



The 1000 Genomes Browser

Graphical visualization of genotype data from the 1000 Genomes Project

<https://www.ncbi.nlm.nih.gov/variation/tools/1000genomes>

National Center for Biotechnology Information • National Library of Medicine • National Institutes of Health • Department of Health and Human Services

Scope

The 1000 Genomes Project aims to produce an extensive public catalog of human genetic variation, including SNPs, structural variants, and their haplotype contexts [1]. The genomes of more than 2500 de-identified people from 25 populations around the world are being sequenced using next-generation sequencing technologies with results of the study freely accessible to researchers worldwide. The data files from this project are mirrored at NCBI through FTP, Aspera, and Amazon Cloud:

- <ftp.ncbi.nlm.nih.gov/1000genomes/>
- www.ncbi.nlm.nih.gov/public/1000genomes/
- s3.amazonaws.com/1000genomes



The **1000 Genomes Browser**, developed from NCBI's Graphical Sequence Viewer (SV) [2], addresses the challenge posed by the influx of this large set of data. This tool provides a convenient display for users to visualize the 1000 Genomes Project data and analyses by mapping the variation data to the reference genome assembly and allowing the retrieval and interactive examination of genotype data for a specific gene or genomic region. Through comparison with existing genome annotations, users can correlate the genotype and population prevalence with gene function.

Interface for data access and display

The 1000 Genomes Browser (www.ncbi.nlm.nih.gov/variation/tools/1000genomes/) enables the selection and display of a selected chromosome (A). Keyword searching (B) zooms the display to the specific location, showing annotation and variation tracks (C), and listing detailed variations with underlying genotype in the table below (D). The "Downloads" widget (E) exports the genotype and alignment data for the selected genomic region. With the "Your data" widget (F), users can upload custom data for private display in the graphical panel. The "Share this page" (G) creates a unique URL for linking to the customized display. Links to FAQ and Help (H) provide additional details about the function of this tool. The browser displays data from the Phase 3 call set. A link at the top (I) provides access to the Phase 1 data.

The screenshot shows the 1000 Genomes Browser interface. At the top, it displays 'NCBI Resources How To Sign in to NCBI' and '1000 Genomes Browser Phase 3'. The main title is 'Homo sapiens: GRCh37.p13 (GCF_000001405.25) Chr 1 (NC_000001.10): 1 - 249.3M'. Below this, there are links for 'Reset All', 'Share this page', 'FAQ', 'Help', and 'Version 3.4'. A warning message states: 'ATTENTION: You are browsing the alignment and genotype data from the Phase 3 May 2013 call set. Data from an earlier release is also available. Click here to browse data from the Phase 1 March 2012 call set.' The interface is divided into several sections: 'Ideogram View' on the left showing chromosomes 1-22, X, and Y; 'Search' with a text input field; 'Subjects', 'Your Data', 'Region Details', 'History', and 'Downloads' on the left sidebar. The main panel shows 'Exon Navigator' with a warning 'There are too many genes in the region (3817). Please narrow the region to enable exon navigation.' Below this are tracks for 'Segmental Duplications on GRCh37', '1000 Genomes Phase 1 Strict Accessibility Mask', 'Genes, NCBI Homo sapiens Annotation Release 105' (listing genes like MTOR, JUN, CRP, etc.), 'ClinVar Short Variations based on dbSNP 142', 'dbSNP 141 (Homo sapiens Annotation Release 105) HapMap Recombination Rate', and '1000 Genomes Phase 3, dbSNP b142'. At the bottom, a table shows 'Go to Selection' and 'Scroll Region' with coordinates and 'Populations / Samples' with allele frequencies for ACB and ASW. A 'Genotype Display Options' dropdown is on the right.

Searching with custom terms

The search box at the left (A) provides a way to quickly navigate to the targeted genomic region of interest. Clicking the “Search examples” (B, activated upon hovering) expands the section to display example query formats. Entering gene symbol “ptpn22” and pressing the return key retrieves a list of matches (C). Hover over an entry to activate the navigation arrow (D), which zooms the viewer to the target region (E) and simultaneously updates the genotype table (F). The gene region displayed is shown at the top (G). The exon navigator (H) allows zooming in to specific exons through a single radio button click, or flipping through exons using the arrows.

Go to Selection	Scroll Region	114,385,308	114,385,478	114,385,519	114,385,570	114,385,623	114,385,669	114,385,679	114,385,770	114,385,829	114,385,908	114,385,952	114,385,999
rs143959041								rs71862845	rs118543141	rs118822609	rs71578795		rs189521575
Populations / Samples		TA...=-0.9521	C=-0.9992	C=-0.9998	G=-0.9996	G=-0.9998	C=-0.9996	G=-0.8053	C=-0.9862	C=0.9964	ATG=-0.3421	T=-0.9998	G=-0.9998
Show: Allele frequencies		T=0.0479	T=0.0008	T=0.0002	A=0.0004	A=0.0002	G=0.0004	A=0.1947	T=0.0138	A=0.0036	AT...=0.2742	C=0.0002	A=0.0002
ACB African Caribbeans ...		TA...=-0.8333	C=-1.0000	C=-1.0000	G=-1.0000	G=-1.0000	C=-0.9896	G=-0.9792	C=-0.9531	C=-1.0000	ATG=-0.2031	T=-1.0000	G=-1.0000
		T=0.1667	T=0.0000	T=0.0000	A=0.0000	A=0.0000	G=0.0104	A=0.0208	T=0.0469	A=0.0000	AT...=0.5104	C=0.0000	A=0.0000

Examining detailed genotypes for specific subjects

Clicking the “Subjects” arrow and the “Tracks in view” arrow (I) expands the section to provide access to read alignment data supporting the genotype calls of specific subjects (individuals). This section can be resized by dragging on its sides.

Track	Aligner	Model	Platform
<input checked="" type="checkbox"/> HG00186 exome (SRR1596708)	bwa	Illumina HiSeq 2000	ILLUMINA
<input checked="" type="checkbox"/> HG00284 exome (SRR1599373)	bwa	Illumina HiSeq 2000	ILLUMINA

Track	Columns	Search:
<input checked="" type="checkbox"/> HG00284 exome (SRR1599373)	<input checked="" type="checkbox"/> Track	
<input type="checkbox"/> HG00308 exome (SRR1602474)	<input type="checkbox"/> Method	
<input checked="" type="checkbox"/> HG00186 exome (SRR1596708)	<input type="checkbox"/> BioProject ID	
<input checked="" type="checkbox"/> HG00277 exome (SRR1596751)	<input checked="" type="checkbox"/> Aligner	
	<input type="checkbox"/> Alignment type	
	<input type="checkbox"/> Population abbr	

An example usage case

Data from many publicly funded genome-wide association studies are available in the dbGaP database at NCBI. The Phenotype and Genotype-Integrator tool ([PheGenI](#), **A**) [3] provides a convenient way to get the genes and SNPs associated with studied phenotypes or disorders. The following example begins using PheGenI to identify candidate genes and SNPs associated with rheumatoid arthritis (relevant PubMed abstracts: <http://1.usa.gov/K7h6Vi>). The retrieved variations are used to examine the genotypes in different populations in data generated by the 1000 Genomes Project.

Searching in PheGenI

Selecting the “Arthritis, Rheumatoid” trait with a p-value cutoff at 10^{-30} (**B**) retrieves a list of genes and SNPs strongly associated with the disorder. The “Association Results” section (**C**) lists these SNPs (rs#) and genes in descending order of significance. The nature of listed SNPs is available in the “Context” column (**D**). Among this group, rs2476601 from PTPN22 is the only missense SNP. The rsID (**E**) is hyper-linked to the SNP record, where a direct link to 1000 Genomes Browser (**F**) allows the interactive examination.

Examining the SNP in the 1000 Genomes Browser

Clicking the magnifying glass icon (**F**) in the SNP record displays a zoomed-in view in the 1000 Genomes Browser, which highlights the target SNP with an added marker (**G**). The genotype table also highlights this SNP by coloring the column in yellow (**H**). SNPs outside the graphic display have their columns colored grey (**I**). Genotyping statistics for a SNP are in the popup (**I**) activated upon hovering over the column header.

Welcome to PheGenI YouTube Tutorial

The Phenotype-Genotype Integrator (PheGenI), merges NHGRI genome-wide association study (GWAS) catalog data with several databases housed at the National Center for Biotechnology Information (NCBI), including Gene, dbGaP, OMIM, GTEx and dbSNP.

Search Criteria

Search Clear Examples...

Phenotype Selection **B**

Traits: arthritis

Arthritis, Juvenile Rheumatoid
Arthritis, Psoriatic
Arthritis, Rheumatoid

Browse...

P-Value: $< 1 \times 10^{-30}$ Source: [Any]

Genotype Selection **i**

Location Gene SNP

Chromosome:

Range (bps):

SNP Functional Class

exon intron nearegene UTR

Clear Invert

Association Results **C** **D**

1 - 10 of 10 Download Modify Search

#	Trait	rs #	Context	Gene	Location	P-value	Source	Study	PubMed
1	Arthritis, Rheumatoid	rs6910071	intron	C6orf10	6: 32,282,854	1.000×10^{-299}	NHGRI		20453842
2	Arthritis, Rheumatoid	rs6457620	intergenic	HLA-DQB1, HLA-DOA2	6: 32,663,999	4.000×10^{-186}	NHGRI		18794853
3	Arthritis, Rheumatoid	rs9268853	intergenic	HLA-DRB9, HLA-DRB5	6: 32,429,643	5.000×10^{-142}	NHGRI		21653640
4	Arthritis, Rheumatoid	rs660895	intergenic	HLA-DRB1, HLA-DOA1	6: 32,577,380	1.000×10^{-108}	NHGRI		17804836
5	Arthritis, Rheumatoid	rs6457617	intergenic	HLA-DQB1, HLA-DOA2	6: 32,663,851	5.000×10^{-12}	NHGRI		17554300
6	Arthritis, Rheumatoid	rs2476601	missense	PTPN22	1: 114,377,568	9.000×10^{-24}	NHGRI		20453842
7	Arthritis, Rheumatoid	rs13192471	intergenic	HLA-DQB1, HLA-DOA2	6: 32,671,103	2.000×10^{-58}	NHGRI		20453841
8	Arthritis, Rheumatoid	rs9272219	intergenic	HLA-DRB1, HLA-DOA1	6: 32,602,269	1.000×10^{-45}	NHGRI		21653640

Genotype Table

Assembly	Annotation Release	Chr	Chr Pos	Contig	Contig Pos	SNP to Chr	Contig allele	Contig to Chr	Neighbor SNP	Map Method
GRCh38.p2	107	1	113834946	NT_032977.10	113248958	Fwd	A	Fwd	view	mapup
GRCh37.p13	105	1	114377568	NT_032977.9	84349486	Fwd	A	Fwd	view	blast

Region PTPN22 NM_012411.4

NC_000001.10: 114M..114M (260bp)

1000 Genomes Phase 1 Strict Accessibility Mask

Genes, NCBI Homo sapiens Annotation Release 105

ClinVar Short Variations based on dbSNP 142 (Homo sapiens Annotation Release 105)

dbSNP 141 (Homo sapiens Annotation Release 105) HapMap Recombination R

1000 Genomes Phase 3, dbSNP b142

SNP Detail **F** **G**

Position: 114377568 Quality: 100
ID: rs2476601

AA	G	DP
AC	4932	EAS_AF 1
AF	0.972644	EUR_AF 0.9056
AFR_AF	0.997	NS 2504
AMR_AF	0.964	SAS_AF 0.9867
AN	5070	SF 0.1

Filter: PASS

Genotype Display Options

Study: The 1000 Genomes Project (phase 3)

Populations / Samples	G=0.9998	C=0.9992	T=0.9998	G=0.9998	C=0.9998	C=0.9998	G=0.9998	C=0.9998	A=0.0274	C=0.9998	A=0.9998
Show: Allele frequencies	A=0.0002	G=0.0008	C=0.0002	A=0.0002	T=0.0002	T=0.0002	C=0.0002	T=0.0002	G=0.9726	A=0.0002	G=0.0002
ACB African Carribeans ...	G=1.0000	C=1.0000	T=1.0000	G=1.0000	C=1.0000	C=1.0000	G=1.0000	C=1.0000	A=0.0052	C=1.0000	A=1.0000
	A=0.0000	G=0.0000	C=0.0000	A=0.0000	T=0.0000	T=0.0000	C=0.0000	T=0.0000	G=0.9948	A=0.0000	G=0.0000
ASW Americans of African...	G=1.0000	C=1.0000	T=1.0000	G=1.0000	C=1.0000	C=1.0000	G=1.0000	C=1.0000	A=0.0164	C=1.0000	A=1.0000
	A=0.0000	G=0.0000	C=0.0000	A=0.0000	T=0.0000	T=0.0000	C=0.0000	T=0.0000	G=0.9836	A=0.0000	G=0.0000

An example usage case (cont.)

The “Genotype” table (shown below) provides the detailed genotype data. Its top row (A) displays the variation calls and their genomic coordinates. The “Populations/Samples” row (B) lists the global allele frequencies (click the arrow to adjust to counts), while rows below (C) list the allele frequency for individual populations. Clicking a population cell (D) toggles open the individual genotypes section for that population. Clicking the checkbox (E) next to an individual triggers the read selector popup from which alignments can be selected for display in the graphical view, allowing for a more detailed examination (F). Upon “Zoom to Sequence at Marker” menu selection (G), aligned reads from the selected individual provide supporting evidence for genotype calls at this position. The prevalence of rheumatoid arthritis reported for different populations from published studies is generally consistent with data from the 1000 Genomes study. This further supports the conclusion of the reported GWAS studies. Two such published reports are available from PubMed (<http://1.usa.gov/K7h6Vi>).

The screenshot displays the 1000 Genomes Browser interface for the marker rs2476601. The main table shows allele frequencies for various populations, including ACB African Caribbeans, CHS Southern Han Chinese, and FIN Finnish in Finland. The table columns represent different alleles (G, C, T, A) and their frequencies. A dropdown menu for 'Genotype Display Options' is visible at the top right. Below the table, a list of individual samples is shown, with checkboxes for selecting specific individuals. A red arrow points to the checkbox for HG00186 (SAMN00016977), which is highlighted. A red arrow also points to the 'Zoom to Sequence at Marker' menu option. The graphical view below shows aligned reads for the selected individual, with a red arrow pointing to the read selector popup. The read selector popup shows the sequence alignment for the selected individual, with a red arrow pointing to the 'Zoom to Sequence at Marker' menu option.

Table 1: Global Allele Frequencies (Populations/Samples)

Population	G	C	T	A
Populations / Samples	G=0.9998	C=0.9992	T=0.9998	A=0.0002
ACB African Caribbeans	G=1.0000	C=1.0000	T=1.0000	A=0.0000
CHS Southern Han Chinese	G=1.0000	C=1.0000	T=1.0000	A=0.0000
FIN Finnish in Finland	G=1.0000	C=1.0000	T=1.0000	A=0.0000

Table 2: Individual Genotypes (Sample HG00186)

Sample	G	C	T	A
HG00186 (SAMN00016977)	G	C	T	A
HG00186 exome (SRR1596708)	G	C	T	A
HG00186 low_coverage (SRR1596709)	G	C	T	A

Table 3: Read Selector Popup (HG00186 exome)

Read	Sequence
1	T C C A A C A G G A A G
2	T C C A A C A G G A A G
3	T C C A A C A G G A A G
4	T C C A A C A G G A A G
5	T C C A A C A G G A A G
6	T C C A A C A G G A A G
7	T C C A A C A G G A A G
8	T C C A A C A G G A A G
9	T C C A A C A G G A A G
10	T C C A A C A G G A A G
11	T C C A A C A G G A A G
12	T C C A A C A G G A A G
13	T C C A A C A G G A A G
14	T C C A A C A G G A A G
15	T C C A A C A G G A A G
16	T C C A A C A G G A A G
17	T C C A A C A G G A A G
18	T C C A A C A G G A A G
19	T C C A A C A G G A A G
20	T C C A A C A G G A A G
21	T C C A A C A G G A A G
22	T C C A A C A G G A A G
23	T C C A A C A G G A A G
24	T C C A A C A G G A A G
25	T C C A A C A G G A A G
26	T C C A A C A G G A A G
27	T C C A A C A G G A A G
28	T C C A A C A G G A A G
29	T C C A A C A G G A A G
30	T C C A A C A G G A A G
31	T C C A A C A G G A A G
32	T C C A A C A G G A A G
33	T C C A A C A G G A A G
34	T C C A A C A G G A A G
35	T C C A A C A G G A A G
36	T C C A A C A G G A A G
37	T C C A A C A G G A A G
38	T C C A A C A G G A A G
39	T C C A A C A G G A A G
40	T C C A A C A G G A A G
41	T C C A A C A G G A A G
42	T C C A A C A G G A A G
43	T C C A A C A G G A A G
44	T C C A A C A G G A A G
45	T C C A A C A G G A A G
46	T C C A A C A G G A A G
47	T C C A A C A G G A A G
48	T C C A A C A G G A A G
49	T C C A A C A G G A A G
50	T C C A A C A G G A A G

References

1. Sabeti PC, *et al.* (2007). Genome-wide detection and characterization of positive selection in human populations. *Nature*. 449(7164): 913-8. [PMID: 17943131](https://pubmed.ncbi.nlm.nih.gov/17943131/)
2. Graphical Sequence Viewer factsheet: ftp.ncbi.nlm.nih.gov/pub/factsheets/Factsheet_Graphical_SV.pdf
3. Phenotype Genotype Integrator factsheet: ftp.ncbi.nlm.nih.gov/pub/factsheets/Factsheet_PheGenI.pdf
4. README for API access to data behind 1000 Genomes browser: <https://www.ncbi.nlm.nih.gov/projects/genotypes/cgi/query.cgi?cmd=readme>