

# The 1000 Genomes Project: A Tutorial

## Browser and Tools Exercises



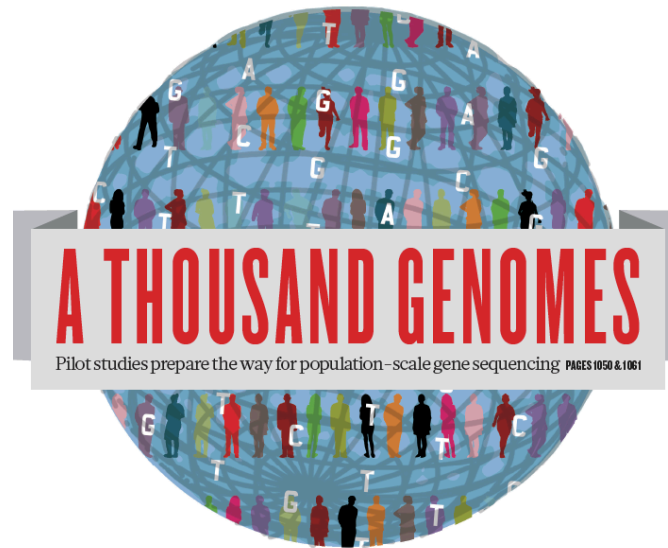
# Browser and Tools Exercises

These exercises use the 1000 genomes website and browser and tools built around it.

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

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





## The 1000 Genomes Project:

Exercise 1: Finding Data and viewing data on the 1000 genomes ftp site



# Finding Data: Exercise

1. Find what VCF files we have containing genotypes from the Illumina Omni platform.  
<http://www.1000genomes.org/ftpsearch>
2. Find the FAQ question which gives you instructions on how to get a sub-section of a VCF file. The Search Box is on the top right hand corner of any website page.  
<http://www.1000genomes.org/>



# Finding Data: Exercise

## 1. Finding Omni VCF Files

The screenshot shows the EBI FTP search interface. At the top, there is a navigation bar with links: Home, About, Data, Analysis, Participants, Contact, Browser, Wiki, and FTP search. A search box is located to the right of the navigation bar. Below the navigation bar, the page title is "SEARCH 1000 GENOMES FTP FILES". The search term "omni\*vcf" is entered in the search box. Below the search box, there is a "Search" button. To the right of the search box, there is a user profile for "LAURA@EBI.AC.UK" with a list of links: My account, Create content, List content, List users, Manage files, Log out, and Frequently Asked Questions. Below the search box, there is a "Help on searching" link. Below the help link, there is a "Search options" section with a list of checkboxes: Use NCBI FTP site, Dump MD5LIST, Exclude FASTQ files, Exclude BAM files, Exclude pilot data, Only pilot data, Exclude index files, and Exclude any .bai, .bas or .tbi file. Two red arrows are pointing to the search options section: one from the left pointing to the "Search options" header, and one from the bottom right pointing to the "Exclude pilot data" checkbox.

Home >  
**SEARCH 1000 GENOMES FTP FILES**

Search term:  
omni\*vcf  
Search for files on the FTP site

[Help on searching](#)

Search options

- Use NCBI FTP site
- Dump MD5LIST
- Exclude FASTQ files
- Exclude BAM files
- Exclude pilot data
- Only pilot data
- Exclude index files
- Exclude any .bai, .bas or .tbi file

Search

LAURA@EBI.AC.UK

- My account
- Create content
- List content
- List users
- Manage files
- Log out
- Frequently Asked Questions



# Finding Data: Exercise

Home About Data Analysis Participants Contact Browser Wiki FTP search

Home >  
**SEARCH 1000 GENOMES FTP FILES**

Search term:  
  
Search for files on the FTP site

[Help on searching](#)

– [Search options](#)

---

## RESULTS

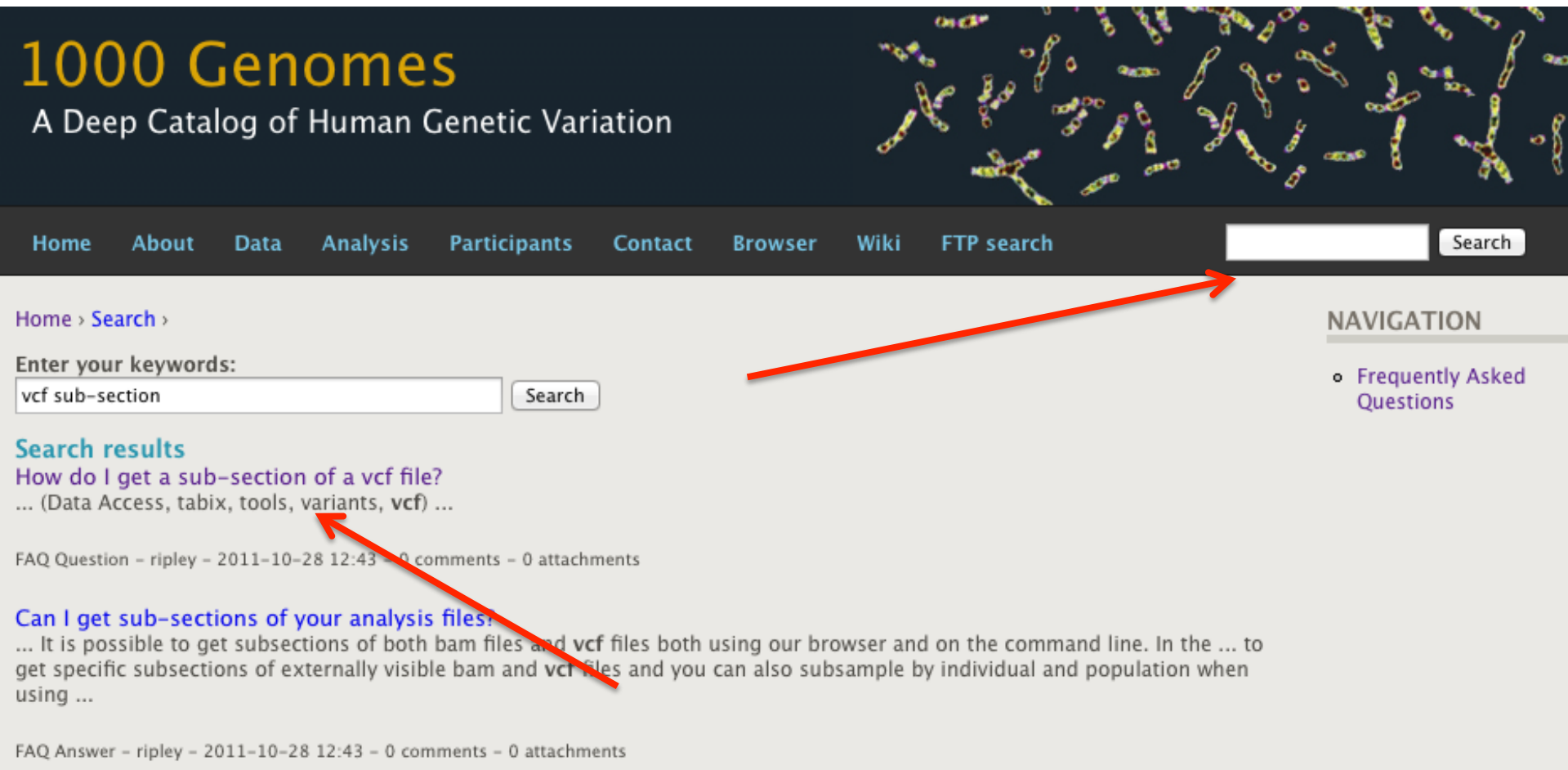
50 files found

File
<a href="ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/working/20120103_omni_shapeit_haplotypes/ALL.chr20.omni_2123_samples_b37_SHAPEIT.20120103.haplotypes.vcf.gz">ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/working/20120103_omni_shapeit_haplotypes/ALL.chr20.omni_2123_samples_b37_SHAPEIT.20120103.haplotypes.vcf.gz</a>
<a href="ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/working/20120103_omni_shapeit_haplotypes/ALL.chr15.omni_2123_samples_b37_SHAPEIT.20120103.haplotypes.vcf.gz">ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/working/20120103_omni_shapeit_haplotypes/ALL.chr15.omni_2123_samples_b37_SHAPEIT.20120103.haplotypes.vcf.gz</a>
<a href="ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/working/20120103_omni_shapeit_haplotypes/ALL.chr4.omni_2123_samples_b37_SHAPEIT.20120103.haplotypes.vcf.gz">ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/working/20120103_omni_shapeit_haplotypes/ALL.chr4.omni_2123_samples_b37_SHAPEIT.20120103.haplotypes.vcf.gz</a>
<a href="ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/working/20120103_omni_shapeit_haplotypes/ALL.chr9.omni_2123_samples_b37_SHAPEIT.20120103.haplotypes.vcf.gz">ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/working/20120103_omni_shapeit_haplotypes/ALL.chr9.omni_2123_samples_b37_SHAPEIT.20120103.haplotypes.vcf.gz</a>
<a href="ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/working/20120103_omni_shapeit_haplotypes/ALL.chr8.omni_2123_samples_b37_SHAPEIT.20120103.haplotypes.vcf.gz">ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/working/20120103_omni_shapeit_haplotypes/ALL.chr8.omni_2123_samples_b37_SHAPEIT.20120103.haplotypes.vcf.gz</a>
<a href="ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/working/20120103_omni_shapeit_haplotypes/ALL.chr12.omni_2123_samples_b37_SHAPEIT.20120103.haplotypes.vcf.gz">ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/working/20120103_omni_shapeit_haplotypes/ALL.chr12.omni_2123_samples_b37_SHAPEIT.20120103.haplotypes.vcf.gz</a>



# Finding Data: Exercise

- Finding help on getting sub-sections of VCF files



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

Home About Data Analysis Participants Contact Browser Wiki FTP search  Search

Home > Search >

Enter your keywords:  
 Search

**Search results**  
[How do I get a sub-section of a vcf file?](#)  
... (Data Access, tabix, tools, variants, vcf) ...

FAQ Question - ripley - 2011-10-28 12:43 - 0 comments - 0 attachments

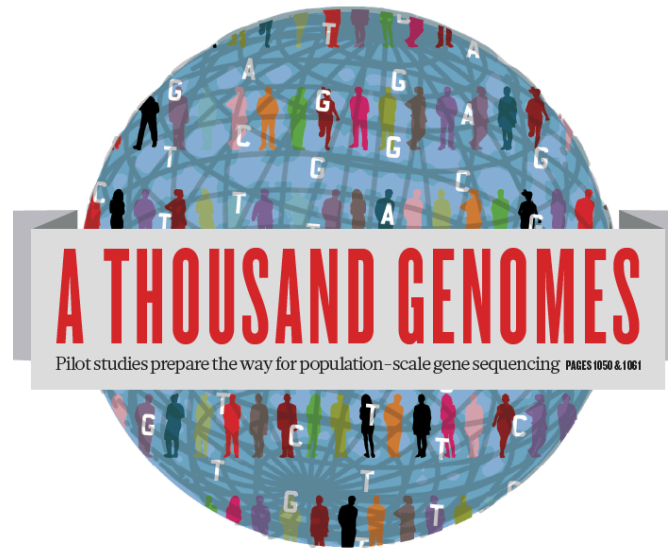
[Can I get sub-sections of your analysis files?](#)  
... It is possible to get subsections of both bam files and vcf files both using our browser and on the command line. In the ... to get specific subsections of externally visible bam and vcf files and you can also subsample by individual and population when using ...

FAQ Answer - ripley - 2011-10-28 12:43 - 0 comments - 0 attachments

**NAVIGATION**

- [Frequently Asked Questions](#)





## The 1000 Genomes Project:

### Exercise 2: Finding Variation Using the Browser





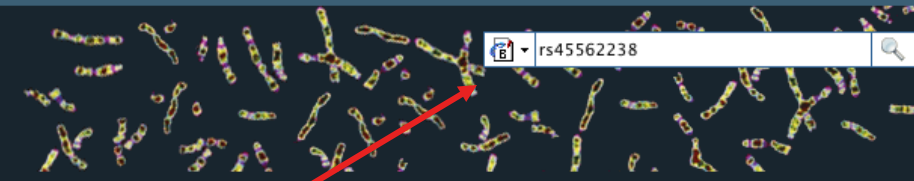
# Exercise: Finding Variation Using the Browser

- Find the variant rs45562238  
*<http://browser.1000genomes.org>*
- In which 1000 Genomes Populations was it detected?
- What are its allele frequencies?
- In which gene is the variant found?



# 1000 Genomes

A Deep Catalog of Human Genetic Variation



rs45562238

Tools | Help

**Search Box**

## Search 1000 Genomes

Go

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

- [View sample data](#)

## 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](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 





Human (GRCh37) ▾

Search 1000 Genomes

New Search

Configure this page

Manage your data

Export data

Get VCF data

Bookmark this page

Results S

You searched for 'rs45562238'

## Gene or Gene Product

0 entrie(s) matched your search strings.

## Genetic Marker

0 entrie(s) matched your search strings.

## Array Probe Set

0 entrie(s) matched your search strings.

## SNP

1 entrie(s) matched your search strings.

1. dbSNP SNP: [rs45562238](#)

SNP Result

## Interpro Domain

0 entrie(s) matched your search strings.

## Gene Family

0 entrie(s) matched your search strings.

## Sequence Aligned to Genome, eg. EST or Protein

0 entrie(s) matched your search strings.

## Genomic Region, eg. Clone or Contig

0 entrie(s) matched your search strings.

1000 Genomes release 9 - September 2011 © [EBI](#)



- Variation displays**
- Flanking sequence
  - Gene/Transcript (1)
  - Population genetics (11)
  - Individual genotypes (2770)
  - Genomic context
  - Phenotype Data
  - Phylogenetic Context
  - External Data
- 
- Configure this page
  - Manage your data
  - Export data
  - Get VCF data
  - Bookmark this page

Variation: rs45562238

Population Genetics

<b>Variation class</b>	SNP ( <a href="#">rs45562238</a> source <a href="#">dbSNP 132</a> - Variants (including SNPs and indels) imported from dbSNP)
<b>Synonyms</b>	OMNI SNP6-133055237 Uniprot <a href="#">VAR_023973</a>
<b>Present in</b>	1000 genomes - Low coverage (1000 genomes - Low coverage - CEU),ALL - interim phase 1 - 1000 Genomes (AMR - interim phase 1 - 1000 Genomes, EUR - interim phase 1 - 1000 Genomes),ENSEMBL:Watson
<b>Alleles</b>	T/C (Ambiguity code: Y)
<b>Ancestral allele</b>	T
<b>Location</b>	This feature maps to 6:133013544 (forward strand)   <a href="#">View in location tab</a>
<b>Validation status</b>	Proven by <b>cluster, frequency, 1000Genome</b>
<b>HGVS names</b> <input type="checkbox"/>	This feature has 3 HGVS names - click the plus to show

Flanking sequence [help](#)

**Flanking Sequence (reference and dbSNP)**

```

AAAAAAAAAAAAAAAAAGATTTAGCAAATAAAATTTAGCCTGTATTTGTAGGCTTCCCAGCTA
AAGTTGAATTTCAAATAAACAACAAAATTTTAGCATACATAAAATCCCAAAAAGTGTGAT
ATGCTTATACTAAAAATCATTTCATTATTTATCTAAACTGTATATTTAACTGAGTGTCTT
TATTTTATCCAGTAACTCTTCTGACATCAAAATATTACCTGTAGATAAATAGCGCCCTCC
ACAGTGTGCAGTCCGTCAAATGCCCTAGAGCGGTACACTTCATTTGGTATGTTCTCAGAC
ATTTTGTAGCTTAAATGACAGCAGAGATCTTTCTGACAAACTGTATAATTTCCCTGCAACT
CCTGTGAGCTTCAAAAAGTGAATTCATCGAAAAAGACAGYGCCTTTAAATTCCTTGT
CCTGATGAGAGCGCTTCTATACTGCTGGCATAGGAAGTCCAGTTCACCACTGCAGAAATGG
GATGGGTGGGAATCCAGTTGCGAGAGGAGGAGTTTCCCTCTCTGTCTTCATATCATAA
TGAAATGCCTTGAAGAATTGGGTGCATAGATGCCACTTCCGGAAGTAATAAGTTGAAA
ACATCATTGAAATGGAAAAGTAAAAAGGGATTTTCATAATTGTTATTACTATTTAACCT
CATATTCACATTTTACTCTCTCTAAGTTATATGTTTTATTAATTTTCAGTAAAAACATGCTG
ATAAAGACAGAGAAGTGAGAGGAAATCAATTAAGTATGAAACCAAGAAGCCAAGAAGGG
AGAGTTGTATGTAAGAAGTCA
    
```

*(Variant highlighted)*



# Gene/Transcript

- Variation displays
  - Flanking sequence
  - Gene/Transcript (1)
  - Population genetics (11)**
  - Individual genotypes (2770)
  - Genomic context
  - Phenotype Data
  - Phylogenetic Context
  - External Data
- Configure this page
- Manage your data
- Export data
- Get VCF data
- Bookmark this page
- Download view as CSV

## Variation: rs45562238

**Variation class** SNP ([rs45562238](#) source [dbSNP 132](#) - Variants (including SNPs and indels) imported from dbSNP [<http://www.ncbi.nlm.nih.gov/projects/SNP/>])

**Synonyms** OMNI SNP6-133055237  
Uniprot [VAR\\_023973](#)

**Present in** 1000 genomes - Low coverage (1000 genomes - Low coverage - CEU),ALL - interim phase 1 - 1000 Genomes (AMR - interim phase 1 - 1000 Genomes, EUR - interim phase 1 - 1000 Genomes),ENSEMBL:Watson

**Alleles** T/C (Ambiguity code: Y)

**Ancestral allele** T

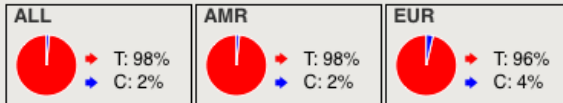
**Location** This feature maps to 6:133013544 (forward strand) | [View in location tab](#)

**Validation status** Proven by **cluster, frequency, 1000Genome**

**HGVS names** This feature has 3 HGVS names - click the plus to show

## Population genetics [help](#)

### 1000 genomes alleles frequencies



## Pie Charts

### 1000 genomes

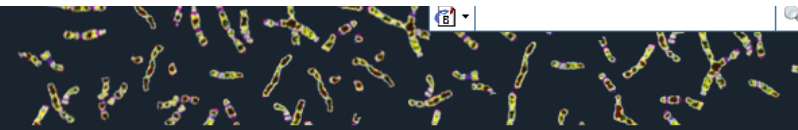
Show/hide columns Filter

Population	Alleles C	Alleles T	Genotypes CIT	Genotypes TIT	Allele count	Genotype count	Genotype detail
1000GENOMES:ALL	0.016	0.984	0.031	0.969	34 (C) / 2154 (T)	34 (CIT) / 1060 (TIT)	<a href="#">Show</a>
1000GENOMES:AMR	0.017	0.983	0.033	0.967		6 (CIT) / 175 (TIT)	<a href="#">Show</a>
1000GENOMES:EUR	0.037	0.963	0.073	0.927		28 (CIT) / 353 (TIT)	<a href="#">Show</a>



# 1000 Genomes

A Deep Catalog of Human Genetic Variation



Human (GRCh37) Location: 6:133,013,044-133,014,044 Variation: rs45562238

Tools | Help

- Variation displays**
- Flanking sequence
  - Gene/transcript (1)**
  - Population genetics (11)
  - Individual genotypes (2770)
  - Genomic context
  - Phenotype Data
  - Phylogenetic Context
  - External Data
- Configure this page
- Manage your data
- Export data
- Get VCF data
- Bookmark this page
- Download view as CSV

## Variation: rs45562238

**Variation class** SNP ([rs45562238](#) source [dbSNP 132](#) - Variants (including SNPs and indels) imported from dbSNP [<http://www.ncbi.nlm.nih.gov/projects/SNP/>])

**Synonyms** OMNI SNP6-133055237  
Uniprot [VAR\\_023973](#)

**Present in** 1000 genomes - Low coverage (1000 genomes - Low coverage - CEU),ALL - interim phase 1 - 1000 Genomes (AMR - interim phase 1 - 1000 Genomes),ENSEMBL:Watson

**Alleles** T/C (Ambiguity code: Y)

**Ancestral allele** T

**Location** This feature maps to 6:133013544 (forward strand) | [View in location tab](#)

**Validation status** Proven by **cluster, frequency, 1000Genome**

**HGVS names** This feature has 3 HGVS names - click the plus to show

Sift/PolyPhen

## Gene/transcript [help](#)

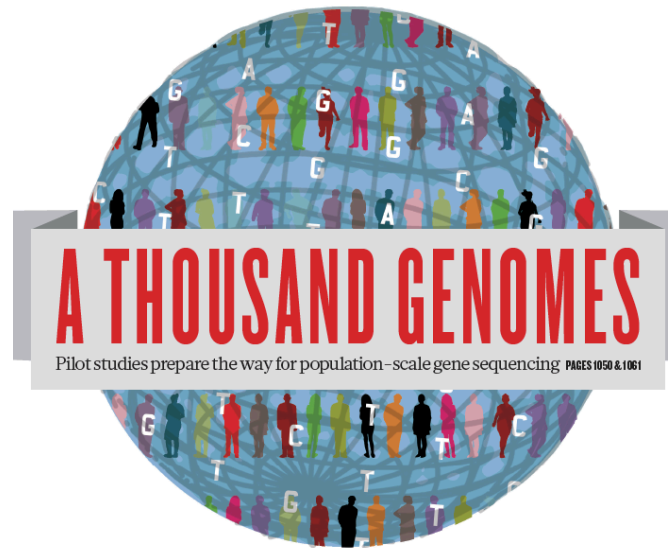
Show/hide columns

Gene	Transcript (strand)	Allele (transcript allele)	Type	HGVS names	Position in transcript	Position in CDS	Position in protein	Amino acid	Codons	SIFT	PolyPhen
<a href="#">ENSG00000112299</a>	<a href="#">ENST00000367928</a>	(-) C (G)	Non-synonymous coding	ENST00000367928.4:c.1006A>G ENSP00000356905.4:p.Thr336Ala	1020	1006	336	T/A	ACT/GCT	tolerated	benign

1000 Genomes release 9 - September 2011 © [EBI](#)

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





# The 1000 Genomes Project:

## Exercise 3: Using 1000 Genomes Tools



# Exercise: Using 1000 Genomes Tools

- Find the gene SLC44A4 using the search box on <http://browser.1000genomes.org>
- Get a VCF file for this Gene using the Get VCF button.
- Uncompress this file
  - You can get a copy at [ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/working/20120120\\_1000genomes\\_tutorial/6.31830969-31846823.ALL.chr6.phase1.projectConsensus.genotypes.vcf](ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/working/20120120_1000genomes_tutorial/6.31830969-31846823.ALL.chr6.phase1.projectConsensus.genotypes.vcf)
- Use this file with the Variant Effect Predictor
  - [http://browser.1000genomes.org/Homo\\_sapiens/UserData/UploadVariations](http://browser.1000genomes.org/Homo_sapiens/UserData/UploadVariations)
- Do any of the variants have deleterious effects according to SIFT or PolyPhen
- Use the example url on the page and the coordinates 6:31830700-31840700 with the Variation Pattern Finder
  - [http://browser.1000genomes.org/Homo\\_sapiens/UserData/VariationsMapVCF](http://browser.1000genomes.org/Homo_sapiens/UserData/VariationsMapVCF)





# 1000 Genomes

A Deep Catalog of Human Genetic Variation



SLC44A4

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.

## 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's on this site are temporary and will NOT be maintained.**

### Links

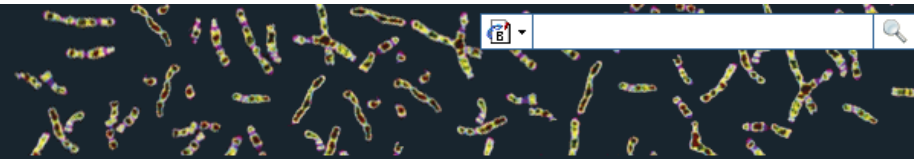


[1000 Genomes](#) →  
More information about the 1000 Genomes Project on the 1000 genomes main site.



[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.





Human (GRCh37) ▾

**Search 1000 Genomes**

New Search

Configure this page

Manage your data

Export data

Get VCF data

Bookmark this page

## Results Summary

You searched for 'SLC44A4'

### Gene or Gene Product

10 entrie(s) matched your search strings.

1. Gene: [ENSG00000204385](#) [Region in detail]  
SLC44A4 - solute carrier fam 44, member 4 [Source:HGNC Symbol;Acc:13941]
2. Transcript: [ENST00000229729](#) [Region in detail]
3. Peptide: [ENSP00000398764](#) [Region in detail]  
SLC44A4
4. Peptide: [ENSP00000392054](#) [Region in detail]  
SLC44A4
5. Peptide: [ENSP00000404572](#) [Region in detail]  
SLC44A4
6. Peptide: [ENSP00000398901](#) [Region in detail]  
SLC44A4
7. Peptide: [ENSP00000415708](#) [Region in detail]  
SLC44A4
8. Peptide: [ENSP00000400263](#) [Region in detail]  
SLC44A4
9. Peptide: [ENSP00000414296](#) [Region in detail]  
SLC44A4
10. Peptide: [ENSP00000399161](#) [Region in detail]  
SLC44A4

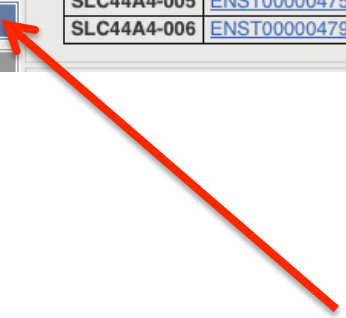


- Gene-based displays
- Gene summary
- Splice variants (9)
- Supporting evidence
- Sequence
- External references
- Regulation
- Genetic Variation
  - Variation Table
  - Structural Variation
  - Variation Image
- External Data
- ID History
  - Gene history
- Configure this page
- Manage your data
- Export data
- Get VCF data
- Bookmark this page

### Gene: SLC44A4 (ENSG00000204385)

**Description** solute carrier family 44, member 4 [Source:HGNC Symbol;Acc:13941]  
**Location** [Chromosome 6: 31,830,969-31,846,823](#) reverse strand.  
**Transcripts**  There are 9 transcripts in this gene

Name	Transcript ID	Length (bp)	Protein ID	Length (aa)	Biotype	CCDS
SLC44A4-001	<a href="#">ENST00000229729</a>	2589	<a href="#">ENSP00000229729</a>	710	Protein coding	<a href="#">CCDS4724</a>
SLC44A4-004	<a href="#">ENST00000414427</a>	1233	<a href="#">ENSP00000398901</a>	411	Protein coding	-
SLC44A4-201	<a href="#">ENST00000375562</a>	2505	<a href="#">ENSP00000364712</a>	668	Protein coding	-
SLC44A4-202	<a href="#">ENST00000544672</a>	2634	<a href="#">ENSP00000444109</a>	634	Protein coding	-
SLC44A4-002	<a href="#">ENST00000465707</a>	681	No protein product	-	Processed transcript	-
SLC44A4-003	<a href="#">ENST00000462671</a>	426	No protein product	-	Processed transcript	-
SLC44A4-007	<a href="#">ENST00000487680</a>	392	No protein product	-	Processed transcript	-
SLC44A4-005	<a href="#">ENST00000475563</a>	575	No protein product	-	Retained intron	-
SLC44A4-006	<a href="#">ENST00000479777</a>	655	No protein product	-	Retained intron	-



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**

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

Click [here](#) for more extensive documentation.

VCF / BAM File URL:

e.g. ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20101123/interim\_phase1\_release/ALL.chr1.phase1.projectConsensus.genotypes.vcf.gz

Region:

( e.g. 1:1-50000 )

Use VCF filters (this doesn't apply to BAM files):

- None
- By individual(s)
- By population(s) \*

(to filter by populations please provide URL to a Sample-Population Mapping File in the box below)

Sample-Population Mapping File URL:

e.g. ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20101123/interim\_phase1\_release/interim\_phase1.20101123.ALL.panel



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

Thank you - your VCF file [\[6.31830969-31846823.ALL.chr6.phase1.projectConsensus.genotypes.vcf.gz\]](#) [Size: 83436] has been generated. Right click on the file name and choose "Save link as ..." from the menu

Preview

```
##fileformat=VCFv4.0
##source=BCM:SNPTools:hapfuse
##reference=1000Genomes-NCBI37
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=AP,Number=2,Type=Float,Description="Allelic Probability, P(Allele=1
#CHROM POS ID REF ALT QUAL FILTER INFO FORMAT HG00096 1
6 31831159 rs3869144 C T 100 PASS .
6 31831167 . T C 100 PASS . GT:AP
```



### Input file

Species: Human (Homo sapiens): GRCh37 ▾

Name for this upload (optional): SLC44A4

Paste file: 

Upload file: /Users/laura/Downloads/6.

or provide file URL:

Input file format: VCF ▾

### Options

Get regulatory region consequences:

Type of consequences to display: Ensembl terms ▾

Check for existing co-located variants: Yes ▾

Return results for variants in coding regions only:

Show HGNC identifier for genes where available:

Show Ensembl protein identifiers where available:

Show HGVS identifiers for variants where available: No ▾

### Non-synonymous SNP predictions (human only)

SIFT predictions: Prediction only ▾

PolyPhen predictions: Prediction only ▾

Condel consensus (SIFT/PolyPhen) predictions: No ▾

6_31833249_A/G	<a href="#">6:31833249</a>	G	<a href="#">ENSG00000204385</a>	<a href="#">ENST00000487680</a>	Transcript	UPSTREAM	-	-	-	-
6_31833249_A/G	<a href="#">6:31833249</a>	G	<a href="#">ENSG00000204385</a>	<a href="#">ENST00000414427</a>	Transcript	DOWNSTREAM	-	-	-	-
6_31833249_A/G	<a href="#">6:31833249</a>	G	<a href="#">ENSG00000204385</a>	<a href="#">ENST00000479777</a>	Transcript	DOWNSTREAM	-	-	-	-
6_31833249_A/G	<a href="#">6:31833249</a>	G	<a href="#">ENSG00000204385</a>	<a href="#">ENST00000475563</a>	Transcript	DOWNSTREAM	-	-	-	-
6_31833357_C/T	<a href="#">6:31833357</a>	T	-	<a href="#">ENSR00000487922</a>	RegulatoryFeature	REGULATORY_REGION	-	-	-	-
6_31833357_C/T	<a href="#">6:31833357</a>	T	<a href="#">ENSG00000204386</a>	<a href="#">ENST00000495807</a>	Transcript	UPSTREAM	-	-	-	-
6_31833357_C/T	<a href="#">6:31833357</a>	T	<a href="#">ENSG00000204386</a>	<a href="#">ENST00000480384</a>	Transcript	UPSTREAM	-	-	-	-
6_31833357_C/T	<a href="#">6:31833357</a>	T	<a href="#">ENSG00000204386</a>	<a href="#">ENST00000491768</a>	Transcript	UPSTREAM	-	-	-	-
6_31833357_C/T	<a href="#">6:31833357</a>	T	<a href="#">ENSG00000204386</a>	<a href="#">ENST00000375631</a>	Transcript	UPSTREAM	-	-	-	-
6_31833357_C/T	<a href="#">6:31833357</a>	T	<a href="#">ENSG00000204386</a>	<a href="#">ENST00000479533</a>	Transcript	UPSTREAM	-	-	-	-
6_31833357_C/T	<a href="#">6:31833357</a>	T	<a href="#">ENSG00000204385</a>	<a href="#">ENST00000229729</a>	Transcript	NON_SYNONYMOUS_CODING	1625	1604	535	R/H
6_31833357_C/T	<a href="#">6:31833357</a>	T	<a href="#">ENSG00000204385</a>	<a href="#">ENST00000375562</a>	Transcript	NON_SYNONYMOUS_CODING	1544	1478	493	R/H
6_31833357_C/T	<a href="#">6:31833357</a>	T	<a href="#">ENSG00000204385</a>	<a href="#">ENST00000544672</a>	Transcript	NON_SYNONYMOUS_CODING	1673	1376	459	R/H
6_31833357_C/T	<a href="#">6:31833357</a>	T	<a href="#">Ei</a>	-	-	<a href="#">1KG 6 31833357</a>	-	-	-	-
6_31833357_C/T	<a href="#">6:31833357</a>	T	<a href="#">Ei</a>	-	-	<a href="#">1KG 6 31833357</a>	-	-	-	-
6_31833357_C/T	<a href="#">6:31833357</a>	T	<a href="#">Ei</a>	-	-	<a href="#">1KG 6 31833357</a>	-	-	-	-
6_31833357_C/T	<a href="#">6:31833357</a>	T	<a href="#">Ei</a>	535	R/H	cGc/cAc	<a href="#">1KG 6 31833357</a>	<b>SIFT=deleterious;</b>	<b>PolyPhen=probably_damaging;</b>	<b>Condel=deleterious</b>
6_31833612_C/G	<a href="#">6:31833612</a>	G	-	-	-	-	-	-	-	-
6_31833612_C/G	<a href="#">6:31833612</a>	G	<a href="#">Ei</a>	493	R/H	cGc/cAc	<a href="#">1KG 6 31833357</a>	<b>SIFT=deleterious;</b>	<b>PolyPhen=possibly_damaging;</b>	<b>Condel=deleterious</b>
6_31833612_C/G	<a href="#">6:31833612</a>	G	<a href="#">Ei</a>	459	R/H	cGc/cAc	<a href="#">1KG 6 31833357</a>	<b>SIFT=deleterious;</b>	<b>PolyPhen=probably_damaging;</b>	<b>Condel=deleterious</b>
6_31833612_C/G	<a href="#">6:31833612</a>	G	<a href="#">Ei</a>	-	-	-	<a href="#">1KG 6 31833357</a>	-	-	-
				-	-	-	<a href="#">1KG 6 31833357</a>	-	-	-
				-	-	-	<a href="#">1KG 6 31833357</a>	-	-	-

## 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**

**i Variation Pattern Finder:**

The Variation Pattern Finder allows one to look for patterns of shared variation between individuals in the same vcf file. The finder looks for distinct variation combinations within the region, as well as individuals associated with each variation combination pattern. Only variants which have potentially functional consequences are considered, both intergenic and intronic snps are excluded. Click [here](#) for more extensive documentation.

The search will be performed on any VCF file you provided. It should be a URL for the file location. Please refer to <http://vcftools.sourceforge.net/specs.html> for VCF format specification. A URL for the latest VCF file for variation calls and genotypes released by the 1000 Genomes Project is displayed as an example below the input box. A mapping file between individual sample and population is required as well. The latest mapping file between individual sample and population released by the 1000 Genomes Project is displayed as well below the input box.

**Upload files****VCF File URL:**[Clear box](#)

e.g. ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20101123/interim\_phase1\_release/ALL.chr6.phase1.projectConsensus.genotypes.vcf.gz

**Sample-Population Mapping File URL:**[Clear box](#)

e.g. ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20101123/interim\_phase1\_release/interim\_phase1.20101123.ALL.panel

**Region:**

e.g. 6:46620015-46620998

**Next >**





# Any questions?

- Please email [info@1000genomes.org](mailto:info@1000genomes.org) if you have any questions or feedback about this resource.



# Thanks

- The 1000 Genomes Project Consortium
- Paul Flicek, Laura Clarke
- Richard Smith, Holly Zheng Bradley and Ian Streeter
- Giulietta Spudich and Bert Overduin

