The 1000 Genomes Project: A Tutorial

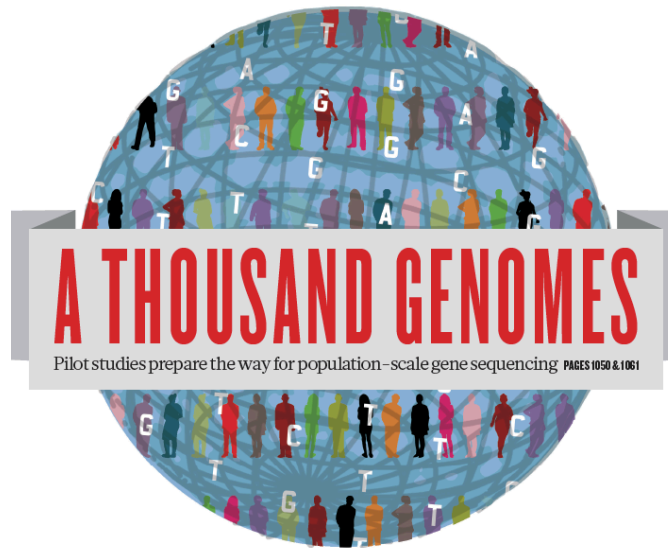# Agenda

- Brief History of the 1000 Genomes Project, data and analysis

- The Raw Data and FTP site

- The Website and Browser

- The 1000 Genomes Tools

EMBL-EBI

Glossary

- **Pilot** : The 1000 Genomes project ran a pilot study between 2008 and 2010

- **Phase 1**: The initial round of exome and low coverage sequencing of 1000 individuals

- **Phase 2**: Expanded sequencing of 1700 individuals and method improvement

- **SAM/BAM**: Sequence Alignment/Map Format, an alignment format

- **VCF**: Variant Call Format, a variant format

EMBL-EBI

The 1000 Genomes Project: A Brief History of Data and Analysis
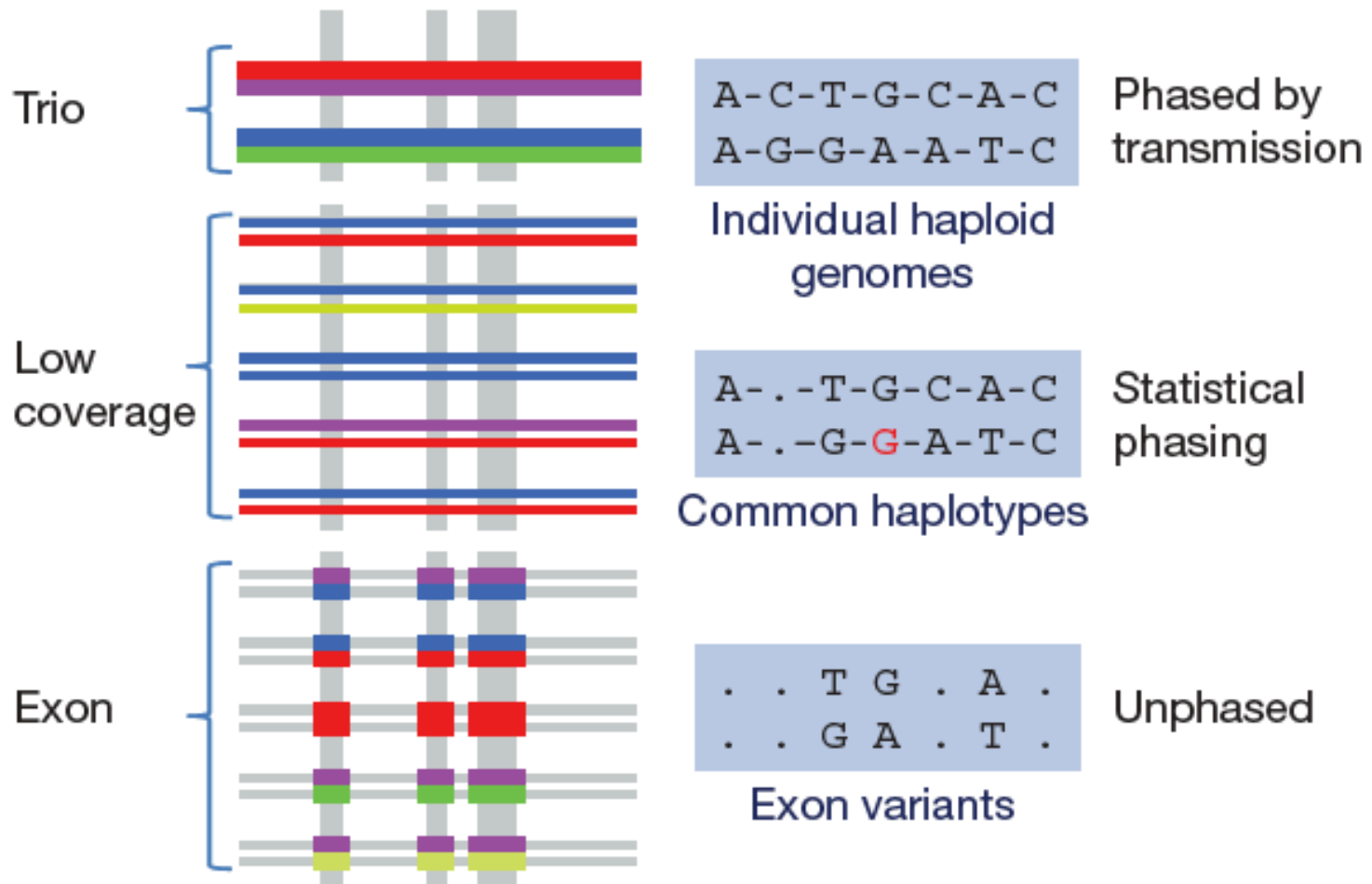
The 1000 Genomes Project

- International project to construct a foundational data set for human genetics
  - Discover virtually all common human variations by investigating many genomes at the base pair level
  - Consortium with multiple centers, platforms, funders
- Aims
  - Discover population level human genetic variations of all types (95% of variation > 1% frequency)
  - Define haplotype structure in the human genome
  - Develop sequence analysis methods, tools, and other reagents that can be transferred to other sequencing projects

EMBL-EBI

3 pilot coverage strategies

Trio

A-C-T-G-C-A-C
A-G-G-A-A-T-C

Phased by transmission

Individual haploid genomes

Low coverage

A-.-T-G-C-A-C
A-.-G-G-A-T-C

Statistical phasing

Common haplotypes

Exon

. . T G . A .
. . G A . T .

Unphased

Exon variants

EMBL-EBI

Main Project Design

- Based on the result of the pilot project, we decided to collect data on 2,500 samples from 5 continental groupings
  - Whole-genome low coverage data (>4x)
  - Full exome data at deep coverage (>20x)
  - A number of deep coverage genomes to be sequenced, with details to be decided
  - High density genotyping at subsets of sites
- Phase 1 Release Integrated Variant Release has been made.

A THOUSAND GENOMES

EMBL-EBI

Phase 1 analysis goal: an **integrated view of human variations**

- Reconstruct haplotypes including all variant types, using all datasets

Deletions

Indels

SNPs (from LC, EX, OMNI)

Goncalo Abecasis
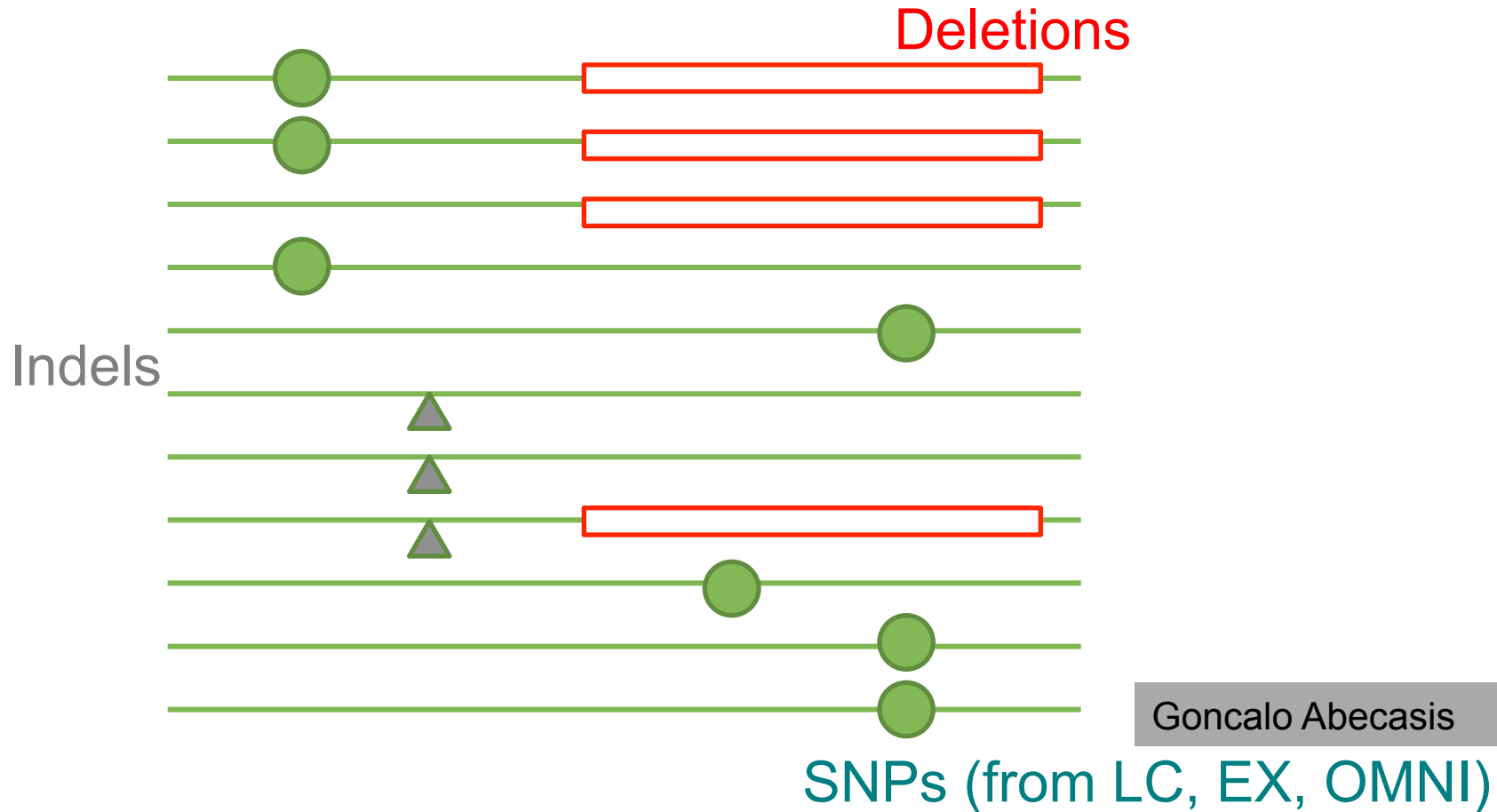
Deep coverage exome data is more sensitive to low-frequency variants

number of sites

● # sites in exomes

● # sites also in low coverage
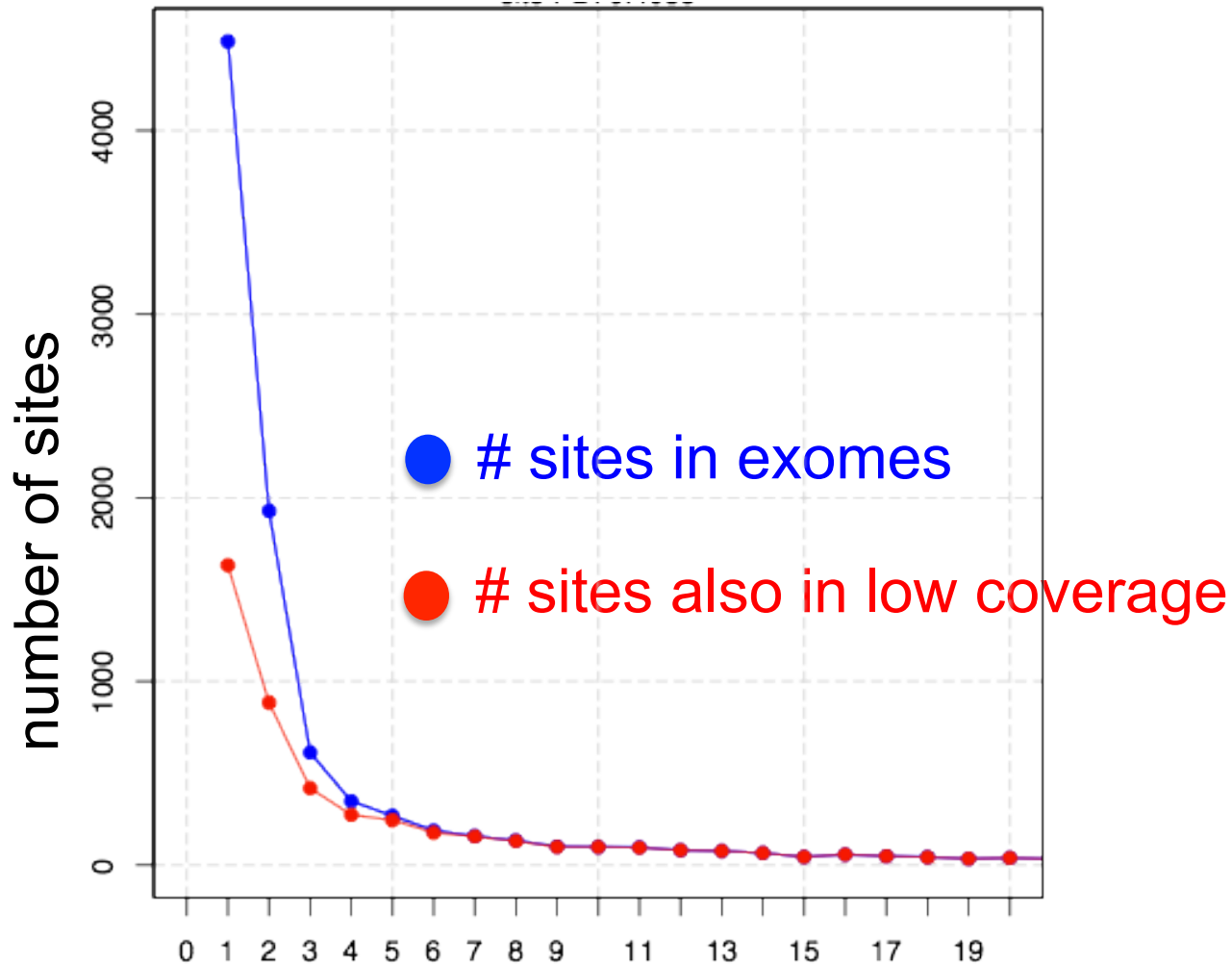
Allele count in 766 exomes (chr. 20, exons only)

Erik Garrison

EMBL-EBI

Newly discovered SNPs are mostly at low frequency and enriched for functional variants

Functional category
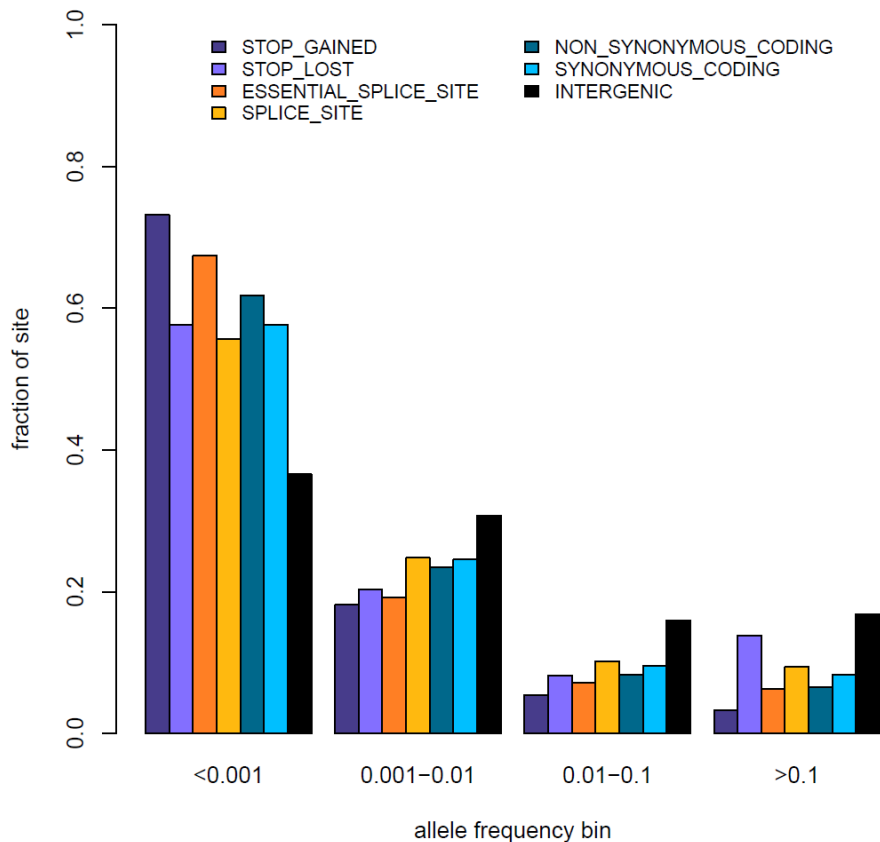
Non-synonymous: Condel score
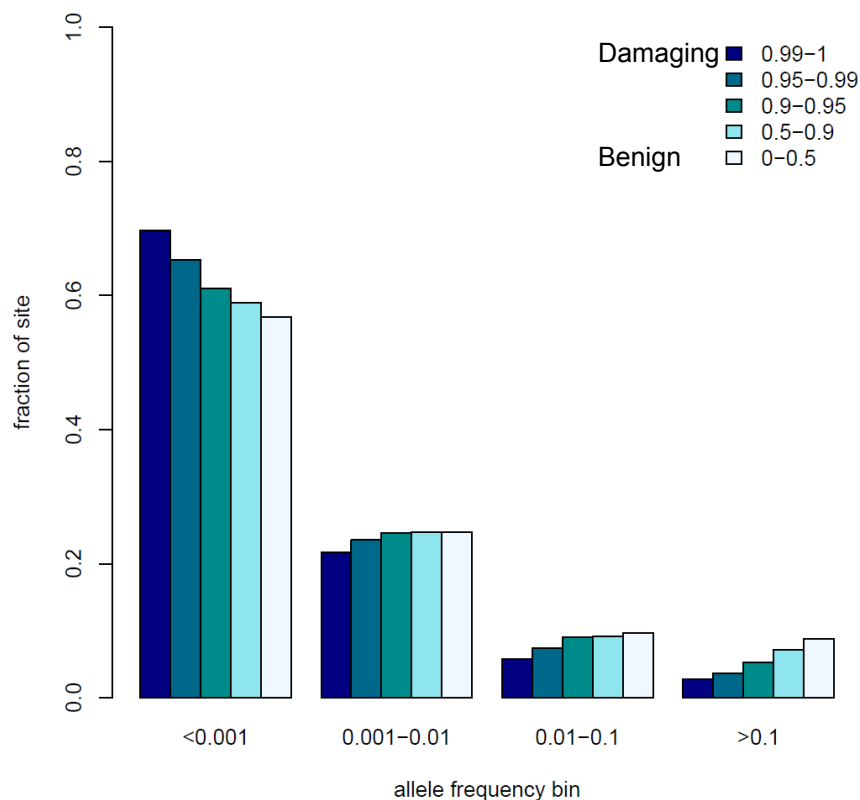
Enza Colonna, Yuan Chen, Yali Xue

Fraction of variant sites present in an individual that are <u>NOT</u> already represented in dbSNP

Date	Fraction <u>not</u> in dbSNP
February, 2000	98%
February, 2001	80%
April, 2008	10%
February, 2011	2%
Now	<1%

Ryan Poplin, David Altshuler

A THOUSAND GENOMES

EMBL-EBI
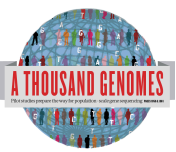
1000 Genomes Project: Present & Future

- First Phase 2 sequence release 14th November 2011
- First Phase 2 alignment release in progress
- First Phase 2 variant site release Summer 2012

- Sample collected expected end to June 2012
- Final Phase 3 Sequence release expected December 2012
- 2013 will represent finalization of 1000 genomes analysis results and final data releases

EMBL-EBI

Hapmap, The Pilot Project and The Main Project

- **Hapmap**
  - Starting in 2002
  - Last release contained ~3m snps
  - 1400 individuals
  - 11 populations
  - High Throughput genotyping chips

- **1000 Genomes Pilot project**
  - Started in 2008
  - Paper release contained ~14 million snps
  - 179 individuals
  - 4 populations
  - Low coverage next generation sequencing

- **1000 Genomes Phase 1**
  - Started in 2009
  - Phase 1 release has 36.6millon snps, 3.8millon indels and 14K deletions
  - 1094 individuals
  - 14 populations
  - Low coverage and exome next generation sequencing

- **1000 Genomes Phase 2**
  - Started in 2011
  - 1715 individuals
  - 19 Populations
  - Low coverage and exome next generation sequencing

A THOUSAND GENOMES

EMBL-EBI

Timeline

- **September 2007:** 1000 Genomes project formally proposed Cambridge, UK
- **April 2008:** First Submission of Data to the Short Read Archive.
- **May 2008:** First public data release.
- **October 2008:** SAM/BAM Format Defined.
- **December 2008:** First High Coverage Variants Released.
- **December 2008:** First 1000 genomes browser released
- **May 2009:** First Indel Calls released.
- **July 2009:** VCF Format defined
- **August 2009:** First Large Scale Deletions released.
- **December 2009:** First Main Project Sequence Data Released.
- **March 2010:** Low Coverage Pilot Variant Release made
- **July 2010:** Phased genotypes for 159 Individuals released.
- **October 2010:** A Map of Human Variation from population scale sequencing is published in Nature.

- **January 2011:** Final Phase 1 Low coverage alignments are released
- **May 2011:** @1000genomes appears on Twitter
- **May 2011:** First Variant Release made on more than 1000 individuals
- **October 2011:** Phase 1 integrated variant release made

EMBL-EBI

Sequencing Data

- The Project contains data from 3 different providers and multiple platforms

Platform	Min Read Length (bp)	Max Read Length (bp)
454 Roche GS FLX Titanium	70	400
Illumina GA	30	81
Illumina GA II	26	160
Illumina HiSeq	50	102
ABI Solid System 2.0	25	35
ABI Solid System 2.5	50	50
ABI Solid System 3.0	50	50

EMBL-EBI

Alignment Data

- The project has made more than 10 releases of Alignment Data
- Pilot Project
  - Aligned to NCBI36
  - Maq and Corona
  - Base Quality Recalibration done
- Phase 1
  - Aligned to GRCh37
  - BWA and Bfast
  - Indel Realignment
- Phase 2
  - Aligned to extended GRCh37
  - Improvements to Base Quality Recalibration

EMBL-EBI

Variant Calling

- Early call sets used a single variant caller
- Intersect approach developed during pilot
- Variant Quality Score Recalibration (VQSR) developed for Phase 1
- Genotype Likelihoods assigned to help with genotype calling
- Integrated genotype calling based on individual variant call sets
- Phase 2 looks to improve site discovery and improve integration

EMBL-EBI

Data Availability

- FTP site: ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/
  - Raw Data Files

- Web site: http://www.1000genomes.org
  - Release Announcements
  - Documentation

- Ensembl Style Browser: http://browser.1000genomes.org
  - Browse 1000 Genomes variants in Genomic Context
  - Variant Effect Predictor
  - Data Slicer
  - Other Tools

EMBL-EBI

The 1000 Genomes Project: The Raw Data

What is available on the ftp site

- Sequence Data
  - Fastq files
  - @ERR050087.1 HS18_6628:8:1108:8213:186084#2/1
  - GGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGG
  - +
  - DCDHKHKKIJGNNHIJIIKLLMCLKMAILIJH3K>HL1I=>MK.D<

- Alignment Data
  - BAM files
  - ERR052835.20962733   163   11   60239   0   100M   =   60609   469

- Variant Calls
  - VCF files
  - 1   10523   .   TCCG   T   152   PASS   VT=INDEL;RSQ=0.5246; AFR_AF=0.01

- Reference Data Sets
  - Reference genome in fasta
  - Annotation sets in bed or gtf

EMBL-EBI

Data formats and key tools

Sequence analysis

The Sequence Alignment/Map format and SAMtools

Heng Li[1,†], Bob Handsaker[2,†], Alec Wysoker[2], Tim Fennell[2], Jue Ruan[3], Nils Homer[4],
Gabor Marth[5], Goncalo Abecasis[6], Richard Durbin[1,*] and 1000 Genome Project Data
Processing Subgroup[7]

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

BAM alignment files

Sequence analysis Advance Access publication June 7, 2011

The variant call format and VCFtools

Petr Danecek[1,†], Adam Auton[2,†], Goncalo Abecasis[3], Cornelis A. Albers[1], Eric Banks[4],
Mark A. DePristo[4], Robert E. Handsaker[4], Gerton Lunter[2], Gabor T. Marth[5],
Stephen T. Sherry[6], Gilean McVean[2,7], Richard Durbin[1,*] and 1000 Genomes Project
Analysis Group[‡]

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

VCF variant files

Sequence analysis Advance Access publication January 5, 2011

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

All indexed for fast retrieval

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

A THOUSAND GENOMES

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	118 KB	05/01/2012 5/01/2012 12 :40:00
README.alignment_data	12 KB	26/01/2011 26/01/2011 12 :00:00
README.ftp_structure	9 KB	04/04/2011 4/04/2011 12 :00:00
README.pilot_data	3 KB	14/...	Documentation	...912 :00:00
README.populations	2 KB	18/02/2010 18/02/2010 12 :00:00
README.sequence_data	7 KB	23/...	Raw Data	...19 :03:00
alignment_indices		14/07/2011 14/07/2011 10 :53:00
changelog_details		05/...	Phase 1 Data	...12 :40:00
current.tree	29933 KB	05/...	...12 :37:00
data		04/...	Pilot Data	...8 :50:00
phase1		14/...	Release Data	...14 :03:00
pilot_data		27/...	...012 :00:00
release		12/...	Technical Data	...13 :18:00
sequence.index	27185 KB	20/12/2011 20/12/2011 12 :26:00
sequence_indices		14/11/2011 14/11/2011 10 :10:00
technical		13/12/2011 13/12/2011 1 :05:00

A THOUSAND GENOMES

EMBL-EBI

The FTP Site: Data

FTP Site: Technical

FTP Site: Release

Index of ftp://ftp.1000genomes...   +

ftp://ftp.1000genomes.**ebi.ac.uk**/vol1/ftp/release/   W ▾ Wikipedi

Maps 17 dates fix lj docs plus gm g fb ds lj NCBI p E SRA C 1KG JIRA Wish am

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

⬆ Up to higher level directory

Name	Size	Last Modified
📁 2008_12		2 :00:00
📁 2009_02	21/02/2009	21/02/200912 :00:00
📁 2009_04	07/05/2009	7/05/200912 :00:00
📁 2009_05	08/06/2009	8/06/200912 :00:00
📁 2009_08	10/08/2009	10/08/200912 :00:00
📁 20100804		0:00
📁 20101123		0:00
📁 2010_11	16/02/2011	16/02/201112 :00:00
📁 20110521	16/12/2011	16/12/201110 :09:00

Older Release Dirs

Sequence Index Dates

20110521.sequence.index	23693 KB	19/07/2011	19/07/201112 :00:00
20110521.sequence.index.exome.stats	48 KB	19/07/2011	19/07/201112 :00:00
20110521.sequence.index.low_coverage.stats	53 KB	21/05/2011	21/05/201112 :00:00
20110521_20110719.exome.stats.csv	2 KB	19/07/2011	19/07/201112 :00:00
20110521_20110719.low_coverage.stats.csv	2 KB	19/07/2011	19/07/201112 :00:00
20110719.sequence.index	23961 KB	19/07/2011	19/07/201112 :00:00
20110719.sequence.index.exome.stats	52 KB	10/10/2011	10/10/201110 :10:00
20110719.sequence.index.low_coverage.stats	54 KB	10/10/2011	10/10/201110 :13:00
20110719_20110920.exome.stats.csv	1 KB	10/10/2011	10/10/20119 :45:00
20110719_20110920.low_coverage.stats.csv	2 KB	10/10/2011	10/10/20119 :45:00

A THOUSAND GENOMES

EMBL-EBI

FTP Site: Pilot Data

The 1000 Genomes Project: Finding Data

Finding Data

- Current.tree file

- ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/current.tree

- Current Tree is updated nightly so can be upto 24 hours out of date

Finding Data

- FTP search and Website Search
- http://www.1000genomes.org/ftpsearch

Viewing Files

- All alignment and variant files are indexed so subsections can be downloaded remotely

- Use samtools to get subsections of bam files
    - **samtools view** http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data/HG01375/alignment/HG01375.mapped.ILLUMINA.bwa.CLM.low_coverage.20111114.bam 6:31833200-31834200

- Use tabix to get subsections of vcf files
    - **tabix -h** ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/working/20120131_omni_genotypes_and_intensities/Omni25_genotypes_2141_samples.b37.vcf.gz 6:31833200-31834200

- You can also use the web Data Slicer interface to do this
    - http://browser.1000genomes.org/Homo_sapiens/UserData/SelectSlice

EMBL-EBI

More Information

- Sam/Bam format
- http://samtools.sourceforge.net/
- samtools-help@lists.sourceforge.net
- VCF format
- http://vcftools.sourceforge.net/
- vcftools-help@lists.sourceforge.net

EMBL-EBI
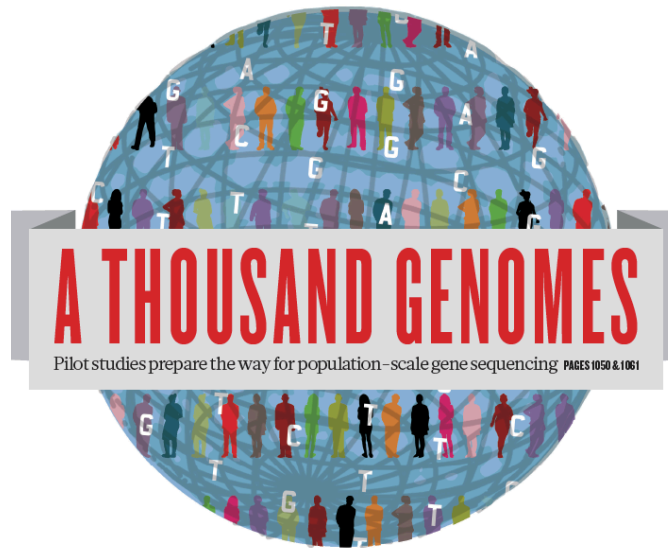
The 1000 Genomes Website and Ensembl- style Browser

EMBL-EBI

http://www.1000genomes.org

http://browser.1000genomes.org

Searching the Browser

- *http://browser.1000genomes.org*

- Search for PTPN22

- Click 'Region in Detail'

Region in Detail

Turning on Tracks

Genes and SNPs

Coding

UTR
Intron

Reverse strand   57.95 Kb

Chromosome bands   133,099,000   133,100,000   133,101,000   133,102,000
ALL - interim phas...
Contigs

Line indicates number of SNPS   Each Line is One SNP

A THOUSAND GENOMES

EMBL-EBI

File upload to view with 1000 Genomes data

Manage your data

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

EMBL-EBI

Uploaded VCF

Example:

EMBL-EBI
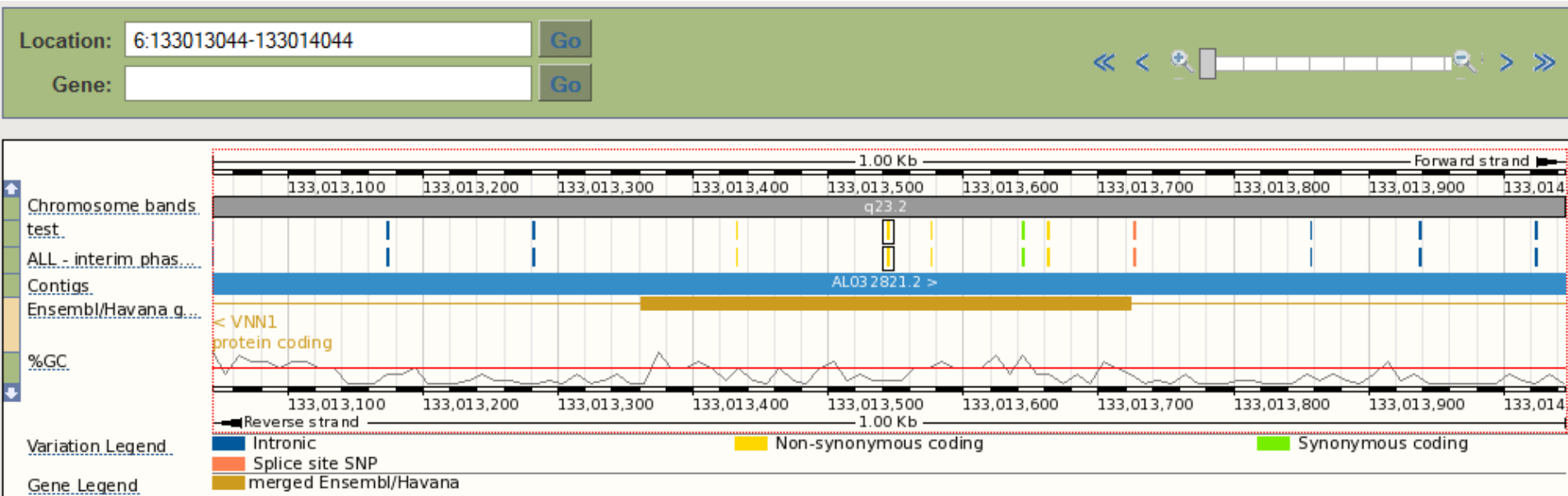
Uploaded BAM

Example:

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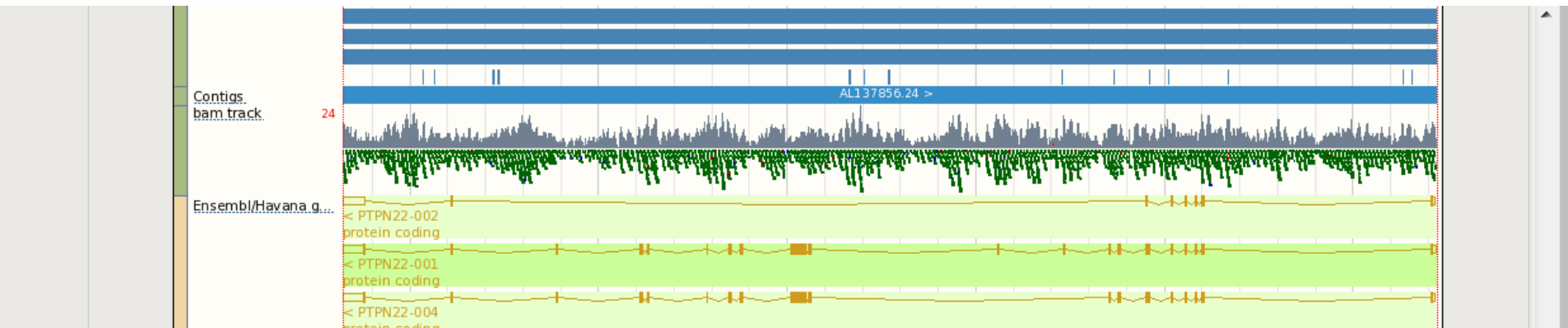
Back to browsing…

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

Download as csv

Get in vcf format

EMBL-EBI

Structural variation (in the Gene tab)

Variation Image

- Gene variation zoom

Transcript Tab: Variations

Effect on Protein:

- SIFT

- PolyPhen

Start again- search for a variation (rs31685)
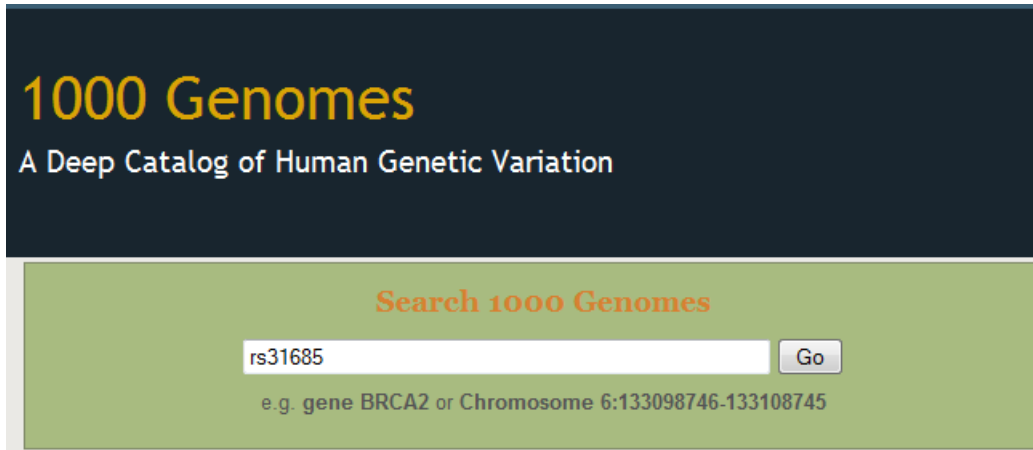
- The Variation tab- left hand links take you to more information

- Population

The Browser: Coming Soon

The 1000 Genomes Project:

The 1000 Genomes Tools

The 1000 Genomes Tools

- Data Slicer

- Variant Effect Predictor

- Variation Pattern Finder

- VCF to PED

- API and Database access

EMBL-EBI

http://browser.1000genomes.org

Tools page

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

EMBL-EBI

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

ℹ **Variant Effect Predictor:**
This tool takes a list of variant positions and alleles, and predicts the effects of each of these on overlapping transcripts and regulatory regions annotated in Ensembl. The tool accepts substitutions, insertions and deletions as input, uploaded as a list of tab separated values, VCF or Pileup format input.

Upload is limited to 750 variants; lines after the limit will be ignored. Users with more than 750 variations can split files into smaller chunks, use the standalone perl script or the variation API. See also full documentation.

Input file

Species:	Human (Homo sapiens): GRCh37 ▼
Name for this upload (optional):	[_____]
Paste file:	[text area]
Upload file:	Choose File no file selected
or provide file URL:	[_____]
Input file format:	Ensembl default ▼

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:	No ▼
PolyPhen predictions:	No ▼
Condel consensus (SIFT/PolyPhen) predictions:	No ▼

Frequency filtering of existing variants (human only)

Filter variants by frequency:	☐

NB: Enabling frequency filtering may be very slow for large datasets

Filter: Exclude ▼ variants with MAF greater than ▼ 0.1 in any 1KG low coverage population ▼

Next >

Variant Effect Predictor

- **perl variant_effect_predictor.pl** -input 6_381831625_3184704.vcf -sift p -polyphen p –check_existing

- less variant_effect_output.txt

```
#Uploaded_variation    Location        Allele  Gene    Feature Feature_type    Consequence
cDNA_position   CDS_position    Protein_position        Amino_acids     Codons  Exi
sting_variation Extra
rs138094825     6:31831667      A       ENSG00000204385 ENST00000414427 Transcript
DOWNSTREAM      -       -       -       -       -       rs138094825     -
rs138094825     6:31831667      A       ENSG00000204385 ENST00000229729 Transcript
INTRONIC        -       -       -       -       -       rs138094825     -
6_31832657_C/T  6:31832657      T       ENSG00000204385 ENST00000229729
Transcript      NON_SYNONYMOUS_CODING   1883    1862    621     R/H     cGc/cAc -
PolyPhen=possibly_damaging;SIFT=deleterious
```

EMBL-EBI

Data Slicing

- Use samtools to get subsections of bam files
  - **samtools view** http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data/ HG01375/alignment/ HG01375.mapped.ILLUMINA.bwa.CLM.low_coverage. 20111114.bam 6:31833625-31833704

- Use tabix to get subsections of vcf files
  - **tabix -h** ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/ working/20120131_omni_genotypes_and_intensities/ Omni25_genotypes_2141_samples.b37.vcf.gz 6:31830969-31846823 | vcf-subset -c HG01375

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

EMBL-EBI

Data Slicing

Variation Pattern Finder

- Remote or local tabix indexed VCF input

- Discovers patterns of Shared Inheritance

- Variants with functional consequences considered by default

- Web output with CSV and Excel downloads

- http://browser.1000genomes.org/Homo_sapiens/UserData/VariationsMapVCF

EMBL-EBI

Variation Pattern Finder

ⓘ 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:

```
ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20101123
/interim_phase1_release
/ALL.chr6.phase1.projectConsensus.genotypes.vcf.gz
```
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:

```
ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20101123
/interim_phase1_release/interim_phase1.20101123.ALL.panel
```
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 >

Variation Pattern Finder

- **perl variant_pattern_finder.pl** -vcf ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20110521/ALL.chr6.phase1_integrated_calls.20101123.snps_indels_svs.genotypes.vcf.gz -sample_panel_file ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20110521/phase1_integrated_calls.20101123.ALL.panel -region 6:31830969-31846823 -expand

Variation Pattern Finder Output

freq	6:31833647_.[T]		6:31833660_rs6915800[G]		samples
freq	ENST00000414427-SPLICE_SITE[],ENST00000544672-SPLICE_SITE[],ENST00000229729-SPLICE_SITE[],ENST00000375562-SPLICE_SITE[]		ENST00000414427-NON_SYNONYMOUS_CODING[R/C],ENST00000229729-NON_SYNONYMOUS_CODING[R/C],ENST00000544672-NON_SYNONYMOUS_CODING[R/C],ENST00000375562-NON_SYNONYMOUS_CODING[R/C]		samples
0.73	REF|REF		G|A	YRI(3)	NA18933, NA19149, NA19098 and 0 others.
0.27	REF|REF		A|G	YRI(2)	NA19146, NA19198
0.18	REF|REF		A|A	LWK(1)	NA19372
0.09	C|T		REF|REF	CHB(1)	NA18592

VCF to PED

- LD Visualization tools like Haploview require PED files
- VCF to PED converts VCF to PED
- Will a file divide by individual or population
- http://browser.1000genomes.org/Homo_sapiens/UserData/Haploview

VCF to PED

VCF to PED

- **perl vcf_to_ped_convert.pl** -vcf ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20110521/ALL.chr6.phase1_integrated_calls.20101123.snps_indels_svs.genotypes.vcf.gz -sample_panel_file ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20110521/phase1_integrated_calls.20101123.ALL.panel -region 6:31830969-31846823 -population CEU
- Output should be two files
- 6_31830969-31846823.info
- 6_31830969-31846823.ped

EMBL-EBI

Haploview

- haploview

Access to backend Ensembl databases

- Public MySQL database at
  - mysql-db.1000genomes.org port 4272

- Full programmatic access with Ensembl API
  - The 1000 Genomes Pilot uses Ensembl v60 databases and the NCBI36 assembly (this is frozen)
  - The 1000 Genomes main project currently uses Ensembl v63 databases
- http://jun2011.archive.ensembl.org/info/docs/api/variation/index.html
- http://www.ensembl.org/info/docs/api/variation/index.html
- http://www.1000genomes.org/node/517

A THOUSAND GENOMES

EMBL-EBI

The 1000 Genomes Project:

Finding out about New Data

Announcements

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

- Please email if you have any questions or feedback about this resource

EMBL-EBI

Thanks

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

EMBL-EBI