to 1000genomes data**



In order to facilitate immediate analysis of the 1000genomes data by the whole scientific community, this browser (based on Ensembl) integrates the SNP calls from [the August 2010 release](#). This data will be submitted to dbSNP, and once rsid's have been allocated, will be absorbed into the UCSC and Ensembl browsers according to their respective release cycles. Until that point **any non rs SNP id's on this site are temporary and will NOT be maintained**.

#### Links

[1000 Genomes](#) →  
More information about the 1000 Genomes Project on the 1000 genomes main site.

[Pilot browser](#) →  
This browser is based on Ensembl release 60 and represents the variant set analysed as part of [A map of human genome variation from population-scale sequencing](#), Nature 467 ,1061.1073.

The 1000 Genomes Project is an international collaborative project described at [www.1000genomes.org](http://www.1000genomes.org). The 1000 Genomes Browser is based on Ensembl web code.

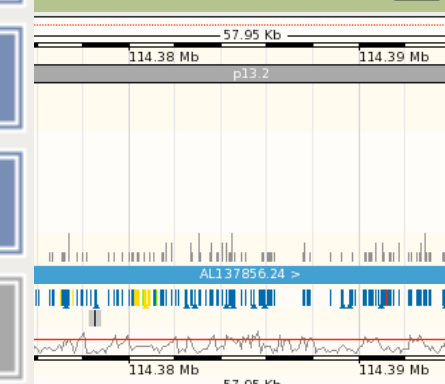
Ensembl is a joint project of EMBL-EBI  and the Wellcome Trust Sanger Institute 

1000 Genomes release 8 - May 2011 © EBI

About 1000 Genomes | Contact Us | Help

- Location-based displays
    - Whole genome
    - Chromosome summary
    - Region overview
    - Region in detail**
  - Genetic Variation
    - Resequencing (2)
    - Linkage Data
    - Markers
  - Configure this page
  - Manage your data
  - Export data
  - Bookmark this page

★ Bookmark this page



<http://browser.1000genomes.org>

# Page configurat

Configure page Main panel Top p

Main panel

Active tracks

- Favourite tracks
- Search Results
- (1/19) 1000 Genomes
- (0/1) 1000 Genomes VCF
- (1/4) Sequence
- (0/1) Markers
- (5/5) Genes
- (0/1) Prediction Transcripts
- (0/5) Protein alignments
- (0/1) Protein features
- (0/2) cDNA/mRNA alignments
- (0/2) RNA alignments
- (0/26) Probe features
- (0/4) Simple features
- (0/7) Misc. regions
- (0/18) Repeats
- (1/84) Germline variation
- (0/1) Somatic Mutations
- (1/34) Functional genomics
- (5/5) Additional decorations
- (5/7) Information

Add custom track

Reset configuration

## 1000 Genomes

☐ Enable/disable all tracks

- ☒ 1000 genomes - August 2010 variations
- ☒ 1000 genomes - August 2010 - AFR variations
- ☒ 1000 genomes - August 2010 - ASN variations
- ☒ 1000 genomes - August 2010 - EUR variations
- ☐ 1000 genomes - High coverage - Trios variations
- ☐ 1000 genomes - High coverage - Trios - CEU variations
- ☐ 1000 genomes - High coverage - Trios - YRI variations
- ☐ 1000 genomes - High coverage exons variations
- ☐ 1000 genomes - High coverage exons - CEU variations
- ☐ 1000 genomes - High coverage exons - CHB variations
- ☐ 1000 genomes - High coverage exons - CHD variations
- ☐ 1000 genomes - High coverage exons - JPT variations
- ☐ 1000 genomes - High coverage exons - LWK variations
- ☐ 1000 genomes - High coverage exons - TSI variations
- ☐ 1000 genomes - High coverage exons - YRI variations
- ☐ 1000 genomes - Low coverage variations
- ☐ 1000 genomes - Low coverage - CEU variations
- ☐ 1000 genomes - Low coverage - CHB+JPT variations
- ☐ 1000 genomes - Low coverage - YRI variations

1000 Ge

Sequenc

Genes

Merge

Merge

Merge

ncRN

ncRN

Germlin

Seque

Function

Reg. I

Additio

%GC

Assembly exceptions

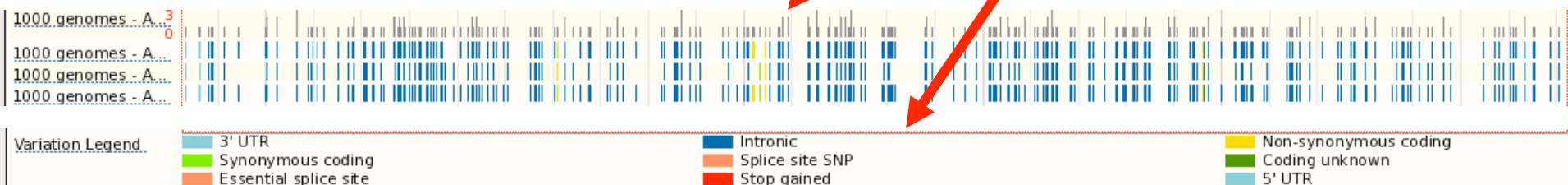
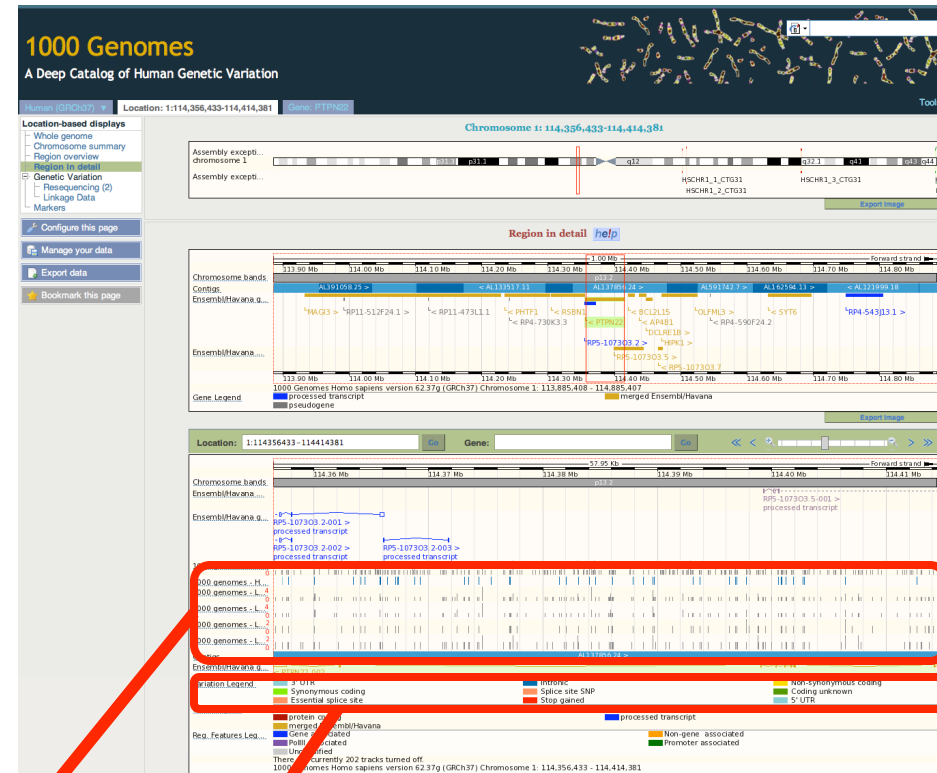
Chromosome bands

Scale bar

Ruler

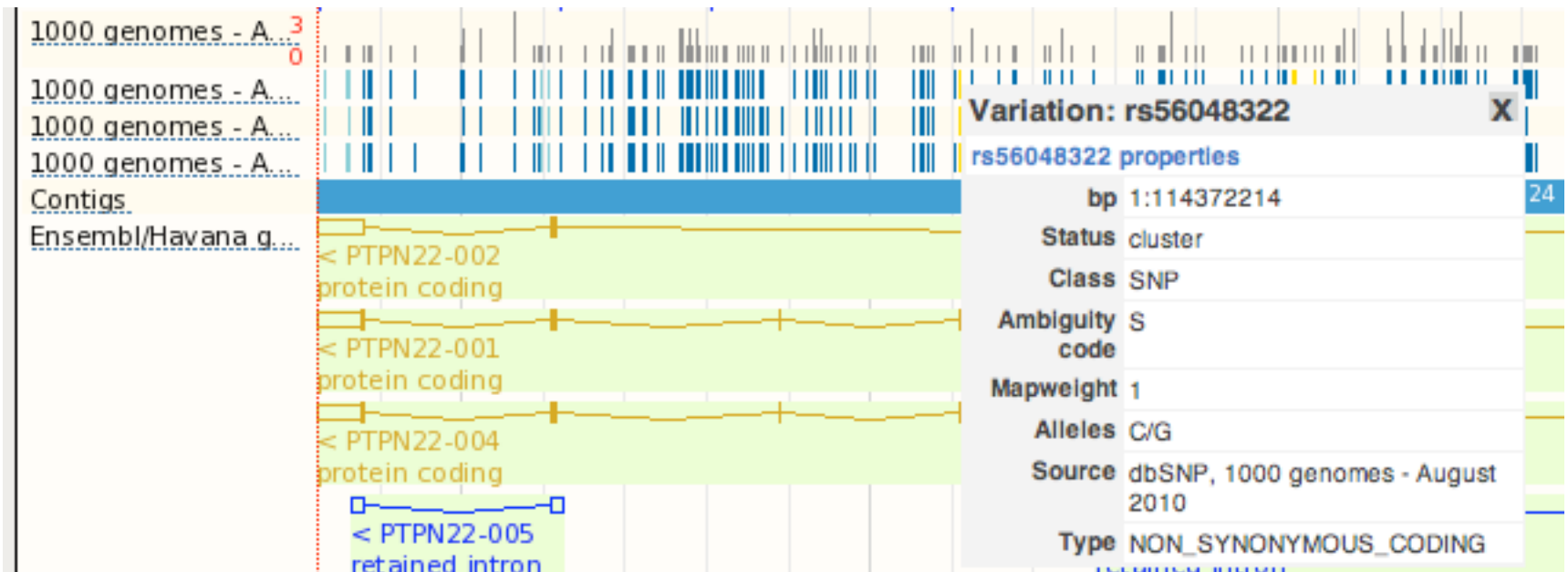
# 1000 Genomes SNPs

- The variation data of the August 2010 release and pilot project data can be viewed in specific tracks on location pages.
- These are selected from the “Configure this page” menu.
- Track can be displayed as “density” or “compact”.
- In “compact” view, variations are coloured by their functions.
- Tracks for all SNPs from dbSNP build 132 are also available

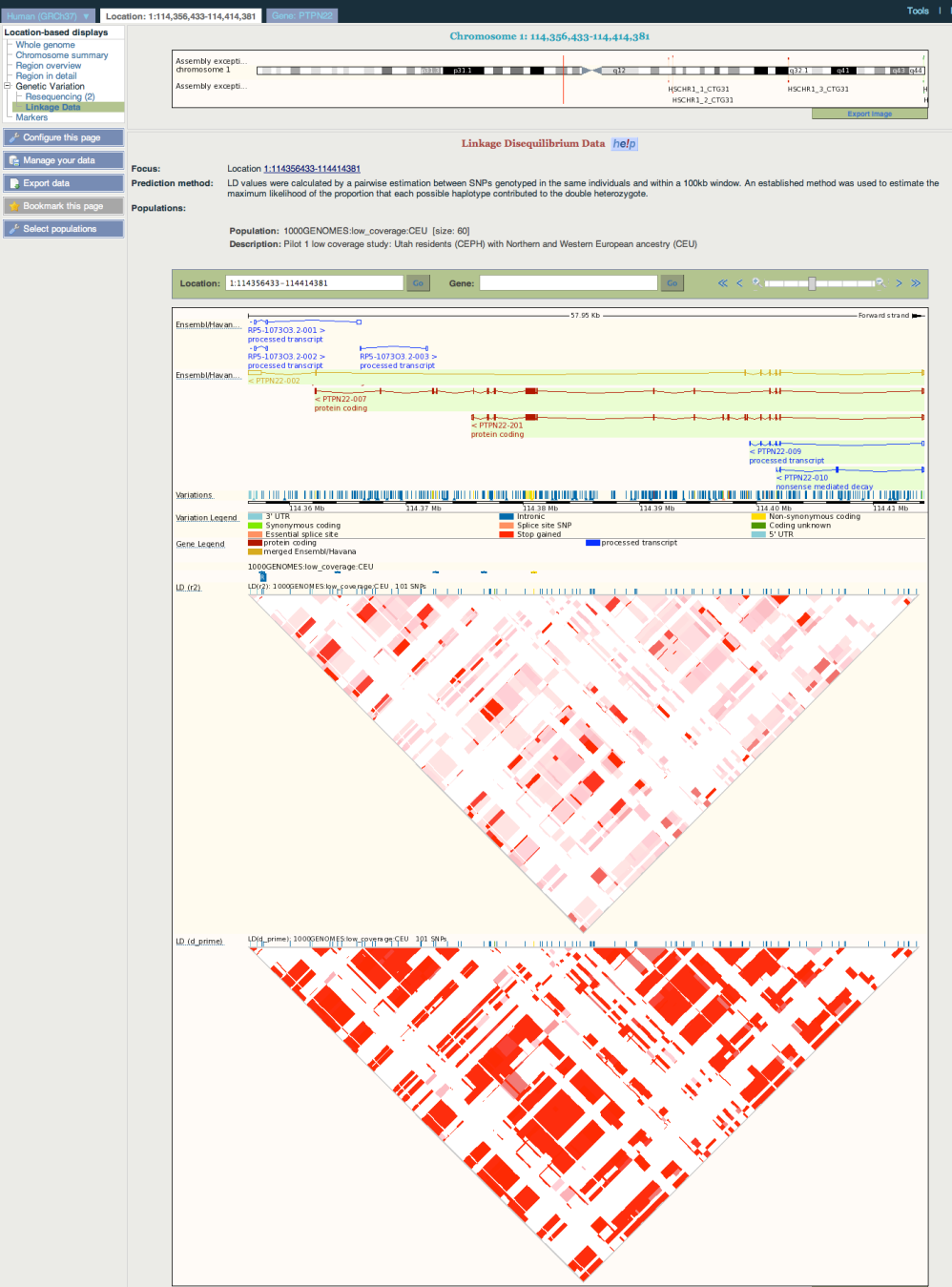


# SNP Information

- SNPs are clickable which brings up a small window with basic information
- The “SNP properties” link leads to a dedicated page for the SNP with detailed information (including information imported from dbSNP)



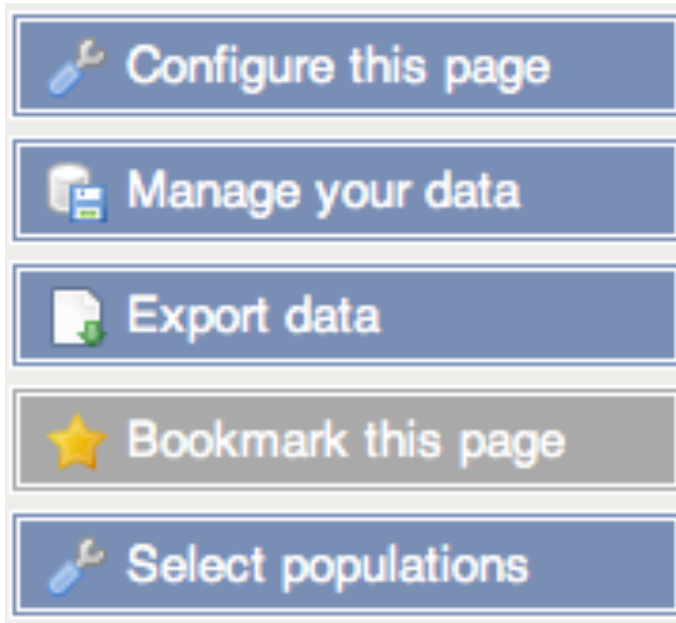




# LD Information

- Currently based on data from HapMap and 1000 genomes project populations
- Populations selectable from drop down tab

# Data Export



- Summary data from the region being viewed can be exported
- Export format (such as html, excel, text) can be specified

## 1000 Genomes

A Deep Catalog of Human Genetic Variation

The image displays a genomic track visualization of the ENST00000441259 gene region, focusing on variations and domain annotations. The top track shows the genomic coordinates (72,357,900 to 72,358,900) and a scale bar indicating 1.13 Kb. Below this, three transcripts are shown: PRF1-001, PRF1-002, and PRF1-003. Each transcript track includes a red bar representing the gene structure and a series of colored boxes indicating domain annotations. The domains shown are PROSITE\_profiles, Pfam domain, Prints domain, SMART domain, and Superfamily do.... The variations are indicated by colored boxes (green for Synonymous coding, yellow for Non-synonymous coding, and red for Stop gained) along the top track. The bottom track shows the variation legend: G/A, G/G, C/T, C/T, G/A, C/T, G/A, C/T, G/A, T/C, G/A, G/A, A/C.

ENST00000441259  
PRF1-001

PROSITE\_profiles

Pfam domain

Prints domain

SMART domain

Superfamily do....

ENST00000473106  
PRF1-002

ENST00000318971  
PRF1-001

PROSITE\_profiles

Pfam domain

Prints domain

SMART domain

Superfamily do....

ENST00000373209  
PRF1-002

PROSITE\_profiles

Pfam domain

Prints domain

SMART domain

Superfamily do....

ENST00000373209  
PRF1-002

PROSITE\_profiles

Pfam domain

Prints domain

SMART domain

Superfamily do....

Variation Legend

Synonymous coding

Non-synonymous coding

Stop gained

# More variation displays

## Transcript Page

**1000 Genomes**  
A Deep Catalog of Human Genetic Variation

Human (GRCh37) Location: 10:72,357,809-72,358,837 Gene: PRF1 Transcript: PRF1-001

Transcript-based displays

- Transcript summary
- Supporting evidence (4)
- Sequence
  - Exons (3)
  - cDNA
  - Protein
- External References
  - General identifiers (35)
  - Oligo probes (24)
- Ontology
  - Ontology chart (19)
  - Ontology table (19)
- Genetic Variation
  - Population comparison
  - Comparison image
- Protein Information
  - Protein summary
  - Domains & features
  - Variations (89)
- External Data
- ID History
  - Transcript history
  - Protein history

Configure this page  
Manage your data  
Export data  
Bookmark this page

Description: perforin 1 (p...)  
Location: Chromosome 10  
Gene: PRF1  
Transcript: PRF1-001

Show/hide columns

Statistics

Exons: This transcript  
Type: Known  
Prediction Method: Transc  
Alternative transcripts: This transcript

Transcript: PRF1-001 (ENST00000441259)

Variations [help](#)

Show All entries

Residue	Variation ID	Variation type	Alleles	Ambiguity code	Residues	Codons	SIFT	PolyPhen
1	<a href="#">CM021667</a>	Coding unknown	HGMD_MUTATION	-	-	-	-	-
1	<a href="#">CM030945</a>	Coding unknown	HGMD_MUTATION	-	-	-	-	-
4	<a href="#">rs35418374</a>	Non-synonymous coding	C/T	Y	R, H	CGT, CAT	tolerated	probably damaging
4	<a href="#">rs12161733</a>	Non-synonymous coding	G/A	R	R, C	CGT, TGT	tolerated	probably damaging
17	<a href="#">1KG INDEL 10_72360608</a>	Frameshift coding	A/-	-	-	-	-	-
17	<a href="#">CD993068</a>	Coding unknown	HGMD_MUTATION	-	-	-	-	-
32	<a href="#">rs2228018</a>	Synonymous coding	C/T	Y	K	AAG, AAA	-	-
38	<a href="#">CM060437</a>	Coding unknown	HGMD_MUTATION	-	-	-	-	-
39	<a href="#">CM023666</a>	Coding unknown	HGMD_MUTATION	-	-	-	-	-
45	<a href="#">CM040245</a>	Coding unknown	HGMD_MUTATION	-	-	-	-	-
45	<a href="#">CM023667</a>	Coding unknown	HGMD_MUTATION	-	-	-	-	-
50	<a href="#">CM010997</a>	Coding unknown	HGMD_MUTATION	-	-	-	-	-
54	<a href="#">CM040246</a>	Coding unknown	HGMD_MUTATION	-	-	-	-	-
64	<a href="#">rs104894180</a>	Stop gained	G/A	R	Q, *	CAA, TAA	-	-
65	<a href="#">CD080649</a>	Coding unknown	HGMD_MUTATION	-	-	-	-	-
69	<a href="#">CD023849</a>	Coding unknown	HGMD_MUTATION	-	-	-	-	-
70	<a href="#">CM040247</a>	Coding unknown	HGMD_MUTATION	-	-	-	-	-
73	<a href="#">CM040248</a>	Coding unknown	HGMD_MUTATION	-	-	-	-	-
89	<a href="#">CM035020</a>	Coding unknown	HGMD_MUTATION	-	-	-	-	-
91	<a href="#">rs35947132</a>	Non-synonymous coding	G/A/T/C	N	A, V	GCG, GTG	deleterious	probably damaging
95	<a href="#">CM012645</a>	Coding unknown	HGMD_MUTATION	-	-	-	-	-
123	<a href="#">COSM32312</a>	Non-synonymous coding	C/T	Y	R, H	CGT, CAT	deleterious	benign
123	<a href="#">1KG 10_72360291</a>	Non-synonymous coding	C/T	Y	R, H	CGT, CAT	deleterious	benign
123	<a href="#">CM071929</a>	Coding unknown	HGMD_MUTATION	-	-	-	-	-
126	<a href="#">rs34279237</a>	Non-synonymous coding	G/A	R	R, C	CGC, TGC	deleterious	probably damaging
135	<a href="#">rs12263572</a>	Non-synonymous coding	C/T	Y	V, M	GTG, ATG	deleterious	probably damaging

# Data Slicer

- Data Slicer allows you to get a fraction of the data within a chromosomal region you specified.
- Accessed from the Data Management page.
- Works on BAMs or VCFs accompanied by appropriate index files.

Configure page

Custom Data

Data Management

- Upload Data
- Attach DAS
- Attach Remote File
- Manage Data
- Features on Karyotype
- Data Converters
  - Assembly Converter
  - ID History Converter
  - Variant Effect Predictor
  - Data Slicer**

Tip

When slicing a VCF or BAM file, both the data file and its index file should be present on the web server and named correctly.  
The VCF file should have a ".vcf.gz" extension, and the index file should have a ".vcf.gz.tbi" extension, E.g: MyData.vcf.gz, MyData.vcf.gz.tbi  
The BAM file should have a ".bam" extension, and the index file should have a ".bam.bai" extension, E.g: MyData.bam, MyData.bam.bai

VCF / BAM File URL:

( e.g. <http://www.example.com/MyProject/MyData.vcf.gz> )

Region:

( e.g. 1:1-50000 )

Next >

Data Management

- Upload Data
- Attach DAS
- Attach Remote File
- Manage Data
- Features on Karyotype
- Data Converters
  - Assembly Converter
  - ID History Converter
  - Variant Effect Predictor
  - Data Slicer**

Thank you - your BAM file [\[1.50000-100000.HG01623.mapped.SOLID.bfast.IBS.low\\_coverage.20101123.bam\]](#) [Size: 280192] has been generated.  
Right click on the file name and choose "Save link as .." from the menu

Preview

@HD	VN:1.0	GO:none	SO:coordinate
@SQ	SN:1	LN:249250621	
@SQ	SN:2	LN:243199373	
@SQ	SN:3	LN:198022430	
@SQ	SN:4	LN:191154276	
@SQ	SN:5	LN:180915260	
@SQ	SN:6	LN:171115067	
@SQ	SN:7	LN:159138663	
@SQ	SN:8	LN:146364022	
@SQ	SN:9	LN:141213431	

# Variant Effect Predictor

[illegible]

# Upload Variant or Alignment Data Into the Browser

Configure page

Custom Data

Data Management

Upload Data

Attach DAS

Attach Remote File

Manage Data

Feature

Data C

Asse

ID H

Varia

Data

Tip

Accessing data via a URL can be slow unless you use an indexed format such as BAM. However it has the advantage that you always see the same data as the file on your own machine.

We currently accept attachment of the following formats: BAM, BED, bedGraph, GBrowse, Generic, GFF, GTF, PSL, VCF, WIG. VCF files must be indexed prior to attachment.

File URL:

( e.g. <http://www.example.com/MyProject/mydata.gff> )

Data format:

-- Choose --

Name for this track:

Next >



# Credits

- Eugene Kulesha, Glenn Proctor
- Natassa Spiridou
- Will McLaren, Fiona Cunningham
- Laura Clarke, Holly Zheng-Bradley, Rick Smith
- [info@1000genomes.org](mailto:info@1000genomes.org)