

The 1000 Genomes Web based Tutorial Exercises.

These are the answers for the Web Based Tutorial Exercises. Please note these are our recommended ways of doing these tasks but there may be other solutions too.

As mentioned in the Exercises document these exercises require you to have haploview installed to be able to complete them.

Haploview is available from: http://www.broadinstitute.org/scientificcommunity/science/programs/medical-and-populationgenetics/haploview/downloads

Finding Data

1a. Find what Omni VCF files we have on our ftp site using the website ftp search.

				<u> </u>	No mar	<i></i>	01854c	~ ~
Home About Data	Analysis Partio	cipants Contac	t Browser	Wiki	FTP search		Search	
Home >							LAURA@EBI.AC.UK	
SEARCH 1000 GE	NOMES FTP FI	_ES					 My account 	
Search term:							Create content	
omni*vcf Search for files on the FTP site							List content List users	
Help on searching							 Manage files Log out 	
- Search options							Frequently Asked Ouestions	
Use NCBI FTP site							Questions	
Dump MD5LIST								
✓ Exclude FASTO files								
✓ Exclude BAM files								
Exclude pilot data	_							
Ozhu zilet dete								
Exclude index files								
 Exclude any .bai, .bas 	or .tbi file							
Search		-						
arch term: nni*vcf								
rch for files on the FTP site								
Ip on searching								
Search options								
earch								
							-	
ESULTS								
0 files found								
ile								
ftp://ftp.1000genomes.	.ebi.ac.uk/vol1/ftp/	technical/workir	ig/20120103	B_omni_s	hapeit_haplotypes			
ftp://ftp.1000genomes. /ALL.chr15.omni 2123	.ebi.ac.uk/vol1/ftp/ samples_b37_SHAP	technical/workir	ig/20120103	gz 8_omni_9 f.az	hapeit_haplotypes			
ftp://ftp.1000genomes. /ALL.chr4.omni 2123 s	.ebi.ac.uk/vol1/ftp/ amples_b37_SHAPE	technical/workir IT.20120103.haj	g/20120103	3_omni_s gz	hapeit_haplotypes			
ftp://ftp.1000genomes. /ALL.chr9.omni_2123_s	.ebi.ac.uk/vol1/ftp/ amples_b37_SHAPE	technical/workir IT.20120103.haj	g/20120103 plotypes.vcf.	3_omni_s gz	hapeit_haplotypes			
ftp://ftp.1000genomes. /ALL.chr8.omni_2123_s	.ebi.ac.uk/vol1/ftp/ amples_b37_SHAPE	technical/workir IT.20120103.haj	g/20120103 plotypes.vcf.	3_omni_s gz	hapeit_haplotypes			
ftp://ftp.1000genomes. /ALL.chr12.omni_2123_	.ebi.ac.uk/vol1/ftp/ _samples_b37_SHAP	technical/workir EIT.20120103.h	g/20120103 aplotypes.vcf	3_omni_s f.gz	hapeit_haplotypes			

1b. Find the most recent Omni VCF file on build 37 from the 31st January 2012

using 31*omni*vcf as a search term should give you two results, one which is b36 and one which is b37

ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/working/20120131_omni_ genotypes_and_intensities/Omni25_genotypes_2141_samples.b37.vcf.gz 2. Use the Website search box found in the top right hand corner of all pages to find the FAQ question about getting subsections of VCF files.



3. Use the Data slicer to get this section of the Omni VCF file 6:31830969-31846823:

http://browser.1000genomes.org/Homo_sapiens/UserData/SelectSlice

Configure Page Custom Data		
Data Management	A Tin	
 Upload Data Attach DAS Attach Remote File 	When slicing a VCF or BAM file, both the data file and its index file should The VCF file should have a ".vcf.gz" extension, and the index file should h	be present on the web server and named correctly. ave a ".vcf.gz.tbi" extension, E.g: MyData.vcf.gz, MyData.vcf.gz.tbi
 Manage Data Features on Karyotype 	The BAM file should have a ".bam" extension, and the index file should ha	we a ".bam.bai" extension, E.g: MyData.bam, MyData.bam.bai
Data Converters Assembly Converter	Click here for more extensive documentation.	
 ID History Converter Variant Effect Predictor 	VCF / BAM File URL:	ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20101123/interim_phase1
Variation Pattern Finder		e.g. ftp://tp.1000genomes.ebi.ac.uk/vol1/ttp/release/20101123 /interim_phase1_release /ALL.chr1.phase1.projectConsensus.genotypes.vcf.gz
	Region:	6:31830969-31846823
	Use VCF filters (this doesn't apply to BAM files):	(e.g. 1:1-50000)
		By individual(s)
		By nonulation(s) *
		(by populations)
		File in the box below)
	Sample-Population Mapping File URL:	ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20101123/interim_phase1
		e.g. ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20101123 /interim_phase1_release/interim_phase1.20101123.ALL.panel
		< Back Next >
Configure Page Custom Data		
Data Management		
 Upload Data Attach DAS 	Thank you - your VCF file [6.31830969-31846823.ALL.chr6.phase1.proje- Right click on the file name and choose "Save link you" from the menu	ctConsensus.genotypes.vcf.gz] [Size: 83436] has been generated.
Attach Remote File Manage Data Features on Karyotype	Preview	
- Assembly Converter	##fileformat=VCFv4.0	
- Variant Effect Predictor	<pre>##source=BCM:SNPTools:hapfuse ##reference=1000Genomes_NCBI37</pre>	
 Data Slicer Variation Pattern Finder 	##FORMAT= <id=gt,number=1,type=string,description="gen< td=""><td>otype"></td></id=gt,number=1,type=string,description="gen<>	otype">
	##FORMAT= <id=ap,number=2,type=float,description="alle #CHROM POS ID REF ALT QUAL FILTE</id=ap,number=2,type=float,description="alle 	lic Probability, P(Allele=1 R INFO FORMAT HG00096 1
	6 31831159 rs3869144 C T	100 PASS . (
	6 31831167 . T C 100	PASS . GT:AP
		1775

Browsing Data

4. Find the variant rs45562238 using http://browser.1000genomes.org.

2 1000 Gene	omes Browser			A 000 (000 - 000 - 000	*
Maps 🔢 dat	wser.1000genomes.org/index.html es 🗌 fix lj 📑 docs 🗌 plus M	gm 🛂 g 🖪 fb 🥎 ds 🚨 lj 😒 NCBI	🚯 p 👯 E	SRA KC M IKG 💥 JIRA 🗍 Wish 🚨 an	Later
1000 C	GENOMES log of Human Genetic Varia	tion	an Ale Ale		12238
	Search 1000 G	enomes	earch	BOX s Browser	enomes data
	e.g. gene BRCA2 or Chromosome	6:133098746-133108745	In order to community 20101123. into the UC non rs SN	acilitate immediate analysis of the 1000 Genomes Project this browser (based on Ensemb) integrates the SNP calls This data has be submitted to dbSNP, and once rsid's have SC and Ensemb) browsers according to their respective re 1d's on this site are temporary and will NOT be mainta	data by the whole scientific from an interim release + been allocated, will be absorbed lease cycles. Until that point any ained.
Start Brov	vsing 1000 Genomes data				
GF	<mark>owse Human</mark> → RCh37		Links		
Pro	otein variations → ew the consequences of sequence variation	on at the level of each protein in the genome.		$\underline{1000 \ Genomes} \rightarrow$ More information about the 1000 Genomes Project on the	e 1000 genomes main site.
Ind Sh	lividual genotypes → ow different individual's genotype, for a va	iriant.	- And	Pilot browser → This browser is based on Ensembl release 60 and repre part of <u>Amap of human genome variation from populatio</u> .1061.1073.	sents the variant set analysed as in-scale sequencing, Nature 467
Browser u	pdate September 2011 erim Main project data from 2010	01123 for 1094 individuals and	1 Alexandre	Tutorial → The 1000 Genomes Browser Tutorial.	
ensembl rele Please see <u>wy</u> here and inst • <u>View sa</u>	ase 63. The data can be found on ww.1000genomes.org for more in ructions for downloading the con umple data	the ftp site. formation about the data presented mplete data set.	The 1000 (<u>www.1000</u> The 1000 (<u>Ensembl</u> is	enomes Project is an international collaborative project de <u>normes.com.</u> enomes Browser is based on Ensembl web code. a joint project of EMBL-EBI	scribed at
1000 Genomes relea	ise 9 - September 2011 © <u>EBI</u>				About 1000 Genomes I Contact Us I Help
A THOUSAND GENOMES				EN	MBL-EBI 🌒
	A Deep Catalog of Hur	nan Genetic Variation		AC 8	
	Search 1000 Genomes			Results Su	
	Configure this page	Gene or Gene Product			
	🕞 Manage your data	0 entrie(s) matched your search stri	ngs.		
	Export data	Genetic Marker			
	re Get VCF data	0 entrie(s) matched your search stri	ngs.		
	😭 Bookmark this page	Array Probe Set			
		0 entrie(s) matched your search stri	ngs.		
		SNP			
		1 entrie(s) matched your search stri	ngs.		
		Interpro Domain		SINF RES	ult
		0 entrie(s) matched your search stri	ngs.		
		Gene Family	-		
		0 entrie(s) matched your search stri	ngs.		
		Sequence Aligned to Genome,	eg. EST or I	rotein	
		0 entrie(s) matched your search stri	ngs.		
		Genomic Region, eg. Clone or	Contig		
		0 entrie(s) matched your search stri	ngs.		
		1000 Genomes release 9 - September 20	011 © <u>EBI</u>		
A REPUBLICA					
A THOUSAND GENOMES	I.			EN	NBL-EBI

5. In what 1000 Genomes Super Population is this variant detected?

American and European

6. What are its global allele frequencies in the 1000 Genomes Data set?

0.02 is the global allele frequency, this is also the American Allele Frequency but it rises to 0.04 in the Europeans. The absence of Asians or Africans in this chart means that the variant was not found in any of our Asian of African individuals.

man (GRCh37) 🔻 🚺 Locatio	on: 6:133,013,044-133	3,014,044 Variation	rs45562238					IDOIS I HE
ation displays				N N	ariation: rs45562	:238		
anking sequence ene/Transcript (1)								
opulation genetics (11)	Variation class	SNP (<u>rs45562238</u> sou	rce dbSNP 132 - V	/ariants (including !	SNPs and indels) impo	orted from dbSNP [http://w	www.ncbi.nlm.nih.gov/proje	cts/SNP/])
dividual genotypes (2770) enomic context	Synonyms	OMNI SNP6-1330552 Uniprot VAR_023973	37					
henotype Data hylogenetic Context	Present in	1000 genomes - Low interim phase 1 - 1000	coverage (1000 ge Genomes),ENSE	momes - Low cove MBL:Watson	rage - CEU),ALL - inte	rim phase 1 - 1000 Geno	omes (AMR - interim phase	1 - 1000 Genomes, EUR -
xtema Data	Alleles	T/C (Ambiguity code:	Y)					
	Ancestral allele	т						
Manage your data	Location	This feature maps to 6	:133013544 (forw	ard strand) I <u>View i</u>	n location tab			
manago your Gata	Validation status	Proven by cluster, fre	quency, 1000Ger	iome				
Export data	HGVS names 🗄	This feature has 3 HG	VS names - click t	he plus to show				
Get VCF data				Po	pulation genetics	help		
Bookmark this page	1000 genomes	alleles frequencies						
Download view as CSV	ALL	AMR	EUR					
	1:98% C:2%	5 T: 98% C: 2%	U : 1: 90 C: 4	5% %		Pie	Charts	
	1000 genomes							
	Show/hide colu	umns					F	ilter
	Population	Alleles C	Alleles	Genotypes CIT	Genotypes TIT	Allele count	Genotype count	Genotype detail
		0.046	0.984	0.031	0.969	34 (C) / 2154 (T)	34 (CIT) / 1060 (TIT)	Show
	1000GENOMES:A	ull 0.016						
	1000GENOMES:A 1000GENOMES:A	MR 0.017	0.983	0.033	0.967		6 (CIT) / 175 (TIT)	Show



EMBL-EBI 🌒

7. In which gene is the variant found?

ENSG00000112299, Vanin 1

1000 Genor A Deep Catalog of Hun	NES nan Genetic V	ariation				er. N			8: 05 8: 05		1-	2 3	× 18 -
Human (GRCh37) V Location	1: 6:133,013,044-133	Variation: 1	rs45562238										roois i neip
Variation displays					Variatio	n: rs45562	238						
Gene/Transcript (1)													
- Population genetics (11)	Variation class	SNP (<u>rs45562238</u> sourc	e dbSNP 132 -	Variants (including	SNPs and indels) imports	ed from dbSN	P [http://www.n	cbi.nlm.nih.gov	projects/SNP/]				
 Individual genotypes (2770) Genomic context 	Synonyms	Uniprot VAR 023973	/										
 Phenotype Data Phylogenetic Context 	Present in	1000 genomes - Low co Genomes),ENSEMBL:V	werage (1000 g Vatson	enomes - Low cove	rage - CEU},ALL - interir	n phase 1 - 1	000 Genomes ((AMR - interim p	has			1	000
- External Data	Alleles	T/C (Ambiguity code: Y))						5	TT/PO	iyPne	n	
Configure this page	Ancestral allele	т											
🕞 Manage vour data	Location	This feature maps to 6:1	133013544 (forw	vard strand) I View	in location tab						<u>۱</u>		
	Validation status	Proven by cluster, freq	uency, 1000Ge	nome							\		
Export data	HGVS names ⊕	This feature has 3 HGV	S names - click	the plus to show							_\		
😭 Get VCF data					Gene/Tr	anscript 👔	e!p						
👷 Bookmark this page											- N		
Download view as CSV	Show/hide coll.										Filter		- 1
	Gene	Transcript (strand)	Allele (transcript A allele)	Туре	HGVS names		Position in transcript	Position in CDS	Position in protein	Amino acid	Codons	SIFT	PolyPhen
	ENSG0000011225	9 ENST0000367928 (+)	C (G)	Non-synonymous coding	ENST00000367928.4 ENSP00000356905.4	c.1006A>G p.Thr336Ala	1020	1006	336	T/A	ACT/GCT	tolerated	benign
	1000 Genomes relea	ise 9 - September 2011 © E	BI								About 1000	Genomes I C	iontact Us I <u>Help</u>





Using the 1000 Genomes Tools

8. Use the browser to find the SLC44A4 gene.



Putting the Gene name in the search box that is found in the top right hand corner of every page should lead you to the results page. You should follow the Gene name link to the Gene page.

9. Use the get VCF button in the left hand menu on the gene page to get a slice of a vcf file for this Gene.



When following the get vcf data button the form automatically fills out the input vcf file and position of the gene in the region box. If you wish to sub select a particular population or individual you would need to tick the appropriate box

Than Right	ik you - your VCF t click on the file r	file [<u>6.3</u> name ar	31830969-31 nd choose "S	<mark>846823.</mark> A ave link a	LL.chr6.pha as" from the	se1.project0 e menu	Consensus	s.genotypes.	vcf.gz] [Si:	ze: 834	36] has been ge	enerated.
Prev	view											
##f:	ileformat=VCF	v4.0										
##se ##re	ource=BCM:SNF eference=1000	Genom	:napruse es-NCBI37									
##F0 ##F0	ORMAT= <id=gt, ORMAT=<id=ap,< th=""><th>Numbe Numbe</th><th>r=1,Type= r=2,Type=</th><th>String, Float,D</th><th>Descripti escriptio</th><th>on="Genot n="Alleli</th><th>ype"> .c Proba</th><th>bility, P</th><th>(Allele</th><th>=1</th><th></th><th></th></id=ap,<></id=gt, 	Numbe Numbe	r=1,Type= r=2,Type=	String, Float,D	Descripti escriptio	on="Genot n="Alleli	ype"> .c Proba	bility, P	(Allele	=1		
#CHI	ROM POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	HG0009	6 1		
	21021150						100	DA <i>GG</i>				
6	31831159		rs3869	144 T	c	100	PASS	PASS	GT:AP	1		
										11.		

The Final page gives you a look at the top few lines of the file and a link to download the complete file

10. Unzip this VCF file using a tool link winzip or Archive Utility.

This should produce a file called 6.31830969-31846823.ALL.chr6.phase1.projectConsensus.genotypes.vcf

11a. Upload this VCF file to the Variant Effect Predictor. http://browser.1000genomes.org/Homo_sapiens/UserData/UploadVariations

Assembly Converter ID History Converter Variant Effect Predictor	Input file			J.
 Data Slicer Variation Pattern Finder 	Species:	Human (Homo sapiens): GRCh37 🗧 🗧		%
	Name for this upload (optional):	SLC44A4	h	40 % %
	Paste file:			1
			l	it can
			l	
	Upload file:	/Users/laura/Downloads/6. Browse	l	
	or provide file URL:		l	
	Input file format:	VCF \$	l	
	Options		l	
	Get regulatory region consequences:		l	
	Type of consequences to display:	Ensembl terms \$	l	
	Check for existing co-located variants:	Yes ‡	l	
	Return results for variants in coding regions only:		l	es I C
	Show HGNC identifier for genes where available:		l	
	Show Ensembl protein identifiers where available:		l	
	Show HGVS identifiers for variants where available:	No ‡	ľ	
	Non-synonymous SNP predictions (human only)			
	SIFT predictions:	Prediction only +		
	PolyPhen predictions:	Prediction only \$		
	Condel consensus (SIFT/PolvPhen) predictions:	No *	//,	

The input form asks you to browse to the location of the vcf file. You need to select vcf as the input format and to see the SIFT and PolyPhen predictions you need to select the appropriate dropdown menus.

6_31833249_A/G	<u>6:31833249</u>	G	ENSG00000204385	ENST00000487680	Transcript	UPSTREAM	-	-	-	-
6_31833249_A/G	<u>6:31833249</u>	G	ENSG00000204385	ENST00000414427	Transcript	DOWNSTREAM	-	-	-	-
6_31833249_A/G	6:31833249	G	ENSG00000204385	ENST00000479777	Transcript	DOWNSTREAM	-	-	-	-
6_31833249_A/G	6:31833249	G	ENSG00000204385	ENST00000475563	Transcript	DOWNSTREAM	-	-	-	-
6_31833357_C/T	<u>6:31833357</u>	Т	-	ENSR00000487922	RegulatoryFeature	REGULATORY_REGION	-	-	-	-
6_31833357_C/T	<u>6:31833357</u>	Т	ENSG00000204386	ENST00000495807	Transcript	UPSTREAM	-	-	-	-
6_31833357_C/T	<u>6:31833357</u>	Т	ENSG00000204386	ENST00000480384	Transcript	UPSTREAM	-	-	-	-
6_31833357_C/T	6:31833357	Т	ENSG00000204386	ENST00000491768	Transcript	UPSTREAM	-	-	-	-
6_31833357_C/T	<u>6:31833357</u>	Т	ENSG00000204386	ENST0000375631	Transcript	UPSTREAM	-	-	-	-
6_31833357_C/T	<u>6:31833357</u>	Т	ENSG00000204386	ENST00000479533	Transcript	UPSTREAM	-	-	-	-
6_31833357_C/T	<u>6:31833357</u>	Т	ENSG00000204385	ENST00000229729	Transcript	NON_SYNONYMOUS_CODING	G 1625	160	4 535	R/H
6 31833357 C/T	6:31833357	т	ENSG00000204385	ENST0000375562	Transcript	NON SYNONYMOUS CODING	3 1544	147	8 493	B/H
0_01000007_0/1	0.01000001	•	21100000000000000		rranoonpr				, 400	1011
		_			_					
6_31833357_C/T	<u>6:31833357</u>	Т	ENSG00000204385	ENST00000544672	Transcript	NON_SYNONYMOUS_CODING	G 1673	137	3 459	R/H
6_31833357_C/T	6:31833357	Т	ENSG00000204385	ENST00000487680	Transcript	UPSTREAM	-	-	-	-
6_31833357_C/T	6:31833357	Т	ENSG00000204385	ENST0000414427	Transcript	DOWNSTREAM	-	-	-	-
6_31833357_C/T	<u>6:31833357</u>	Т	ENSG00000204385	ENST00000479777	Transcript	DOWNSTREAM	-	-	-	-
6_31833357_C/T	<u>6:31833357</u>	Т	ENSG00000204385	ENST00000475563	Transcript	DOWNSTREAM	-	-	-	-
6_31833612_C/G	<u>6:31833612</u>	G	-	ENSR00000487922	RegulatoryFeature	REGULATORY_REGION	-	-	-	-
6_31833612_C/G	6:31833612	G	ENSG00000204386	ENST00000495807	Transcript	UPSTREAM	-	-	-	-
6_31833612_C/G	6:31833612	G	ENSG00000204386	ENST00000480384	Transcript	UPSTREAM	-	-	-	-

The output from the Variation Effect Precictor gives the provided identifier for the variant (or uses the position and allele string to create one), the position of the variant, the alternative allele then information about the feature it overlaps with and that effect that causes.

-	-	-	<u>1KG 6 31833357</u> -
-	-	-	1KG 6 31833357 -
535	R/H	cGc/cAc	<u>1KG 6 31833357</u> SIFT=deleterious; PolyPhen=probably_damaging; Condel=deleterious
493	R/H	cGc/cAc	<u>1KG 6 31833357</u> SIFT=deleterious; PolyPhen=possibly_damaging; Condel=deleterious
459	R/H	cGc/cAc	<u>1KG 6 31833357</u> SIFT=deleterious; PolyPhen=probably_damaging; Condel=deleterious
-	-	-	1KG 6 31833357 -
-	-	-	1KG 6 31833357 -
-	-	-	<u>1KG 6 31833357</u> -

If you scroll along the page you will see additional information that you requested on the form like any variants in the Ensembl database yours overlaps with and what the SIFT and PolyPhen results are.

11b. Do any of the variants have negative SIFT or PolyPhen predictions?

Yes, There are several variants which have negative SIFT or PolyPhen predictions including 6_31833357_C/T which overlaps with 3 different transcripts all with deleterious non synonymous codon changes, ENST00000229729, ENST00000375562, ENST00000544672

12. Using the example URLs on the Variation Pattern Finder tool menu look at the patterns of inheritance for this region: 6:31830700-31840700 http://browser.1000genomes.org/Homo_sapiens/UserData/VariationsMapVCF

Custom Data	
Data Management Upload Data Attach DAts Attach DAts Batach Remote File Hanage Data Features on Karyotype E Data Converter - Datatory Converter - Variant Effect Predictor - Data Sicer	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 there 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 file you provided. It should be a URL for the file location. Please refer to http://vcftools.sourceforge.net/specs.html for VCF file you provided. It should be a URL for the file location. Please refer to http://vcftools.sourceforge.net/specs.html for VCF file you provided. It should be a URL for the file location. Please refer to http://vcftools.sourceforge.net/specs.html for VCF file you provided. It should be a URL for the file location. Please refer to http://vcftools.sourceforge.net/specs.html for VCF file you provided. It should be a URL for the file location. Please refer to http://vcftools.sourceforge.net/specs.html for VCF file you provided. It should be a URL for the file location. Please refer to http://vcftools.sourceforge.net/specs.html for VCF file you provided. It should be a URL for the file location. Please refer to http://vcftools.sourceforge.net/specs.html for yariation calls and genotypes released by the 1000 Genomes
- Variation Pattern Finder	Upload files
	VCF File URL: /interim_phasel_release /ALL.chr6.phasel.projectConsensus.genotypes.vcf.gz
	e.g. ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/released of U1123/interim_phase1_release/ALL.chr6.phase1.projectConsensus.genotypes.vcf.gz
	Sample-Population Mapping File URL: <pre>ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20101123 /interim_phasel_release/interim_phasel.20101123.ALL.panel</pre>
	e.g. ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20101123/interim_phase1_release/interim_phase1.20101123.ALL.panel
	Region: 6:31830700-31840700 e.g. 6:46620015-46620998 Next >

As this gene is on chr 6 the default URLs in the form when you first click on it should be fine. You need to add your region to the region box and then after clicking next you will see:

Data Management	Madati a Battan Etada						
Attack DAO	variation Pattern Finder						
 Attach DAS Attach Remote File Manage Data 	Export data: CSV Excel						
Features on Karyotype Data Converters Assembly Converter	Go to collapsed view						
 ID History Converter Variant Effect Predictor 	Population ASW	CEU	Freq		rs116706632:G/A	rs117127493:G/C	rs644827:T/C
 Data Slicer 					6:31836976	6:31837009	6:31838441
Variation Pattern Finder				_CODING:R/C	ENST00000229729 NON_SYNONYMOUS_CODING:P/S	ENST00000229729 NON_SYNONYMOUS_CODING:Q/E	ENST00000229729 NON_SYNONYMOU
				_CODING:R/C	ENST00000375562 NON_SYNONYMOUS_CODING:P/S	ENST00000375562 NON_SYNONYMOUS_CODING:Q/E	ENST00000375562 NON_SYNONYMOU
				CODING:R/C	ENST00000544672 NON_SYNONYMOUS_CODING:P/S	ENST00000544672 NON_SYNONYMOUS_CODING:Q/E	ENST00000544672 NON_SYNONYMOU
				_CODING:R/C	ENST00000414427 NON_SYNONYMOUS_CODING:P/S	ENST00000414427 NON_SYNONYMOUS_CODING:Q/E	
	•						
	NA20289, NA20296 and 13 other(s)	NA069	0.293		GIG	GIG	CIC
	NA20127, NA19703 and 9 other(s)	NA128	0.203		GIG	GIG	CIT
	NA20314, NA20317 and 6 other(s)	NA120	0.195		GIG	GIG	TIC
	NA19920, NA19700 and 2 other(s)		0.032		GIG	GIG	CIC
	NA19819, NA20281 and 2 other(s)		0.026		GIG	GIG	CIC
	NA20291, NA20356 and 3 other(s)		0.016		GIG	GIG	TIC
	NA19908	NA122	0.013		GIG	GIG	CIT
			0.008		GIG	CIG	CIC
			0.005		GIG	GIC	TIC
		NA119	0.005		GIG	GIC	CIC
	NA19916		0.004		GIG	GIG	CIC
	NA19711 NA20340		0.003		GIG	GIG	CIC
			0.003		GIG	GIG	CIT
		NA119	0.003		GIA	GIG	CIC
			0.003		GIG	CIG	CIT

The grey headline row has population names and varianat names and alleles. The following rows contain the functional consequences of these variants. This tool only considers variants with functional consequences. The rows then contain a list of individuals who are part of that population, the global frequency this pattern occurs in and the actual pattern of genotypes in those individuals. 13a. For the same region use the VCF to PED tool to produce a ped and info file for the CEU population.

lick <u>here</u> for more extensive documentation.	
Jpload files	
/CF File URL:	ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20101123 /interim_phasel_release /ALL.chr6.phasel.projectConsensus.genotypes.vcf.gz
	<u>Clear box</u>
e.g. ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20101123/interi	m_phase1_release/ALL.chr6.phase1.projectConsensus.genotypes.vcf.gz
Sample-Population Mapping File URL:	ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20101123 /interim_phasel_release/interim_phase1.20101123.ALL.panel
	<u>Clear box</u>
e.g. ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20101123/interi	m_phase1_release/interim_phase1.20101123.ALL.panel
Region:	6-31830700-31840700
	0.51650700-51640700
.g. 6:46620015-46620998	Next >
.g. 6:46620015-46620998 VCF filter by population(s)	Next >
v.g. 6:46620015-46620998 VCF filter by population(s) Select one or more populations from the scrollable list:	ASW CEU CHB CHS CLM FIN GBR IBS JPT LWK

Again as we are considering a gene on chromosome 6 the default URLs in the box should work. Once you have put the region in the region box and clicked next you should see a list of populations including CEU. If you select CEU and click next you will then be presented with two links to files you can right click on and download.

13b. Look at these files in haploview.

000	Welcome to HaploView	
Linkage Format Ha	ps Format HapMap Format	HapMap PHASE
Data File	wnloads/6_31830700-31840700.ped	Browse
Locus Information File	wploads /6_31830700_31840700 info	Browse
		biowse
	Chromosome Do association test	
• F	amily trio data 🔿 Case/Control data	
	Standard TDT O ParenTDT	
Test list file (optional)		Browse
Ignore pairwise	comparisons of markers > 500	kb apart.
Exclude ind	ividuals with > 50 % missing genoty	/pes.
	OK Cancel	Proxy Settings
ile Display Analysis Help	OK Cancel Haploview 4.2 6_31830700-31840700.ped	Proxy Settings
File Display Analysis Help BIOCK 2 (5 KD)	OK Cancel Haploview 4.2 6_31830700-31840700.ped LD Plot Haplotypes Check Markers Tagger	Proxy Settings
File Display Analysis Help 5 16 17 18 19 20 21 22 2	OK Cancel Haploview 4.2 6_31830700-31840700.ped D Plot Haplotypes Check Markers Tagger 24 25 26 27 28 29 30 31 32 33	Proxy Settings
File Display Analysis Help 5 16 17 18 19 20 21 22 22 74 16 74	OK Cancel Haploview 4.2 6_31830700-31840700.ped DPlot Haplovypes Check Markers Tagger 24 25 26 27 28 29 30 31 32 33	Proxy Settings Ka 34 35 36 37 38 39
File Display Analysis Help 5 16 17 18 19 20 21 22 2 74 16 74 12 5 72	OK Cancel Haploview 4.2 6_31830700-31840700.ped ID Plot Haplotypes Check Markers Tagger 24 25 26 27 28 29 30 31 32 33	Proxy Settings Ka 34 35 36 37 38 39 90
File Display Analysis Help 5 16 17 18 19 20 21 22 2 74 16 74 74 77 72 72 72	OK Cancel Haploview 4.2 6_31830700-31840700.ped DPor Haplotypes Check Markers Tagger 24 25 26 27 28 29 30 31 32 33	Proxy Settings Ka 34 35 36 37 38 39 90 90 90 90 6
File Display Analysis Help 5 16 17 900CK 2 13 KD) 74 16 77 74 16 77 72 72 72 13 63 77 72 72 13	OK Cancel Haploview 4.2 6_31830700-31840700.ped IDPot 10 Plot Haplovypes Check Markers Tagger 2 2 2 30 31 32 33 4 4 25 26 27 28 29 30 31 32 33 4	Proxy Settings Ka 34 35 36 37 38 39 50 50 50 50 50 50 50 50 50 50
File Display Analysis Help 5 16 17 16 19 20 21 22 2 74 16 7 12 72 72 7 63 63 72 72 72	OK Cancel Haploview 4.2 6_31830700-31840700.ped 1 Plot Haploview 4.2 6_2 31830700-31840700.ped 24 25 26 27 28 29 30 31 32 33 24 25 26 27 28 29 30 31 32 33	Proxy Settings Ka 34 35 36 37 38 39 90 90 90 90 90 90 90 90 90 9
File Display Analysis Help 5 16 17 18 19 20 21 22 2 74 12 72 72 72 73 13 63 63 72 72 72 72 72 72 72 72 72 72 72 72 72	OK Cancel Haploview 4.2 6.31830700-31840700.ped 1 24 25 26 27 28 29 30 31 32 33 24 25 26 27 28 29 30 31 32 33	Proxy Settings K 34 35 36 37 38 39 90 90 90 90 90 90 6 80 82 85 89 82 85
File Display Analysis Help 5 16 17 18 19 20 21 22 2 74 16 74 12 5 72 72 72 63 63 77 72 63 63 77 72 63 63 77 72 63 63 77 72	OK Cancel Haploview 4.2 6_31830700-31840700.ped 1 24 25 26 27 28 29 30 31 32 33 24 25 26 27 28 29 30 31 32 33 4	Proxy Settings
File Display Analysis Help 5 16 17 900K 2 (3 KD) 20 21 22 2 74 12 72 72 72 13 63 63 72 72 72 72 72 72 72 72 72 72 72 72 72	OK Cancel Haploview 4.2 6_31830700-31840700.ped ID Plot 10 Plot Haplotypes Check Markers Tagger 24 25 26 27 28 29 30 31 32 33 4 25 26 27 28 29 30 31 32 33 4	Proxy Settings 34 35 36 37 38 39 90 90 90 90 6 90 6 82 83 82 88 85 82 88 85 82 88 85 85 85 85 85 85 85 85 85

Loading the data (ped) and locus information file (info) into haploview gives you the ability to look at the ld plot for the region.





The Haplotypes button views you a view of the haplotype blocks which exist in that region. In this case there are 2 haplotype blocks.