

The 1000 Genomes Tutorial The Website and Browser

Laura Clarke
17th February 2012



These slides should give you an overview of the 1000 genomes website and our Ensembl style browser



Glossary

- **Pilot** : The 1000 Genomes project ran a pilot study between 2008 and 2010
- **Phase 1**: The initial round of exome and low coverage sequencing of 1000 individuals
- **Phase 2**: Expanded sequencing of 1700 individuals and method improvement
- **SAM/BAM**: Sequence Alignment/Map Format, an alignment format
- **VCF**: Variant Call Format, a variant format



Summary

- 1000 Genomes Website
- 1000 Genomes Browser
- Gene and SNP display on Region in Detail
- Searching for a Gene
- Region in Detail
- Turning on Tracks
- File upload
- Gene View
- Structural Variation
- Gene Variation Zoom
- Transcript View
- Searching for a Variation
- Population Genetics
- New Variation Views
- Data Availability
- Announcements



http://www.1000genomes.org

1000 Genomes
A Deep Catalog of Human Genetic Variation

Home About Data Analysis Participants Contact **Browser** Wiki FTP search Search

LATEST ANNOUNCEMENTS

WEDNESDAY OCTOBER 12, 2011

October 2011 Integrated Variant Set release #ICHG2011

This **October 2011** release represents an integrated set of variant calls and phased genotypes including SNPS, short INDELS and Deletions based on low coverage and exome sequencing data across 1092 individuals.

Our [FAQ](#) contains instructions on how to get [smaller subsections](#) of these files

Data access links: [EBI](#) / [NCBI](#)

Link to additional information: [README file](#)

THURSDAY JUNE 23, 2011

June 2011 Data Release

Genotypes for 1094 individuals for the [May 2011 snp calls](#) from the 20101123 sequence and alignment release of the 1000 genomes project has now been made. This release is based on the GRCh37 assembly of the human genome and is released in the format [VCF 4.0](#)

Our [FAQ](#) contains instructions on how to get [smaller subsections](#) of these files

Data access links: [EBI](#) / [NCBI](#)

Link to additional information: [README file](#)

NAVIGATION

- [Frequently Asked Questions](#)

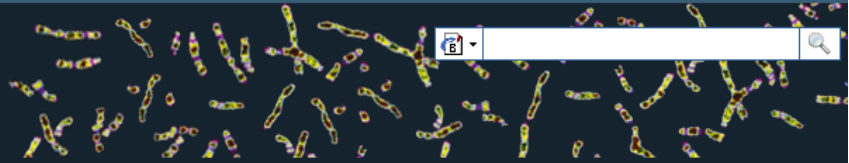
LINKS

- [All Project Announcements](#)
- [Sample and Project Information](#)
- [Media Archive](#)
- [Download the 1000 Genomes Pilot Paper](#)
- [Project Contacts](#)



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 September 2011

based on interim Main project data from 20101123 for 1094 individuals and ensembl release 63. The data can be found on [the ftp site](#).

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 1000 Genomes Project data by the whole scientific community, this browser (based on Ensembl) integrates the SNP calls from an [interim release 20101123](#). 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 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.



[Tutorial](#) →
The 1000 Genomes Browser Tutorial.

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 EMBL-EBI  and the [Wellcome Trust Sanger Institute](#)



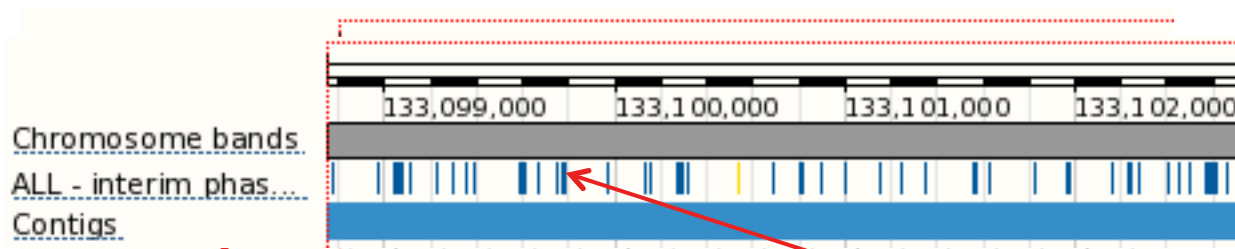
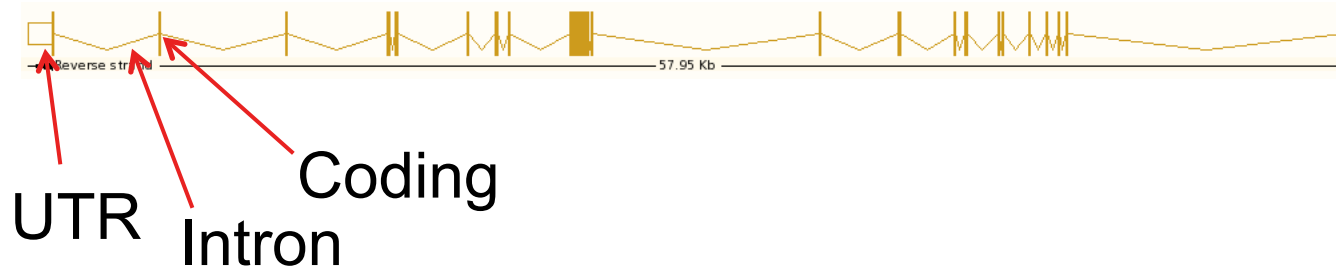
1000 Genomes release 10 - October 2011 © [EBI](#)

[About 1000 Genomes](#) | [Contact Us](#) | [Help](#)

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



Genes and SNPs

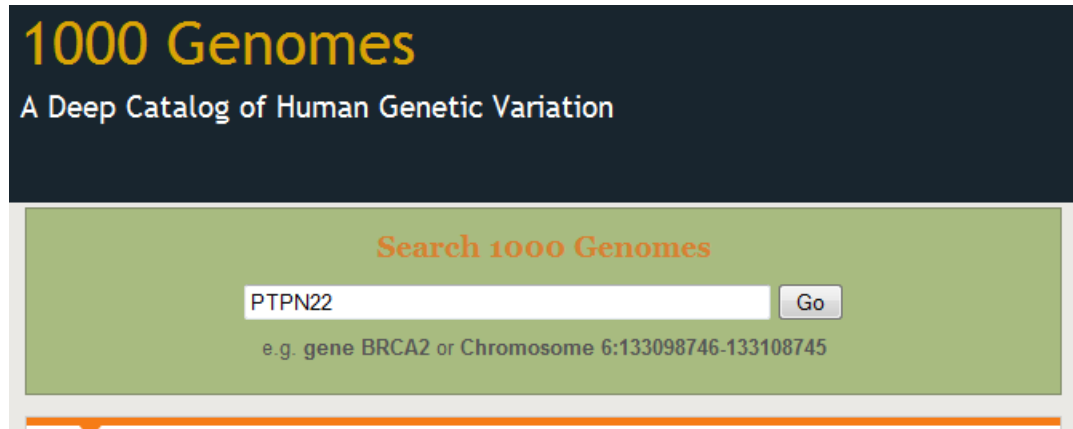


Line indicates number of SNPS

Each Line is One SNP

Searching the Browser

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



1000 Genomes
A Deep Catalog of Human Genetic Variation

Search 1000 Genomes

PTPN22

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

- Search for PTPN22
- Click 'Region in Detail'

You searched for 'PTPN22'

Gene or Gene Product

6 entrie(s) matched your search strings.

1. **Gene:** [ENSG00000134242](#) [[Region in detail](#)]
PTPN22 - protein tyrosine phosphatase, non-receptor type 22 (lymphoid) [Source:HGNC Symbol;Acc:9652]
2. **Variations in gene ENSG00000134242:** [[Variations in gene](#)]
3. **Transcript:** [ENST00000359785](#) [[Region in detail](#)]
4. **Peptide:** [ENSP00000435176](#) [[Region in detail](#)]
PTPN22
5. **Peptide:** [ENSP00000352833](#) [[Region in detail](#)]
PTPN22
6. **Peptide:** [ENSP00000346621](#) [[Region in detail](#)]
PTPN22



Region in Detail

Ensembl genome browser 9: Homo sapiens - Region in detail - Chromosome 1: 114,356,433-114,414,381

browser.1000genomes.org/Homo_sapiens/Location/View?db=core;g=ENSG00000134242;r=1:114356433-114414381

1000 Genomes

A Deep Catalog of Human Genetic Variation

Human (GRCh37) Location: 1:114,356,433-114,414,381 Gene: PTPN22

Location-based displays

- Whole genome
- Chromosome summary
- Region overview
- Region in detail**
- Genetic Variation
 - Resequencing (20)
 - Linkage Data
 - Markers

Configure this page
Manage your data
Export data
Get VCF data
Bookmark this page
View in Ensembl

Chromosome 1: 114,356,433-114,414,381

Assembly exceptions chromosome 1

Assembly exceptions

HSCHR1_1_CTG31 HSCHR1_2_CTG31 HSCHR1_3_CTG31

Region in detail help

Chromosome bands

Contigs

Ensembl/Havana g...
MAGI3
RP11-512F24.1
RP11-473L1.1
PHTF1
RP4-730...
PTPN22
BCL2L15
OLFML3
AP4B1
DCLRE1B
SYT6
RP4-543J13.1
RP5-1073O3.2
HIPK1
RP5-1073O3.5
RP5-1073O3.7

Ensembl/Havana g...
1000 Genomes Homo sapiens version 63.37 (GRCh37) Chromosome 1: 113,885,408 - 114,885,407
processed transcript
merged Ensembl/Havana

Gene Legend
processed transcript
pseudogene

Location: 1:114356433-114414381 Go

Gene: Go

Ensembl/Havana g...
RPS-1073O3.2-001 > processed transcript
RPS-1073O3.2-002 > processed transcript
RPS-1073O3.2-003 > processed transcript

Contigs

Ensembl/Havana g...
PTPN22-002 protein coding
PTPN22-001 protein coding
PTPN22-004

Turning on Tracks

Configure this page



Configure Region Image | **Configure Overview Image** | Custom Data

Configure view

- Image options
 - Active tracks
 - Favourite tracks
 - Track order
 - Search results
 - 1000 Genomes (2/5)**
 - 1000 Genomes VCF (0/1)
 - Sequence (2/4)
 - Markers (1/1)
 - Genes (5/5)
 - Prediction transcripts (0/1)
 - Protein alignments (0/5)
 - Protein features (4/5)
 - cDNA/mRNA alignments (0/2)
 - RNA alignments (0/2)

1000 Genomes

- Enable/disable all tracks
- ALL - interim phase 1 - 1000 Genomes variations
- AFR - interim phase 1 - 1000 Genomes variations
- AMR - interim phase 1 - 1000 Genomes variations
- ASN - interim phase 1 - 1000 Genomes variations
- EUR - interim phase 1 - 1000 Genomes variations



Configure Region Image | **Configure Overview Image** | Custom Data

Configure view

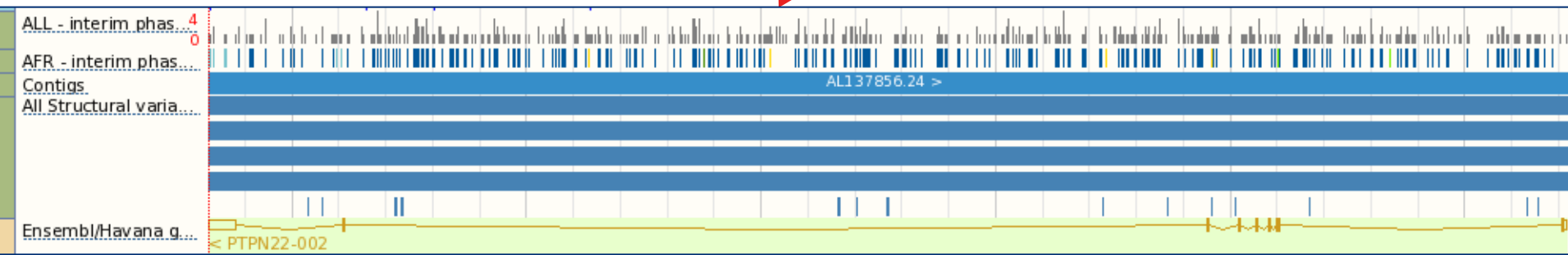
- Image options
- Favourite tracks
- Track order
- Search results**
- 1000 Genomes (2/5)
- 1000 Genomes VCF (0/1)
- Sequence (1/4)
- Markers (0/1)
- Genes (5/5)
- Prediction transcripts (0/1)

Germline variation

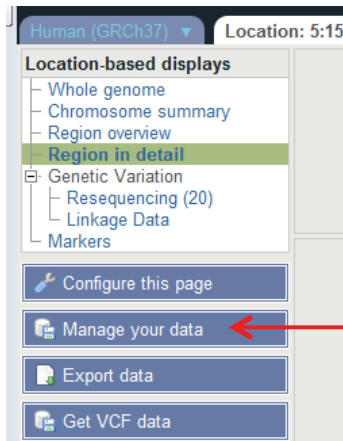
- Structural variants (all sources)
- DGVa structural variations

Key

- Track style



File upload to view with 1000 Genomes data



Manage your data

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
 - Variation Pattern Finder

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:

Name for this track:

Next >

- Supports popular file types:
 - BAM, BED, bedGraph, BigWig, GBrowse, Generic, GFF, GTF, PSL, VCF*, WIG
- * VCF must be indexed



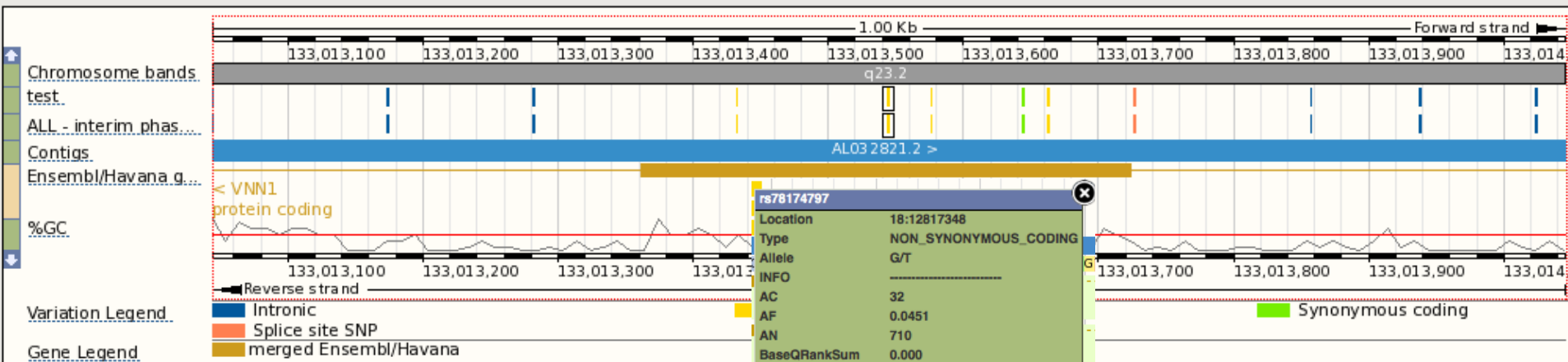

Uploaded VCF

Example:

ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20110521/ALL.wgs.phase1_release_v2.20101123.snps_indels_sv.sites.vcf.gz

Location:

Gene:



Uploaded BAM

Example:

http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data/HG01375/alignment/HG01375.mapped.ILLUMINA.bwa.CLM.low_coverage.20111114.bam



Back to browsing...

Click the Gene tab, then 'Variation Table' or 'Variation Image'

Gene Tab

Human (GRCh37) Location: 1:114,362,205-114,362,276 Gene: PTPN22

Gene: PTPN22 (ENSG00000134242)

Description: protein tyrosine phosphatase, non-receptor type 22 (lymphoid) [Source:HGNC Symbol;Acc:9652]
Location: [Chromosome 1: 114,356,433-114,414,381 reverse strand.](#)
Transcripts: There are 12 transcripts in this gene
Click the plus to show the transcript table

Variation Table [help](#)

Summary of variations in ENSG00000134242 by consequence type

Show entries

Number of variants	Type	Description
19 Show	Essential splice site	In the first 2 or the last 2 basepairs of an intron
9 Show	Stop gained	In coding sequence, resulting in the gain of a stop codon
0 -	Stop lost	In coding sequence, resulting in the loss of a stop codon
0 -	Complex in/del	Insertion or deletion that spans an exon/intron or coding sequence/UTR border
0 -	Frameshift coding	In coding sequence, resulting in a frameshift
160 Show	Non-synonymous coding	In coding sequence and results in an amino acid change in the encoded peptide sequence
65 Show	Splice site	1-3 bps into an exon or 3-8 bps into an intron
0 -	Partial codon	Located within the final, incomplete codon of a transcript whose end coordinate is unknown
83 Show	Synonymous coding	In coding sequence, not resulting in an amino acid change (silent mutation)

Get VCF data

Download as csv

Get in vcf format

Structural variation (in the Gene tab)

Human (GRCh37) Location: 1:114,356,433-114,414,381 Gene: PTPN22

Gene: **PTPN22 (ENSG00000134242)**

Description: protein tyrosine phosphatase, non-receptor type 22 (lymphoid) [Source:HGNC Symbol;Acc:9652]
 Location: [Chromosome 1: 114,356,433-114,414,381](#) reverse strand.
 Transcripts: There are 12 transcripts in this gene

Structural Variation

Name	Transcript ID	Length (bp)	Protein ID	Length (aa)	Biotype	CCDS
PTPN22-001	ENST00000359785	3654	ENSP00000352833	807	Protein coding	CCDS863
PTPN22-002	ENST00000460620	1794	ENSP00000433141	179	Protein coding	-
PTPN22-004	ENST00000528414	3424	ENSP00000435176	752	Protein coding	-
PTPN22-006	ENST00000420377	2726	ENSP00000388229	795	Protein coding	-
PTPN22-007	ENST00000525799	2118	ENSP00000432674	668	Protein coding	-
PTPN22-201	ENST00000354605	2347	ENSP00000346621	691	Protein coding	CCDS864
PTPN22-202	ENST00000538253	2414	ENSP00000439372	563	Protein coding	-
PTPN22-008	ENST00000532224	2421	ENSP00000431249	135	Nonsense mediated decay	-
PTPN22-010	ENST00000529045	527	ENSP00000434932	92	Nonsense mediated decay	-
PTPN22-009	ENST00000534519	565	No protein product	-	Processed transcript	-
PTPN22-003	ENST00000484147	2258	No protein product	-	Retained intron	-
PTPN22-005	ENST00000469077	562	No protein product	-	Retained intron	-

All Structural varia...

Structural variants

Name	Chr:bp	Genomic size (bp)	Class	Source Study	Study description
nsv435973	1.81610203-127449918	45,839,716	SV	DGVa.nstd16	Database of Genomic Variants Archive: Korbel 2007 "Paired-end mapping reveals extensive structural variation in the human genome." PMID:17901297 [remapped from build NCBI36]
esv705	1.113157135-116741372	3,584,238	SV	DGVa.estd1	Database of Genomic Variants Archive: Redon 2006 "Global variation in copy number in the human genome." PMID:17122850 [remapped from build NCBI35]
esv21206	1.113862952-114901117	1,038,166	SV	DGVa.estd20	Database of Genomic Variants Archive: Conrad 2009 "Origins and functional impact of copy number variation in the human genome." PMID:19812545 [remapped from build NCBI36]
esv23869	1.113862952-114901117	1,038,166	SV	DGVa.estd20	Database of Genomic Variants Archive: Conrad 2009 "Origins and functional impact of copy number variation in the human genome." PMID:19812545 [remapped from build NCBI36]
CN_447814	1.114360689-114360713	25	CNV_PROBE	Affy	Copy Number Variation (CNV) probes from the Affymetrix Genome-Wide Human SNP Array 6.0

Table



Variation Image

- Gene variation zoom

1000 Genomes

A Deep Catalog of Human Genetic Variation

Human (GRCh37) Location: 13:32,890,598-32,890,664 Gene: BRCA2

Gene-based displays

- Gene summary
- Splice variants (6)
- Supporting evidence
- Sequence
- External references
- Regulation
- Genetic Variation
- Variation Table
- Variation Image**
- External Data
- ID History
- Gene history

Gene: BRCA2 (ENSG00000139618)

Description: breast cancer 2, early onset [Source:HGNC Symbol;Acc:1101]

Location: Chromosome 13: 32,889,811-32,973,805 forward strand.

Transcripts: There are 6 transcripts in this gene

Name	Transcript ID	Length (bp)	Protein ID	Length (aa)	Biotype	CCDS
BRCA2-001	ENST00000380152	10930	ENSP00000369407	3418	Protein coding	CCDS9344
BRCA2-003	ENST00000530893	2009	ENSP00000435689	602	Protein coding	-
BRCA2-201	ENST00000544465	10984	ENSP00000439902	3418	Protein coding	CCDS9344
BRCA2-002	ENST00000470094	842	ENSP00000434988	188	Nonsense mediated decay	-
BRCA2-005	ENST00000507292	495	ENSP00000433188	64	Nonsense mediated decay	-
BRCA2-006	ENST00000533776	523	No protein product	-	Retained intron	-

Transcript and Gene level displays

In 1000 Genomes we provide displays at two levels:

- Transcript views which provide information specific to an individual transcript such as the cDNA and CDS sequences and protein domain annotation.
- Gene views which provide displays for data associated at the gene level such as orthologues, paralogues, regulatory regions and splice variants.

This view is a gene level view. To access the transcript level displays select a Transcript ID in the table above and then navigate to the information you want using the menu at the left hand side of the page. To return to viewing gene level information click on the Gene tab in the menu bar at the top of the page.

Location: 13:32890598-32890664 Go Variation ID: Go

Variations

ENST00000380152
BRCA2-001

PIRSF domain
PIRSF002397
DNA_recomb/repair_BRCA2

PROSITE_profiles

Pfam_domain

Superfamily do...

ENST00000470094
BRCA2-002

Pfam_domain

Superfamily do...

ENST00000530893
BRCA2-003

Stop gained

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. 5 of the 20 variations in this region have been filtered out by the Source, Class and Type filters. None of the intronic variations are removed by the Context filter.



Transcript Tab: Variations

Effect on Protein:

- SIFT
- PolyPhen

1000 Genomes A Deep Catalog of Human Genetic Variation

Human (GRCh37) Location: 1:114,359,433-114,414,381 Gene: PTPN22 Transcript: PTPN22-001

Transcript: PTPN22-001 (ENST00000359785)

Description: protein tyrosine phosphatase, non-receptor type 22 (lymphoid) [Source:HGNC Symbol;Acc:9652]
 Location: Chromosome 1:114,359,433-114,414,381 reverse strand.
 Gene: This transcript is a product of gene [ENSG00000134242](#) - There are 12 transcripts in this gene

Name	Transcript ID	Length (bp)	Protein ID	Length (aa)	Biotype	CCDS
PTPN22-001	ENST00000359785	3654	ENSP00000352833	807	Protein coding	CCDS8883
PTPN22-002	ENST00000460620	1794	ENSP00000433141	179	Protein coding	-
PTPN22-004	ENST00000528414	3424	ENSP00000435176	752	Protein coding	-
PTPN22-006	ENST00000420377	2726	ENSP00000388229	795	Protein coding	-
PTPN22-007	ENST00000525799	2118	ENSP00000432674	668	Protein coding	-
PTPN22-201	ENST00000354605	2347	ENSP00000346621	691	Protein coding	CCDS884
PTPN22-202	ENST00000538253	2414	ENSP00000439372	563	Protein coding	-
PTPN22-008	ENST00000532224	2421	ENSP00000431249	135	Nonsense mediated decay	-
PTPN22-010	ENST00000529045	527	ENSP00000434932	92	Nonsense mediated decay	-
PTPN22-009	ENST00000534519	565	No protein product	-	Processed transcript	-
PTPN22-003	ENST00000484147	2258	No protein product	-	Retained intron	-
PTPN22-005	ENST00000469077	562	No protein product	-	Retained intron	-

Transcript and Gene level displays

VIEWS IN 1000 GENOMES ARE SEPARATED INTO GENE BASED VIEWS AND TRANSCRIPT BASED VIEWS ACCORDING TO WHICH LEVEL THE INFORMATION IS MORE APPROPRIATELY ASSOCIATED WITH. 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.

Variations [help](#)

Download view as CSV

Residue	Variation ID	Variation type	Alleles	Ambiguity code	Residues	Codons	SIFT	PolyPhen
16	rs74163639	Synonymous coding	G/A	R	S	AGC, AGT	-	-
49	rs61745743	Synonymous coding	A/G	R	A	GCT, GCC	-	-
71	rs74163642	Non-synonymous coding	A/G	R	V, A	GTA, GCA	deleterious	probably damaging
141	rs115552198	Non-synonymous coding	G/A	R	R, C	CGC, TGC	deleterious	probably damaging
177	1KG_1_114399013	Synonymous coding	C/T	Y	K	AAG, AAA	-	-
183	rs34590413	Stop gained	G/A	R	R, *	CGA, TGA	-	-
201	rs74163647	Non-synonymous coding	G/A	R	S, F	TCT, TTT	deleterious	probably damaging
206	rs61738614	Non-synonymous coding	A/C	M	L, R	CTT, CGT	deleterious	probably damaging
232	rs78195073	Synonymous coding	T/C	Y	G	GGA, GGG	-	-
247	rs35910094	Synonymous coding	T/G	K	L	CTA, CTC	-	-
263	rs33996649	Non-synonymous coding	C/T	Y	R, Q	CGG, CAG	tolerated	benign
266	rs72650670	Non-synonymous coding	G/A	R	R, W	CGG, TGG	deleterious	probably damaging
277	rs72483511	Stop gained, Splice site	C/A	M	E, *	GAA, TAA	-	-
324	rs113984534	Synonymous coding	A/G	R	Y	TAT, TAC	-	-
366	rs74163654	Synonymous coding	C/T	Y	E	GAG, GAA	-	-
370	rs72650671	Non-synonymous coding	G/T	K	H, N	CAC, AAC	deleterious	possibly damaging
388	rs77913785	Non-synonymous coding	G/T	K	D, E	GAC, GAA	deleterious	benign
413	1KG_1_114380784	Non-synonymous coding	T/G	K	Q, P	CAA, CCA	deleterious	benign
414	1KG_1_114380780	Synonymous coding	A/G	R	S	AGT, AGC	-	-
427	rs112873647	Non-synonymous coding	-ATT	-	-, N	-, AAT	-	-
444	rs74163655	Non-synonymous coding	T/A	W	I, L	ATA, TTA	tolerated	benign
447	rs112191110	Non-synonymous coding	G/A	R	T, I	ACC, ATC	deleterious	probably damaging
452	rs56174946	Synonymous coding	A/G	R	F	TTT, TTC	-	-
456	rs72650672	Non-synonymous coding	G/C	S	Q, E	CAG, GAG	deleterious	possibly damaging
477	rs74163656	Synonymous coding	A/G	R	L	CAT, CAC	-	-

Start again- search for a variation (rs31685)

1000 Genomes
A Deep Catalog of Human Genetic Variation

Search 1000 Genomes

rs31685

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

- The Variation tab- left hand links take you to more information

Human (GRCh37) Location: 5:159,283,673-159,284,673 Variation: rs31685

Variation displays

- Flanking sequence
- Gene/Transcript (1)
- Population genetics (117)
- Individual genotypes (4343)
- Genomic context
- Phenotype Data
- Phylogenetic Context
- External Data

Variation: rs31685

Variation class SNP ([rs31685](#) source [dbSNP_132](#) - Variants (including SNPs and indels) imported from dbSNP [<http://www.ncbi.nlm.nih.gov/projects/SNP/>])

Synonyms Affy GeneChip 100K Array SNP_A-1683078
Affy GeneChip 500K Array SNP_A-4265358
Affy GenomeWideSNP_6.0 AFFY_6_1M_SNP_A-4265358, SNP_A-4265358
dbSNP [rs17746160](#), [rs60752908](#), [rs713581](#), [rs58941657](#)
ENSEMBL ENSSNP12948257, ENSSNP9597299

Present in + This feature is present in **1000 genomes** and 3 other sets - click the plus to show all sets

Alleles G/A (Ambiguity code: R)

Ancestral allele A

Location This feature maps to 5:159284173 (forward strand) | [View in location tab](#)

Validation status Proven by **cluster, frequency, doublehit, 1000Genome HapMap variant**

HGVS names + This feature has 2 HGVS names - click the plus to show

[Configure this page](#)

[Manage your data](#)

[Export data](#)

[Get VCF data](#)

- Population

1000 Genomes

A Deep Catalog of Human Genetic Variation

Human (GRCh37) Location: 6:74,125,388-74,126,388 Variation: rs311685

Variation displays

- Flanking sequence
- Gene/Transcript (3)
- Population genetics (46)**
- Individual genotypes (2769)
- Genomic context
- Phenotype Data
- Phylogenetic Context
- External Data

Variation class SNP (rs311685 source dbSNP_132 - Variants (including SNPs and indels) imported from dbSNP [http://www.ncbi.nlm.nih.gov/projects/SNP/])

Synonyms Affy GeneChip 100K Array SNP_A-1679873
Affy GenomeWideSNP_6.0 AFFY_6_1M_SNP_A-8668494, SNP_A-8668494
dbSNP rs58378291, rs17756820, rs52794514, rs524803, rs3173186, rs11567000, rs17421786
ENSEMBL ENSNP9062281
Illumina_Human1M-duoV3 rs311685
Uniprot VAR_057235

Present in 1000 genomes - High coverage - Trios (1000 genomes - High coverage - Trios - CEU, 1000 genomes - High coverage - Trios - YRI), 1000 genomes - Low coverage (1000 genomes - Low coverage - CEU, 1000 genomes - Low coverage - CHB+JPT, 1000 genomes - Low coverage - YRI), ALL - interim phase 1 - 1000 Genomes (AFR - interim phase 1 - 1000 Genomes, AMR - interim phase 1 - 1000 Genomes, ASN - interim phase 1 - 1000 Genomes, EUR - interim phase 1 - 1000 Genomes), ENSEMBL:Venter,HapMap

Alleles A/G (Ambiguity code: R)

Ancestral allele A

Location This feature maps to 6:74125888 (forward strand) | [View in location tab](#)

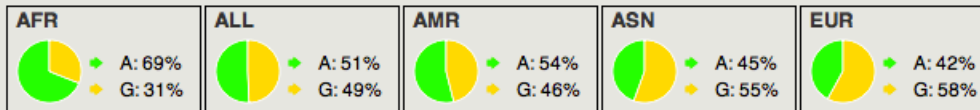
Validation status Proven by cluster, frequency, doublehit, 1000Genome HapMap variant

HGVS names This feature has 4 HGVS names - click the plus to show

Population genetics [help](#)



1000 genomes alleles frequencies



1000 genomes

Show/hide columns Filter

Population	Alleles A	Alleles G	Genotypes A/A	Genotypes A/G	Genotypes G/G	Count
1000GENOMES:AFR	0.689	0.311	0.463	0.451	0.085	114
1000GENOMES:ALL	0.507	0.493	0.269	0.477	0.254	294
1000GENOMES:AMR	0.539	0.461	0.293	0.492	0.215	53
1000GENOMES:ASN	0.446	0.554	0.199	0.493	0.308	57
1000GENOMES:EUR	0.421	0.579	0.184	0.475	0.341	70

1000 genomes pilot

Show/hide columns Filter

Population	ssID	Submitter	Alleles A	Alleles G	Count
1000GENOMES:pilot 1 CEU low coverage panel	ss233534774	1000GENOMES	0.458	0.542	
1000GENOMES:pilot 1 CHB+JPT low coverage panel	ss240577229	1000GENOMES	0.400	0.600	
1000GENOMES:pilot 1 YRI low coverage panel	ss222470667	1000GENOMES	0.729	0.271	

The Browser: Coming Soon

e!Ensembl BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors Login · Register

Human (GRCh37) Location: 9:22,125,003-22,126,003 Variation: rs1333049

Variation displays

- Explore this variation
- Genomic context
 - Gene/Transcript (2)
- Population genetics (28)
- Individual genotypes (1737)
- Linkage disequilibrium
- Phenotype Data (8)
- Phylogenetic Context (4)
- Flanking sequence
- External Data

rs1333049 SNP

Source [dbSNP 134](#) - Variants (including SNPs and indels) imported from [dbSNP](#)

Alleles Reference/Alternative: **G/C** | Ancestral: **C** | Ambiguity code: **S** | MAF: **0.40** (C)

Location Chromosome **9:22125503** (forward strand) | [View in location tab](#)

Validation status This variation is validated by **1000 Genomes**, **HapMap** and also cluster, doublehit, frequency, precious, submitter

Synonyms This feature has **7** synonyms - click the plus to show

HGVS name [g.22125503G>C](#)

[Configure this page](#)
[Manage your data](#)
[Export data](#)
[Bookmark this page](#)

Explore this variation [help](#)

- Genomic context**
- Gene / Transcript**
- Population genetics**
- Individual genotypes**
- Linkage disequilibrium**
- Phenotype data**
- Phylogenetic context**
- Flanking sequence**

Help with variations

YouTube videos

- [SNPs and other Variations - 1 of 2](#)
- [SNPs and other Variations - 2 of 2](#)
- [Clip: Genome Variation](#)
- [BioMart: Variation IDs to HGNC Symbols](#)

Reference materials

- [Ensembl variation data: background and terminology](#)
- [Variation Quick Reference card](#)

Additional resources

- [Accessing variation data with the Variation API](#)
- [Genomes and SNPs in Malaria](#)



Data Availability

- FTP site: <ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/>
 - Raw Data Files
- Web site: <http://www.1000genomes.org>
 - Release Announcements
 - Documentation
- Ensembl Style Browser: <http://browser.1000genomes.org>
 - Browse 1000 Genomes variants in Genomic Context
 - Variant Effect Predictor
 - Data Slicer
 - Other Tools



Announcements

- <http://1000genomes.org>
- 1000announce@1000genomes.org
- <http://www.1000genomes.org/1000-genomes-announcement-mailing-list>
- <http://www.1000genomes.org/announcements/rss.xml>
- <http://twitter.com/#!/1000genomes>



Questions

Please send any questions about this presentation and any other material on our website to info@1000genomes.org



Thanks

- The 1000 Genomes Project Consortium
- Paul Flicek
- Richard Smith
- Holly Zheng Bradley
- Ian Streeter
- Giulietta Spudich

