

1000 Genomes Browser Orientation and Pilot Data

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

Based on Pilot Project Data

Paul Flicek

European Bioinformatics Institute

3 November 2010

Overview

- Based on version 54 of Ensembl code and NCBI36 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 Pilot data
 - Some details and additional configuration options will be added over the coming weeks
- Please send questions to info@1000genomes.org
- 1000 Genomes full project data will be released in a new browser as the data becomes available

1000 Genomes

A Deep Catalog of Human Genetic Variation



Home

Search 1000Genomes

e.g. gene BRCA2 or AL032821.2.1.143563

Go

Start Browsing 1000 Genomes data



[Browse Human](#) →
NCBI 36

[Transcript SNP view](#) →
View the consequences of sequence variation at the level of each transcript in the genome.

[SeqAlignView](#) →
Shows read-depth data alongside SNPs

[Other sites using Ensembl software...](#)

Browser update November 2010

based on the full pilot project data described in [A map of human genome variation from population-scale sequencing](#), Nature 467, 1061.1073.

Please see 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 March 2010 release](#). All of this data has been 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 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.

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

Ensembl is a joint project of  and the [Wellcome Trust Sanger Institute](#)



1000 Genomes Browser Home Page

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

Main loc

Location-based displays

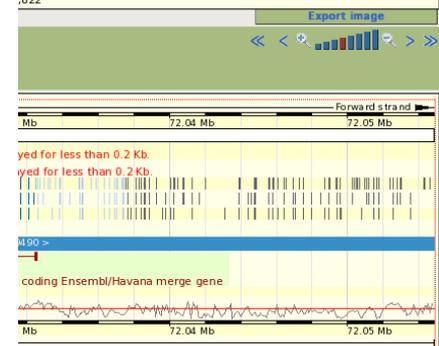
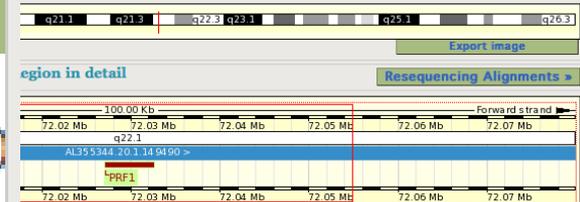
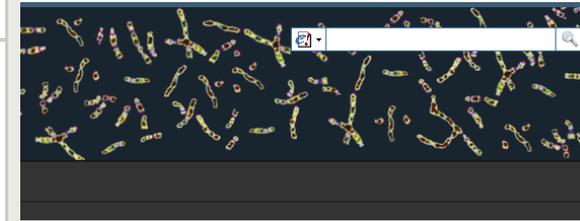
- Whole genome
- Chromosome summary
- Region overview
- Region in detail**

Comparative Genomics

Genomic alignments (0)

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

- Built on Ensembl
- Navigation is on the left hand side
- Page configuration and data expert links
- Includes only human data in current release



Variation Legend			
Intergenic	Synonymous coding	Downstream	3' UTR
5' UTR	Intronic	Upstream	Non-synonymous coding

There are currently 87 tracks turned off.
Ensembl Homo sapiens version 54.361 (NCBI36) Chromosome 10: 72,004,824 - 72,054,823

Configuring the display
You currently have 1 tracks in the overview panel and 87 tracks in the main panel turned off. To change the tracks you are displaying, use the "Configure this page" link on the left.

Page configuration

Configure page | Main panel | Top panel | Custom Data

Main panel

Active tracks

- (11/27) 1000 Genomes
- (5/5) Resembl
- (1/4) Sequence
- (0/1) Markers
- (3/3) Genes
- (0/1) Prediction Transcrip
- (0/6) Protein alignments
- (0/1) Protein features
- (1/3) cDNA/mRNA alignm
- (0/3) EST alignments
- (0/2) RNA alignments
- (0/16) Oligo features
- (0/4) Simple features
- (0/6) Misc. regions
- (0/12) Repeats
- (0/13) Variation features
- (5/5) Additional decorati
- (5/5) Information

1000 Genomes

- low_coverage_YRI variations [Show info](#)
- low_coverage_CEU variations [Show info](#)
- low_coverage_CHBJPT variations [Show info](#)
- YRI TD low coverage [Show info](#)
- CHBJPT TD low coverage [Show info](#)
- CEU TD low coverage [Show info](#)
- YRI low coverage ME insertion [Show info](#)
- CHBJPT low coverage ME insertion [Show info](#)
- 1KG Low coverage deletions [Show info](#)
- 1KG Recombination hotspots [Show info](#)
- CEU low coverage ME insertion [Show info](#)

Resembl

- Sequences [Show info](#)
- Consensus [Show info](#)
- Reference [Show info](#)
- Coverage [Show info](#)
- Reads [Show info](#)

Sequence

- Contigs [Show info](#)

Genes

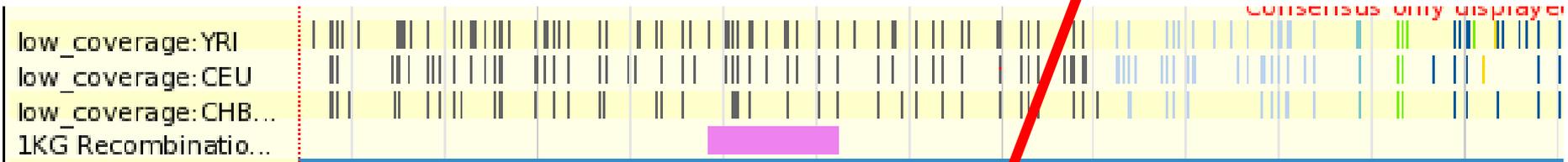
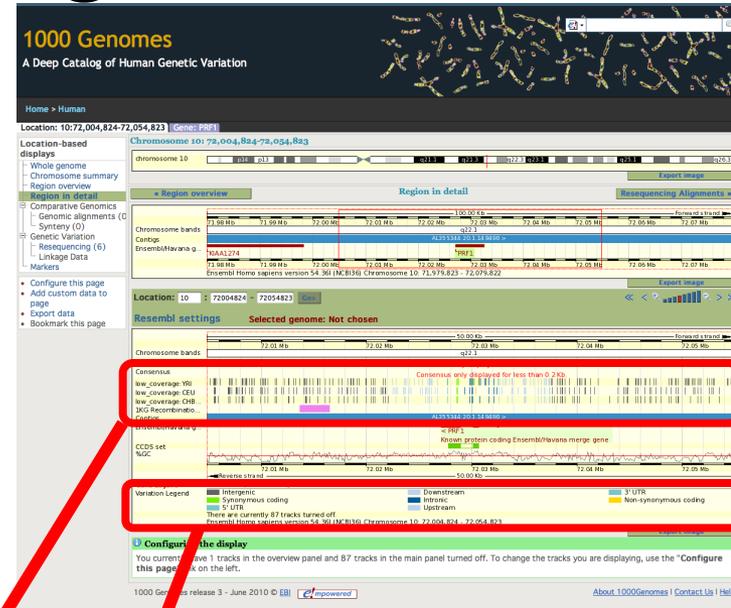
- Ig/T-cell receptor gene [Show info](#)
- Merged Ensembl and Havana Genes [Show info](#)
- ncRNA gene [Show info](#)

cDNA/mRNA alignments

- low_coverage_YRI variations [Show info](#)
- exon_CHD variations [Show info](#)
- exon_JPT variations [Show info](#)
- exon_LWK variations [Show info](#)
- exon_TSI variations [Show info](#)
- exon_YRI variations [Show info](#)
- trio_CEU variations [Show info](#)
- trio_YRI variations [Show info](#)
- low_coverage_CEU variations [Show info](#)
- low_coverage_CHBJPT variations [Show info](#)
- exon_CHB variations [Show info](#)
- exon_CEU variations [Show info](#)
- YRI trio novel insertions [Show info](#)
- CEU trio novel insertions [Show info](#)
- YRI TD trio [Show info](#)
- CEU TD trio [Show info](#)
- YRI TD low coverage [Show info](#)
- CHBJPT TD low coverage [Show info](#)
- CEU TD low coverage [Show info](#)
- YRI low coverage ME insertion [Show info](#)
- CHBJPT low coverage ME insertion [Show info](#)
- YRI trio ME insertion [Show info](#)
- CEU trio ME insertion [Show info](#)
- 1KG trio deletions [Show info](#)
- 1KG Low coverage deletions [Show info](#)
- 1KG Recombination hotspots [Show info](#)
- CEU low coverage ME insertion [Show info](#)

1000 Genomes SNPs

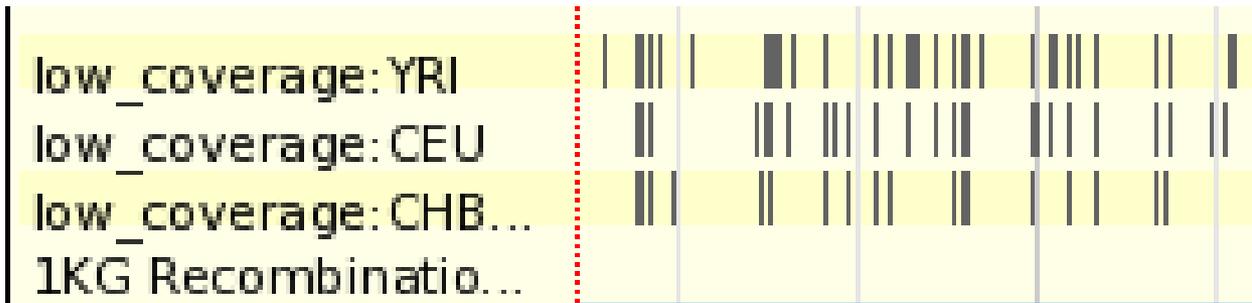
- The high coverage trios, low coverage populations and exon capture data can be viewed in specific tracks on location pages
- These are selected from the “Configure this page” menu and appear as tracks near the middle of the display
- Tracks for all SNPs from dbSNP build 129 are also available



Variation Legend	
	Intergenic
	Synonymous coding
	5' UTR
	Downstream
	Intronic
	Upstream

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)



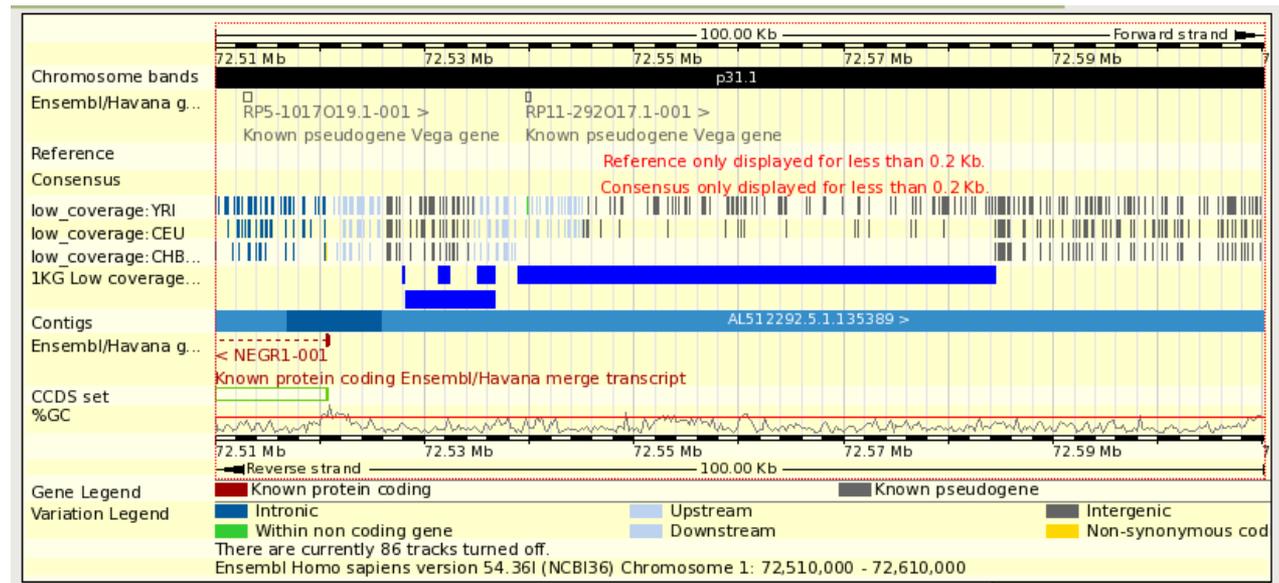
Variation: rs10740355 X

Variation Properties

bp:	72019982
status:	cluster, freq, hapmap
class:	snp
ambiguity	R
code:	
mapweight:	1
alleles:	A/G
source:	ENSEMBL:Watson, trio:YRI, trio:CEU, low_coverage:CEU, low_coverage:CHBJPT, low_coverage:YRI, ENSEMBL:celera, ENSEMBL:Venter, dbSNP
type:	INTERGENIC

SV Information

- Structural variants of specific types are selectable via the “configure this page” link
- Pop-up menus from clicking on the variant list whether the variant has been validated



Resequencing alignment

- View any region of the genome in alignment with reference for the 6 high coverage trio individuals
- Assumption made that if there is sequence coverage and not a SNP called, the base is the same as the reference
 - Not a de novo assembly
- Use “Resequencing” link on the left side of pages to access view

Location-based displays

- Whole genome
- Chromosome summary
- Region overview
- **Region in detail**
- Comparative Genomics
 - Genomic alignments (0)
 - Synteny (0)
- Genetic Variation
 - Resequencing (6)
 - Linkage Data
- Markers

Resequencing alignment options

Configure page Custom Data SAVE and close

Configure view
Configure

Configuration for: "Resequencing Alignments"

Number of base pairs per row: 120 bps

Exons to highlight: Display exons in both orientations

Matching basepairs: Replace matching bp with dots

Show variations: Yes

Line numbering: Relative to this sequence

Codons: Do not show codons

Display pop-up information on mouseover: No

Reference individual: refNCBI36

Options for resequenced Homo_sapiens individuals

trio:NA12878	<input checked="" type="checkbox"/>
trio:NA12891	<input checked="" type="checkbox"/>
trio:NA12892	<input checked="" type="checkbox"/>
trio:NA19238	<input checked="" type="checkbox"/>
trio:NA19239	<input checked="" type="checkbox"/>
trio:NA19240	<input checked="" type="checkbox"/>

To update this configuration, select your tracks and other options in the box above and close this popup window. Your view will then be updated automatically.
[Reset configuration for Resequencing Alignments to default settings.](#)

Resequencing alignment output

THIS STYLE: Location of selected exons

THIS STYLE: Location of SNPs

THIS STYLE: Resequencing coverage

- Basepairs in secondary individuals matching the reference strain are replaced with dots

~ No resequencing coverage at this position

Homo_sapiens > [chromosome:NCBI36:1:1084501:1130500:1](#)

```
refNCBI36      1 TGGTGGCTGCAGTGAGCCGAGATCGCACCCTGCCTCCAGCCTGGGCAACAGAGCCAGACTCCATTTAAAAAAGAAAAGCAGGTGAGGACGTGTGAGCAAGTCTGGGCCATGCTGCCA 120
trio:NA12878  1 .....G..... 120
trio:NA12891  1 .....R..... 120
trio:NA12892  1 .....G..... 120
trio:NA19238  1 .....G..... 120
trio:NA19239  1 .....~..... 120
trio:NA19240  1 .....R..... 120

refNCBI36     121 AGCTCCCCTTCTCCGAGCAGCTCCACACAAGGAGCAGAGGCAGCTCCAGTCCACAGCCAGACACAGTCATTTCCCTCTACTCAGGAATTAGGCAGGATGGTGTGGGGCCTGAGTGTC 240
trio:NA12878  121 .....~..... 240
trio:NA12891  121 .....~..... 240
trio:NA12892  121 .....~..... 240
trio:NA19238  121 .....~..... 240
trio:NA19239  121 .....~..... 240
trio:NA19240  121 .....~..... 240

refNCBI36     241 ACCATGAAGCCGGGAAGCAGGTGTTTCAGCCAGACGCAAAATATTCCTCCTCAAAGCCAGGGGAGCGGCCACAGTGGATTTTATTATGGGGGACGGGGGCACGCCAGTGCCCCGCCACTC 360
trio:NA12878  241 .....A..... 360
trio:NA12891  241 .....R..... 360
trio:NA12892  241 .....A..... 360
trio:NA19238  241 .....A..... 360
trio:NA19239  241 .....~..... 360
trio:NA19240  241 .....R..... 360
```

- SNPs are annotated
 - Heterozygous SNPs with ambiguity codes
- “~” means no coverage

Sequence level data (trios only)

1000 Genomes
A Deep Catalog of Human Genetic Variation

Home > Human

Location: 10:72,004,824-72,054,823 Gene: PRF1

Chromosome 10: 72,004,824-72,054,823

Location-based displays

- Whole genome
- Chromosome summary
- Region overview
- Region in detail**
- Comparative Genomics
 - Genomic alignments (0)
 - Synteny (0)
- Genetic Variation
 - Resequencing (6)
 - Linkage Data
 - Markers

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

Resembl settings Selected genome: Not chosen

Chromosome bands
Reference
Consensus
low_coverage:YRI
low_coverage:CEU
low_coverage:CHB...
1KG Recombinatio...
Contigs
Ensembl/Havana g...
CCDS set
%GC
Gene Legend
Variation Legend

There are currently 87 tracks turned off.
Ensembl Homo sapiens version 54.361 (NCBI36)

Configuring the display
You currently have 1 tracks in the overview panel and 87 tra
this page" link on the left.

1000 Genomes release 3 - June 2010 © EBI

Reference
Consensus
Sequences

Coverage 60

Reference only displayed for less than 0.2 Kb.
Consensus only displayed for less than 0.2 Kb.
Sequences is displayed only for regions less then 0.2 Kb (314)

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



LD Information

- Currently based on data from HapMap and Perlegen populations
- Populations selectable from drop down tab

Data Export

- [Configure this page](#)
- [Add custom data to page](#)
- [Export data](#)
- [Bookmark this page](#)

- Summary data from the region being viewed can be exported
- Additional export configuration options coming soon

More variation displays

Gene Page

Home > Human

Location: 10:72,026,110-72,032,537

Gene: PRF1 | Transcript: PRF1-001

Gene: PRF1

- Gene summary
- Splice variants (2)
- Supporting evidence
- Sequence
- External references (3)
- Regulation
- Comparative Genomics
- Genetic Variation
- Variation Table
- Variation Image

Perforin-1 Precursor (P1) (Lymphocyte pore-forming protein) (PPF) (Cytolysin) (EnsemblProtein/Swiss-Prot P14222)

Location: Chromosome 10: 72,027,110-72,032,537 reverse strand.

Transcripts: There are 2 transcripts in this gene. [show transcripts](#)

Name	Transcript ID	Protein ID	Description
PRF1-001	ENST00000373209	ENSP00000362305	protein_coding
PRF1-201	ENST00000318971	ENSP00000316746	protein_coding

Gene summary

Name: PRF1 (HGNC (curated))

Synonyms: HPLH2, P1, PPF [To view all Ensembl genes linked to the name [click here](#).]

CCDS: This gene is a member of the Human CCDS set: [CCDS7305](#)

Gene type: Known protein coding

Prediction: Gene containing both Ensembl genbuild transcripts and [Havana](#) manual curation, see [article](#).

Method

Transcripts

Contigs: 72.02 Mb to 72.04 Mb

Ensembl/Havana g...: Known protein coding Ensembl/Havana merge transcript for PRF1-001 and PRF1-201.

Export image

Configuring the display

Tip: use the "Configure this page" link on the left to show additional data in this region.

1000 Genomes release 3 - June 2010 © EBI | [mpowered](#) | [About 1000Genomes | Contact Us](#)

1000 Genomes

A Deep Catalog of Human Genetic Variation

Home > Human

Location: 10:72,026,110-72,033,537 | Gene: PRF1 | Transcript: PRF1-201

Gene: PRF1 (ENSG00000180644)

- Gene summary
- Splice variants (2)
- Supporting evidence
- Sequence
- External references (3)
- Regulation
- Comparative Genomics
- Genetic Variation
- Variation Table
- Variation Image
- External Data
- ID History
- Gene history

Perforin-1 Precursor (P1) (Lymphocyte pore-forming protein) (PPF) (Cytolysin) Source: UniProtKB/Swiss-Prot P14222

Location: Chromosome 10: 72,027,110-72,032,537 reverse strand.

Transcripts: There are 2 transcripts in this gene. [show transcripts](#)

Variation Table

Variations: Ensembl/Havana...

ENST00000373209 PRF1-001

ENST00000318971 PRF1-201

Pfam domain

3' UTR, 5' UTR, Intronic, Synonymous coding, Non-synonymous coding, Upstream

Configuration legend

Configuring the display

Tip: use the "Configure this page" link on the left to customise the protein domains and types of variations displayed above.

Please note the default 'Context' settings will probably filter out some intronic SNPs. None of the variations are filtered out by the Source, Class and Type filters. 23 intronic variations are removed by the Context filter.

1000 Genomes release 3 - June 2010 © EBI | [mpowered](#) | [About 1000Genomes | Contact Us | Help](#)

More variation displays

Transcript Page

Home > Human

Location: 10:72,026,110-72,033,537 Gene: PRF1 Transcript: PRF1-001

Transcript-based displays

Transcript summary

Supporting evidence (8)

Sequence

- Exons (3)
- cDNA
- Protein

External References

- General identifiers (35)
- Oligo probes (23)
- Gene ontology (10)

Population comparison

Comparison image

Domains & features (4)

Variations (16)

Perforin-1 Precursor (P1)(Lymphocyte pore-forming protein)(PPP)(Cytolysin) Source: UniProtKB/Swiss-Prot P14222

Location Chromosome 10: 72,027,110-72,032,521 reverse strand

Gene This transcript is a product of gene ENSG00000180644 - There are 2 transcripts in this gene: [hide transcripts](#)

Name	Transcript ID	Protein ID	Description
PRF1-001	ENST00000373209	ENSP00000362305	protein_coding
PRF1-201	ENST00000318971	ENSP00000316746	protein_coding

Transcript and Gene level displays

In Ensembl a gene is made up of one or more transcripts. Views in Ensembl are separated into Gene based views and Transcript based views according to which level the information is more appropriately associated. This view is a transcript level view. To flip between the two sets of views you can click on the Gene and Transcript tabs in the menu bar at the top of the page.

Transcript summary

Supporting evidence

Exons: 3 Transcript length: 2,488 bps Translation length: 555 residues

This transcript is a member of the Human CCDS set: [CCDS7305](#)

Known protein coding

Transcript where the Ensembl genebuild transcript and the Vega manual annotation have the same sequence, for every base [article](#).

This Ensembl/Havana merge transcript entry corresponds to the following database identifiers:

Havana transcript: [Q1THUMT00000048517](#) [view all locations]

Imported HAVANA transcripts (Shares CDS with ENST): [ENST00000318971](#) [view all locations]

1000 Genomes

A Deep Catalog of Human Genetic Variation

Home > Human

Location: 10:72,026,110-72,033,537 Gene: PRF1 Transcript: PRF1-201

Transcript: PRF1-201 (ENST00000318971)

Perforin-1 Precursor (P1)(Lymphocyte pore-forming protein)(PPP)(Cytolysin) Source: UniProtKB/Swiss-Prot P14222

Location Chromosome 10: 72,027,111-72,032,537 reverse strand

Gene This transcript is a product of gene ENSG00000180644 - There are 2 transcripts in this gene: [hide transcripts](#)

Name	Transcript ID	Protein ID	Description
PRF1-001	ENST00000373209	ENSP00000362305	protein_coding
PRF1-201	ENST00000318971	ENSP00000316746	protein_coding

Transcript and Gene level displays

In Ensembl a gene is made up of one or more transcripts. Views in Ensembl are separated into Gene based views and Transcript based views according to which level the information is more appropriately associated. This view is a transcript level view. To flip between the two sets of views you can click on the Gene and Transcript tabs in the menu bar at the top of the page.

Population comparison

Comparison image

Protein summary

Variations

ENST00000318971

refNCBI36 Resequencing cov.

trioNAL2878 Resequencing cov.

trioNAL2891 Resequencing cov.

trioNAL2892 Resequencing cov.

trioNAL9238 Resequencing cov.

trioNAL9239 Resequencing cov.

trioNAL9240 Resequencing cov.

545 bp

Legend:
■ Synonymous coding
■ Non-synonymous coding
■ 5' UTR
■ Missing data
■ Heterozygous

the left to customise the exon context and types of variations displayed above.
Source, Class and Type filters.
Text filter:
ons within 100 bp of exon boundaries.

About 1000Genomes | Contact Us | Help

Residue	SNP ID	SNP type	Alleles	Ambiguity code	Alternate residues
4	rs12161733	Non-synonymous	A/G	RY	H, C, R, Y
32	rs2228018	Synonymous	T/C	Y	-
91	rs35947132	Non-synonymous	A/G	R	A, V
126	rs34279237	Non-synonymous	A/G	R	C, R
135	rs12263572	Non-synonymous	T/C	Y	M, V
145	trio:YRI:10:72030230	Synonymous	T/C	Y	-
154	low_coverage:YRI:10:72030203	Synonymous	T/C	Y	-
225	rs28933973	Non-synonymous	A/G	R	W, R
242	rs35329429	Synonymous	A/G	R	-
252	rs28933375	Non-synonymous	T/C	Y	S, N
274	rs885821	Synonymous	A/G	R	-
300	rs885822	Synonymous	A/G	R	-
345	rs28933374	Non-synonymous	A/G	R	P, L
426	rs1042652	Non-synonymous	T/C	Y	S, G
435	rs28933376	Non-synonymous	A/G	R	T, M
459	rs2228019	Synonymous	A/G	R	-

Credits

- Eugene Kulesha, Glenn Proctor
- Natassa Spiridou
- Will McLaren, Fiona Cunningham
- Laura Clarke, Holly Zheng-Bradley, Rick Smith

- info@1000genomes.org