



The 1000 Genomes Project How to Access the Data?

7/11/2012

Laura Clarke



How Much Data?



How Much Data?

- There are 333,332 files on the ftp site
- There are 367T of data on the ftp site
- There are 26 populations
- There are 2,208 samples
- There are 51,913 gigabases of low coverage sequence
- 18877x coverage in low coverage
- There are 23,860 gigabases of exome sequence



Access

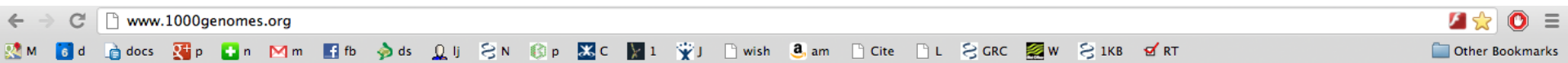


Access

- Website
 - <http://www.1000genomes.org>
- Two mirrored ftp sites
 - <ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp>
 - <ftp://ftp-trace.ncbi.nih.gov/1000genomes/ftp>
- NCBI site is direct mirror of EBI site
- Can be up to 24 hours out of date
- Both also accessible using aspera
 - <http://asperasoft.com/>
- EBI site has http mirror
 - <http://ftp.1000genomes.ebi.ac.uk/vol1/ftp>
- Amazon Cloud public data set

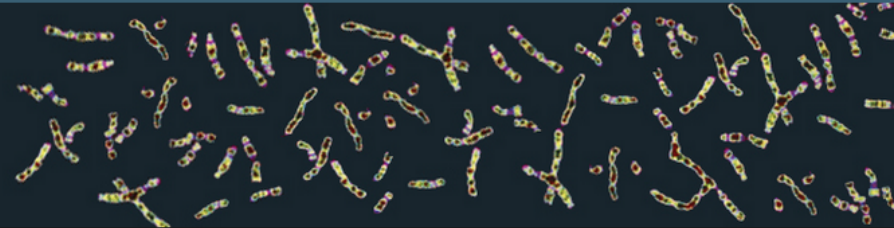


Website <http://www.1000genomes.org>



1000 Genomes

A Deep Catalog of Human Genetic Variation



[Home](#) [About](#) [Data](#) [Analysis](#) [Participants](#) [Contact](#) [Browser](#) [Wiki](#) [FTP search](#)

[View](#) [Edit](#) [Outline](#) [Revisions](#) [Grant](#)

LATEST ANNOUNCEMENTS

WEDNESDAY OCTOBER 31, 2012

An integrated map of genetic variation from 1092 human genomes

The Phase 1 publication, *An Integrated map of genetic variation from 1092 human genomes* is now available from Nature and can be downloaded directly from the ftp site. The paper is distributed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported licence. Please share our paper appropriately.

All the data files associated with this paper can be found in our [phase1 analysis results directory](#).

Recent project announcements

TUESDAY OCTOBER 23, 2012

#ASHG2012 1000 Genomes Tutorial, Wednesday 7th November 7-9:30pm San Francisco Marriott Marquis

The 1000 Genomes Project is holding a tutorial during ASHG 2012 on Wednesday 7th November 7:00 to 9:30pm at the San Francisco Marriot Marquis.

The 1000 Genomes Project has released the sequence data and an integrated set of variants, genotypes, and haplotypes for the 1092 samples in the phase 1 set, and the sequence data for the phase 2 set. This tutorial describes the data sets, how to access them, and how to use them.

LAURA@EBI.AC.UK

- [My account](#)
- ▶ [Create content](#)
- [List content](#)
- [List users](#)
- [Manage files](#)
- [Log out](#)
- [Frequently Asked Questions](#)

LINKS



[All Project Announcements](#)



[Sample and Project Information](#)



[Media Archive](#)



EMBL-EBI



















ftp://ftp.1000genomes.ebi.ac.uk

ftp://ftp-trace.ncbi.nih.gov/1000genomes/ftp

Index of ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/

 Up to higher level directory

Name	Size	Last Modified
 CHANGELOG		
 README.alignment_data		
 README.ftp_structure		
 README.pilot_data		
 README.populations		
 README.sequence_data		
 alignment_indices		
 changelog_details		
 current.tree		
 data		
 phase1		
 pilot_data		
 release		
 sequence.index		
 sequence_indices		
 technical		

Name	Size	Last Modified
Site documentation	12 KB	26/01/2011 26/01/201112 :00:00
Sequence and Alignment by Sample ID	9 KB	04/04/2011 4/04/201112 :00:00
Phase 1 Data Sets	3 KB	14/08/2009 14/08/200912 :00:00
Official Releases	2 KB	18/02/2010 18/02/201012 :00:00
Pre Release material and Reference Data Sets	27185 KB	20/12/2011 20/12/201112 :26:00



Phase I complete; paper just published

ARTICLE

doi:10.1038/nature11632

An integrated map of genetic variation from 1,092 human genomes

The 1000 Genomes Project Consortium*

Nature. 2012. 491:56-65














FTP Site: Phase1 Data

Index of ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/phase1/

 [Up to higher level directory](#)

Name

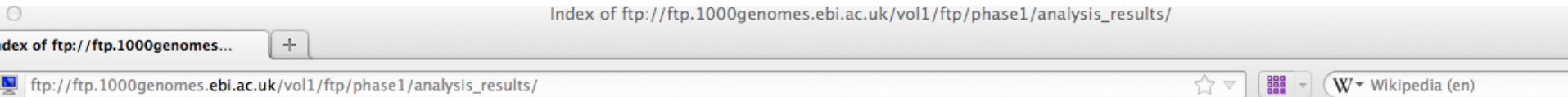
 README.phase1_alignment_data	11 KB	08/02/2012	8/02/2012 12:00:00
 analysis_results		20/06/2012	20/06/2012 9:45:00
 data			
 phase1.alignment.index	1140 KB	27/03/2012	27/03/2012 12:00:00
 phase1.alignment.index.bas.gz	642 KB	27/03/2012	27/03/2012 12:00:00
 phase1.exome.alignment.index	389 KB	14/12/2011	14/12/2011 12:00:00
 phase1.exome.alignment.index.HsMetrics.gz	141 KB	14/12/2011	14/12/2011 12:00:00
 phase1.exome.alignment.index.HsMetrics.stats	1 KB	14/12/2011	14/12/2011 12:00:00
 phase1.exome.alignment.index.bas.gz	414 KB	14/12/2011	14/12/2011 12:00:00
 phase1.exome.alignment.index_stats.csv	1 KB	14/12/2011	14/12/2011 12:00:00
 technical		14/12/2011	14/12/2011 12:00:00

Phase 1 Analysis Results

Phase 1 Alignment Files



FTP Site: Phase1 Data



Index of ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/phase1/analysis_results/

[Up to higher level directory](#)

Name	Size	Last Modified
README_20120614_phase1_analysis_sets	5 KB	04/07/2012 4/07/2012 3 :01:00
ancestry_deconvolution		13/06/2012 13/06/2012 2 :43:00
consensus_call_sets		20/06/2012 20/06/2012 1 :42:00
experimental_validation		21/06/2012 21/06/2012 10 :59:00
functional_annotation		21/06/2012 21/06/2012 10 :53:00
input_call_sets		21/06/2012 21/06/2012 2 :13:00
integrated_call_sets		9/10/2012 9/10/2012 8 :34:00
paper		2/11/2012 2/11/2012 3 :55:00
supporting		26/09/2012 26/09/2012 1 :03:00

Functional Annotation

Input Call Sets

Integrated Call Sets

Paper



Data formats and key tools

BIOINFORMATICS APPLICATIONS NOTE

Vol. 25 no. 16 2009, pages 2078–2079
doi:10.1093/bioinformatics/btp352

Sequence analysis

The Sequence Alignment/Map format and SAMtools

Heng Li^{1,†}, Bob Handsaker^{2,†}, Alec Wysoker², Tim Fennell², Jue Ruan³, Nils Homer⁴, Gabor Marth⁵, Goncalo Abecasis⁶, Richard Durbin^{1,*} and 1000 Genome Project Data Processing Subgroup⁷

¹Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Cambridge, CB10 1SA, UK, ²Broad Institute of MIT and Harvard, Cambridge, MA 02141, USA, ³Beijing Institute of Genomics, Chinese Academy of Science, Beijing 100029, China, ⁴Department of Computer Science, University of California Los Angeles, Los Angeles, CA 90095, ⁵Department of Biology, Boston College, Chestnut Hill, MA 02467, ⁶Center for Statistical Genetics, Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA and ⁷<http://1000genomes.org>

Received on April 28, 2009; revised on May 28, 2009; accepted on May 30, 2009

Advance Access publication June 8, 2009

Associate Editor: Alfonso Valencia

BAM alignment files

BIOINFORMATICS APPLICATIONS NOTE

Vol. 27 no. 15 2011, pages 2156–2158
doi:10.1093/bioinformatics/btr330

Sequence analysis

Advance Access publication June 7, 2011

The variant call format and VCFtools

Petr Danecek^{1,†}, Adam Auton^{2,†}, Goncalo Abecasis³, Cornelis A. Albers¹, Eric Banks⁴, Mark A. DePristo⁴, Robert E. Handsaker⁴, Gerton Lunter², Gabor T. Marth⁵, Stephen T. Sherry⁶, Gilean McVean^{2,7}, Richard Durbin^{1,*} and 1000 Genomes Project Analysis Group[†]

¹Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Cambridge CB10 1SA, ²Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK, ³Center for Statistical Genetics, Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, ⁴Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA 02141, ⁵Department of Biology, Boston College, MA 02467, ⁶National Institutes of Health National Center for Biotechnology Information, MD 20894, USA and ⁷Department of Statistics, University of Oxford, Oxford OX1 3TG, UK

Associate Editor: John Quackenbush

VCF variant files

BIOINFORMATICS APPLICATIONS NOTE

Vol. 27 no. 5 2011, pages 718–719
doi:10.1093/bioinformatics/btq671

Sequence analysis

Advance Access publication January 5, 2011

All indexed for fast retrieval

Tabix: fast retrieval of sequence features from generic TAB-delimited files

Heng Li

Program in Medical Population Genetics, The Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA

Associate Editor: Dmitriy Frishman



EMBL-EBI



Amazon ECS



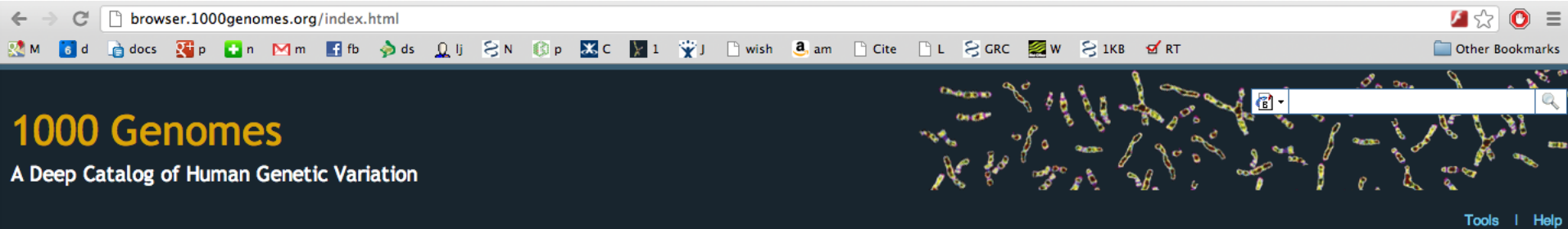
- FTP site mirror
- AMI image for our tutorial
- Preconfigured mount in CloudBioLinux
- <http://www.1000genomes.org/using-1000-genomes-data-amazon-web-service-cloud>



Browser



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



← → ↻ browser.1000genomes.org/index.html

M d docs p n m fb ds lj N p C 1 J wish am Cite L GRC W 1KB RT Other Bookmarks

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 May 2012

based on the [phase 1 integrated release for 1092 individuals and ensembl release 65](#). 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 and INDEL calls from the [phase 1 integrated release](#). 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](#) 



EMBL-EBI



Transcript Tab: Variations

Effect on Protein:

- SIFT
- PolyPhen

1000 Genomes
A Deep Catalog of Human Genetic Variation

Human (GRCh37) Location: 1:114,356,433-114,414,381 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 1:114,356,433-114,414,381 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](#)

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	-	-

1000 Genomes release 10 - October 2011 © EBI



Variation

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](#)

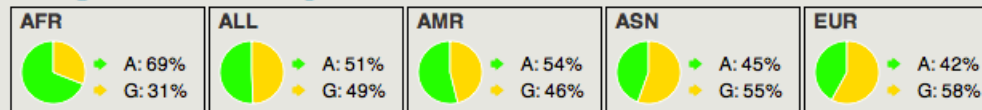
Validation status Proven by cluster, frequency, doublehit, 1000Genome HapMap variant

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	

Tools

- Data Slicer
- Variant Effect Predictor
- Variation Pattern Finder
- VCF to PED converter

1000 Genomes
A Deep Catalog of Human Genetic Variation

Search 1000 Genomes

Go

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

Start Browsing 1000 Genomes data

[Browse Human](#) →
GRCh37

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 and INDEL calls from the [phase 1 integrated release](#). 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

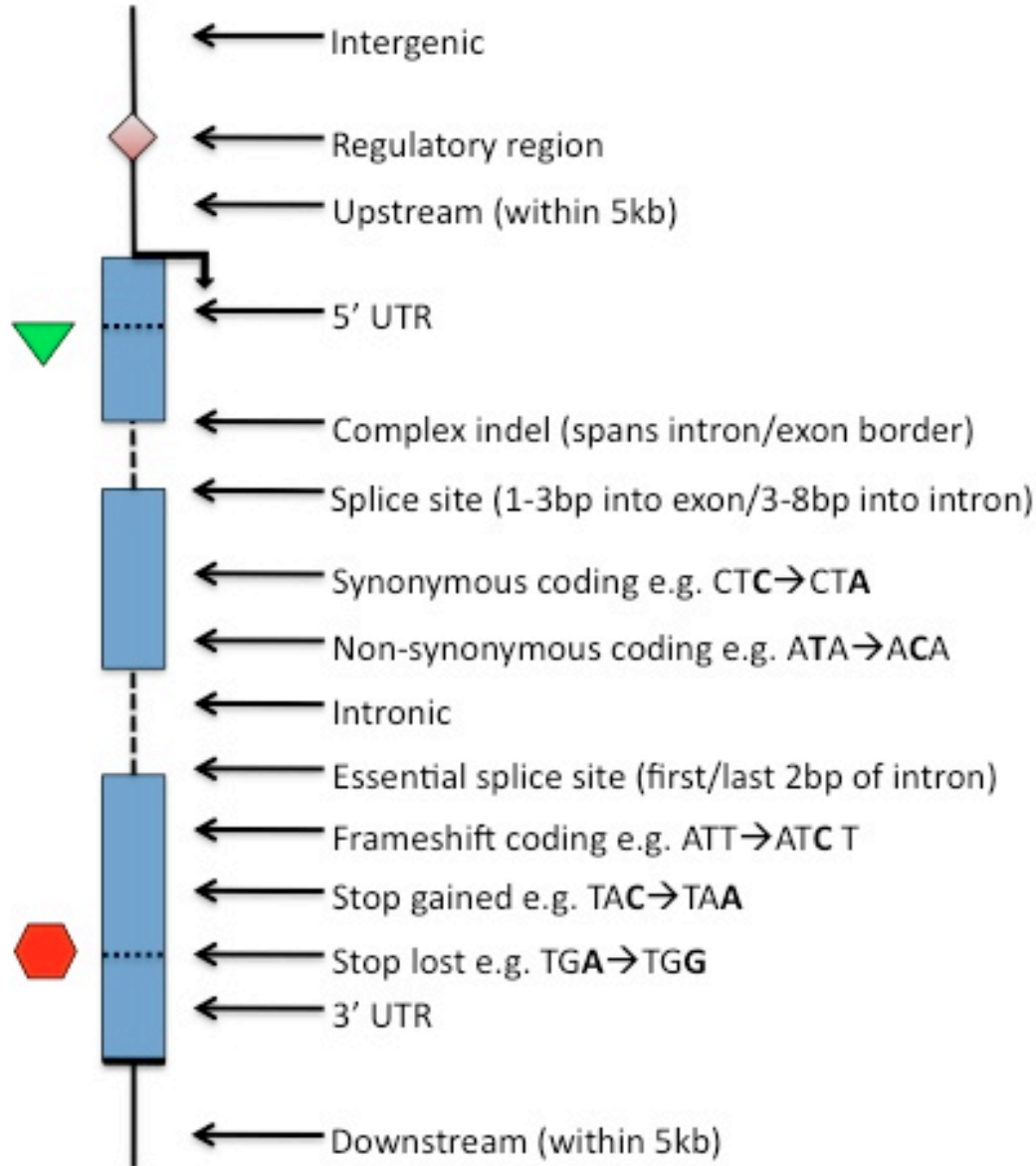


Variant Effect Predictor

- Predicts Functional Consequences of Variants
- Both Web Front end and API script
- Can provide
 - sift/polyphen/condel consequences
 - Refseq gene names
 - HGVS output
- Can run from a cache as well as Database
- Convert from one input format to another
- Script available for download from:
 - ftp://ftp.ensembl.org/pub/misc-scripts/Variant_effect_predictor/
 - http://browser.1000genomes.org/Homo_sapiens/UserData/UploadVariations



Variant Effect Predictor



Others: Within non-coding gene, Within mature miRNA, NMD transcript

Variation Effect Predictor Output

6_31833357_C/T	6:31833357	T	ENSG00000204386	ENST00000480384	Transcript	UPSTREAM	-	-	-	-	-	-	
6_31833357_C/T	6:31833357	T	ENSG00000204386	ENST00000491768	Transcript	UPSTREAM	-	-	-	-	-	-	
6_31833357_C/T	6:31833357	T	ENSG00000204386	ENST00000375631	Transcript	UPSTREAM	-	-	-	-	-	-	
6_31833357_C/T	6:31833357	T	ENSG00000204386	ENST00000479533	Transcript	UPSTREAM	-	-	-	-	-	-	
6_31833357_C/T	6:31833357	T	ENSG00000204385	ENST00000229729	Transcript	NON_SYNONYMOUS_CODING	1625	1604	535	R/H	cGc/cAc	1KG 6 31833357	SIFT=deleterious; PolyPhen=probably_damaging
6_31833357_C/T	6:31833357	T	ENSG00000204385	ENST00000375562	Transcript	NON_SYNONYMOUS_CODING	1544	1478	493	R/H	cGc/cAc	1KG 6 31833357	SIFT=deleterious; PolyPhen=possibly_damaging
6_31833357_C/T	6:31833357	T	ENSG00000204385	ENST00000544672	Transcript	NON_SYNONYMOUS_CODING	1673	1376	459	R/H	cGc/cAc	1KG 6 31833357	SIFT=deleterious; PolyPhen=probably_damaging
6_31833357_C/T	6:31833357	T	ENSG00000204385	ENST00000487680	Transcript	UPSTREAM	-	-	-	-	-	1KG 6 31833357	-
6_31833357_C/T	6:31833357	T	ENSG00000204385	ENST00000414427	Transcript	DOWNSTREAM	-	-	-	-	-	1KG 6 31833357	-
6_31833357_C/T	6:31833357	T	ENSG00000204385	ENST00000479777	Transcript	DOWNSTREAM	-	-	-	-	-	1KG 6 31833357	-
6_31833357_C/T	6:31833357	T	ENSG00000204385	ENST00000475563	Transcript	DOWNSTREAM	-	-	-	-	-	1KG 6 31833357	-
0204386	ENST00000491768	Transcript	UPSTREAM	-	-	-	-	-	-	-	-	1KG 6 31833357	-
0204386	ENST00000375631	Transcript	UPSTREAM	-	-	-	-	-	-	-	-	1KG 6 31833357	-
0204386	ENST00000479533	Transcript	UPSTREAM	-	-	-	-	-	-	-	-	1KG 6 31833357	-
0204385	ENST00000229729	Transcript	NON_SYNONYMOUS_CODING	1625	1604	535	R/H	cGc/cAc	1KG 6 31833357	SIFT=deleterious; PolyPhen=probably_damaging			
0204385	ENST00000375562	Transcript	NON_SYNONYMOUS_CODING	1544	1478	493	R/H	cGc/cAc	1KG 6 31833357	SIFT=deleterious; PolyPhen=possibly_damaging			
0204385	ENST00000544672	Transcript	NON_SYNONYMOUS_CODING	1673	1376	459	R/H	cGc/cAc	1KG 6 31833357	SIFT=deleterious; PolyPhen=probably_damaging			
0204385	ENST00000487680	Transcript	UPSTREAM	-	-	-	-	-	-	-	-	1KG 6 31833357	-
0204385	ENST00000414427	Transcript	DOWNSTREAM	-	-	-	-	-	-	-	-	1KG 6 31833357	-
0204385	ENST00000479777	Transcript	DOWNSTREAM	-	-	-	-	-	-	-	-	1KG 6 31833357	-
0204385	ENST00000475563	Transcript	DOWNSTREAM	-	-	-	-	-	-	-	-	1KG 6 31833357	-



Variation Pattern Finder

- http://browser.1000genomes.org/Homo_sapiens/UserData/VariationsMapVCF
- VCF input
- Discovers patterns of Shared Inheritance
- Variants with functional consequences considered
- Web output with csv and excel downloads



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

Variation Pattern Finder

Export data: [CSV](#) [Excel](#)

Go to collapsed view

Population ASW	CEU	Freq	Variation info rs9369628:C/T	rs61661828:C/T	rs12192544:C/G	rs599
			6:46620135	6:46620240	6:46620252	6:466
			ENST00000275016 SPLICE_SITE	ENST00000275016 NON_SYNONYMOUS_CODING:R/H	ENST00000275016 NON_SYNONYMOUS_CODING:R/P	ENST0 NON_S
NA20314, NA20322	NA12348, N	0.095	CIC	CIC	GIC	GIG
NA20356, NA19625 and 1 other(s)	NA11919, N	0.092	CIC	CIC	CIG	GIG
NA20291, NA19985 and 5 other(s)		0.069	CIT	CIC	CIC	GIG
NA20289, NA20294 and 4 other(s)		0.057	TIC	CIC	CIC	GIG
	NA12546, N	0.026	CIC	CIC	GIG	GIG
NA19819		0.012	TIT	CIC	CIC	GIG
	NA12283	0.011	TIC	CIC	CIG	GIG
NA19908, NA20278		0.011	CIT	CIC	GIC	GIG
NA19703		0.008	CIC	CIC	CIC	GIG
NA20351		0.007	CIC	CIC	CIC	GIG
		0.006	CIC	CIC	CIG	GIG
NA19712		0.004	CIC	CIC	CIC	CIG
		0.003	CIC	CIC	GIC	GIG
		0.003	TIC	CIC	CIC	GIG
		0.002	CIC	CIC	CIC	GIG



More Help

- <http://www.1000genomes.org/using-1000-genomes-data>
- ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/browser/1000genomes_browser_main_project_20110521/The_1000_Genomes_Browser_Tutorial.ensembl_65.doc
- <http://www.1000genomes.org/ashg-2012-poster>
- <http://www.1000genomes.org/faq>
- **Twitter @1000genomes.org**
- 1000announce@1000genomes.org
- info@1000genomes.org



Questions?

info@1000genomes.org

