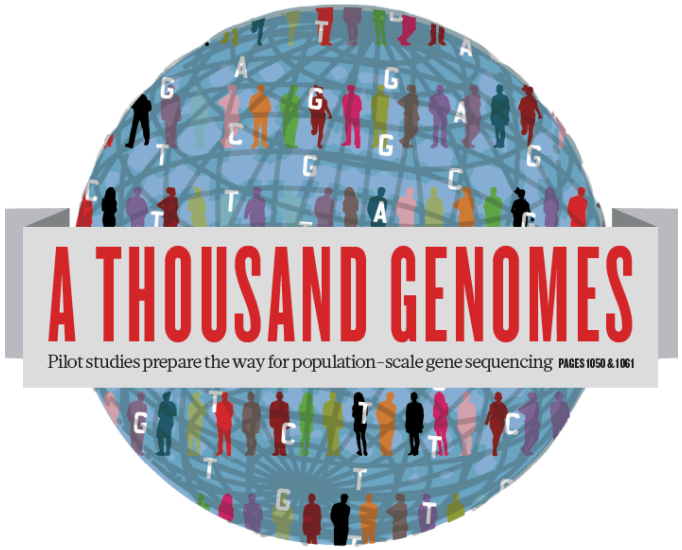


The 1000 Genomes Tutorial
The Website and Browser

Laura Clarke

17th February 2012

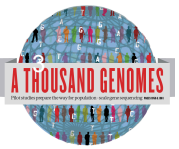
EMBL-EBI

These slides should give you an overview of the 1000 genomes website and our Ensembl style browser

EMBL-EBI

Glossary

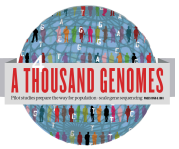
- **Pilot** : The 1000 Genomes project ran a pilot study between 2008 and 2010
- **Phase 1**: The initial round of exome and low coverage sequencing of 1000 individuals
- **Phase 2**: Expanded sequencing of 1700 individuals and method improvement
- **SAM/BAM**: Sequence Alignment/Map Format, an alignment format
- **VCF**: Variant Call Format, a variant format



EMBL-EBI

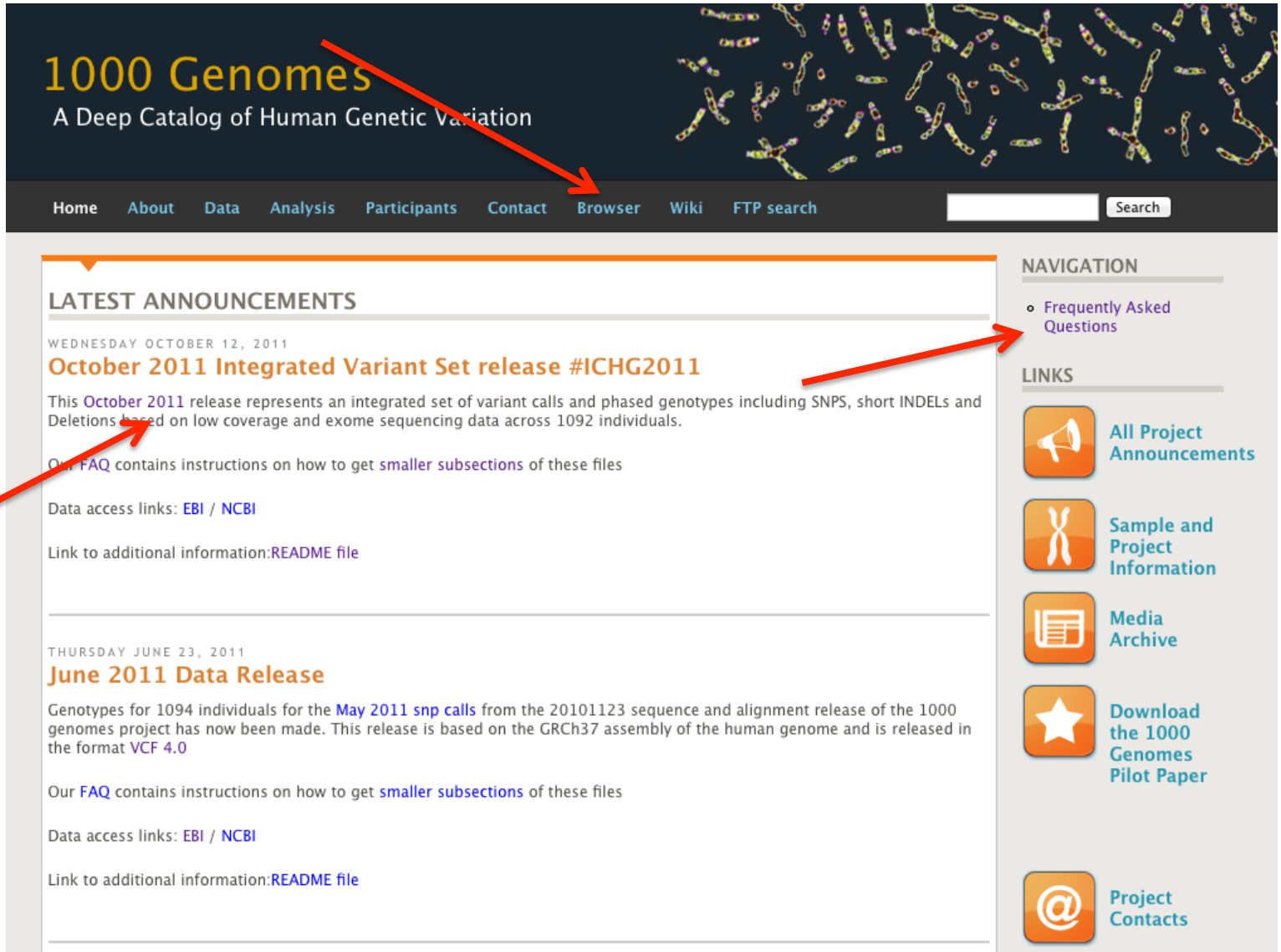
Summary

- 1000 Genomes Website
- 1000 Genomes Browser
- Gene and SNP display on Region in Detail
- Searching for a Gene
- Region in Detail
- Turning on Tracks
- File upload
- Gene View
- Structural Variation
- Gene Variation Zoom
- Transcript View
- Searching for a Variation
- Population Genetics
- New Variation Views
- Data Availability
- Announcements



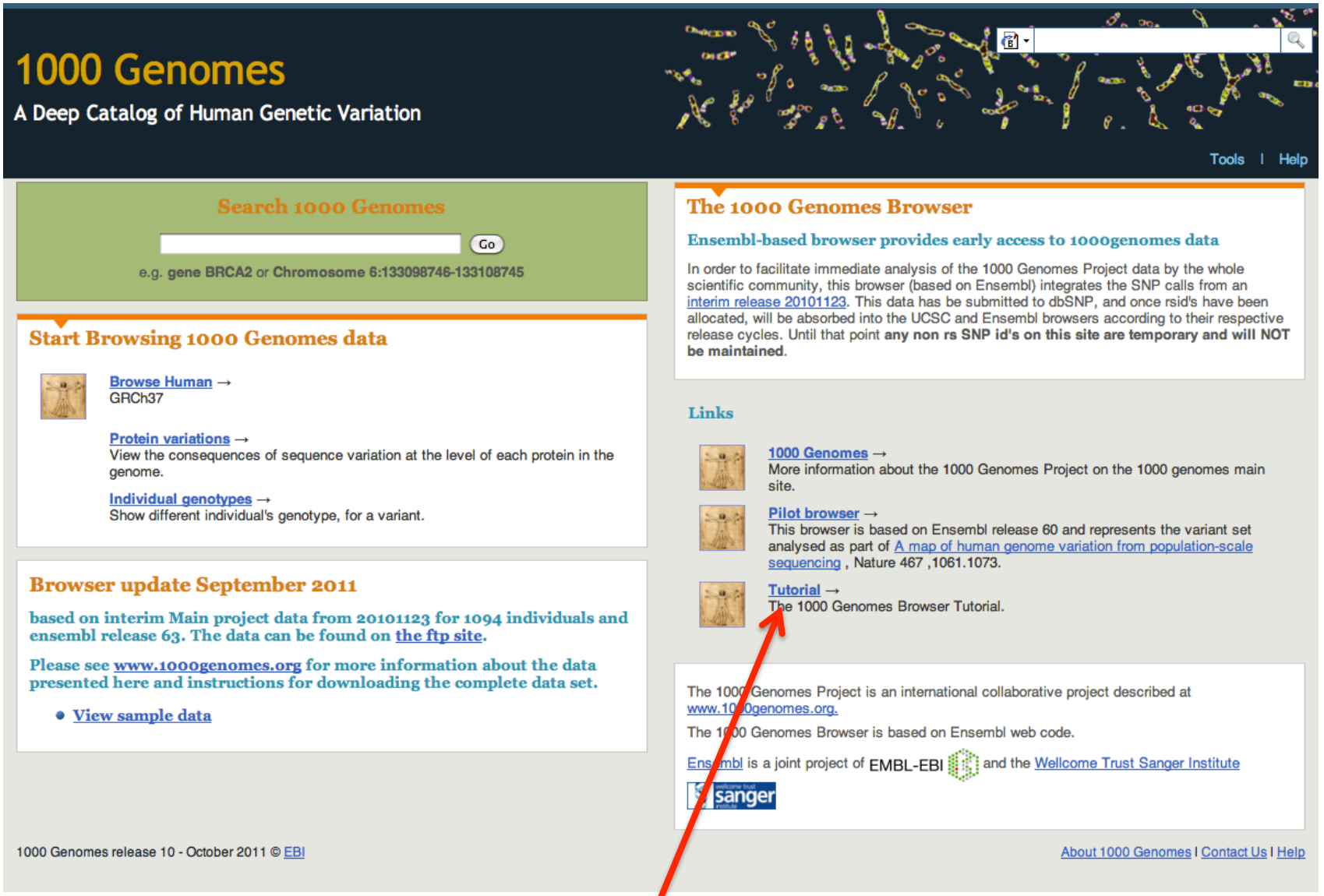
EMBL-EBI

http://www.1000genomes.org



A THOUSAND GENOMES

EMBL-EBI



http://browser.1000genomes.org

A THOUSAND GENOMES

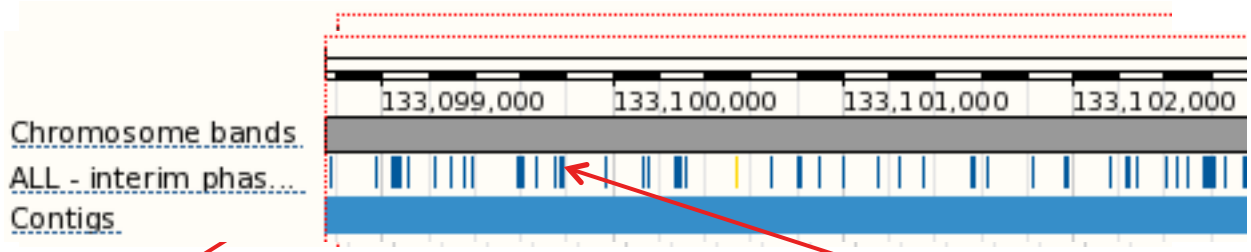
EMBL-EBI

Genes and SNPs

UTR

Intron

Coding

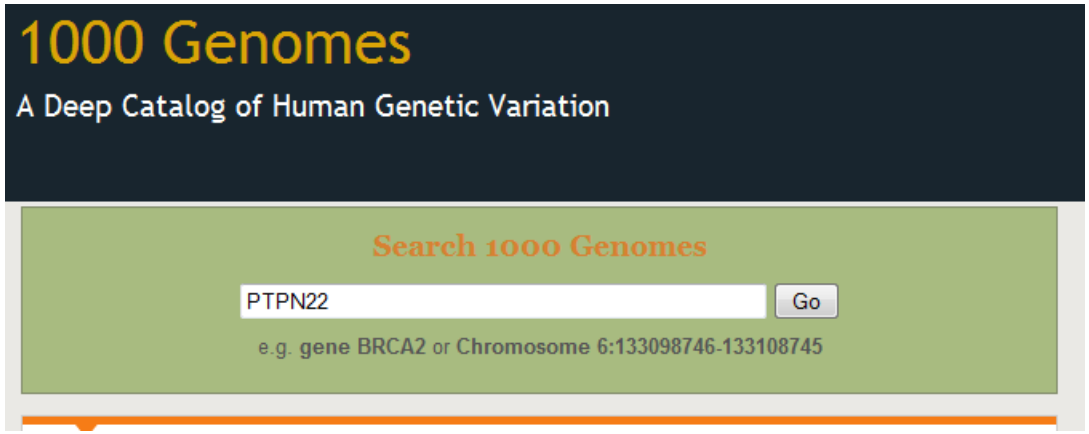


Line indicates number of SNPS

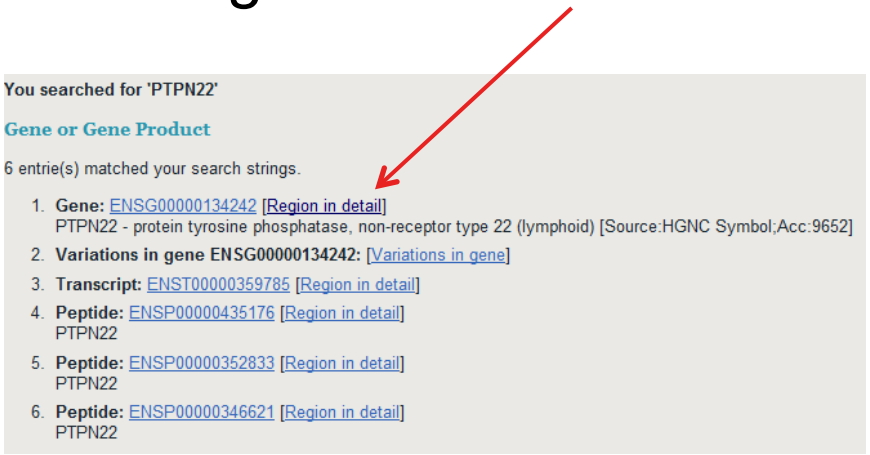
Each Line is One SNP

Searching the Browser

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



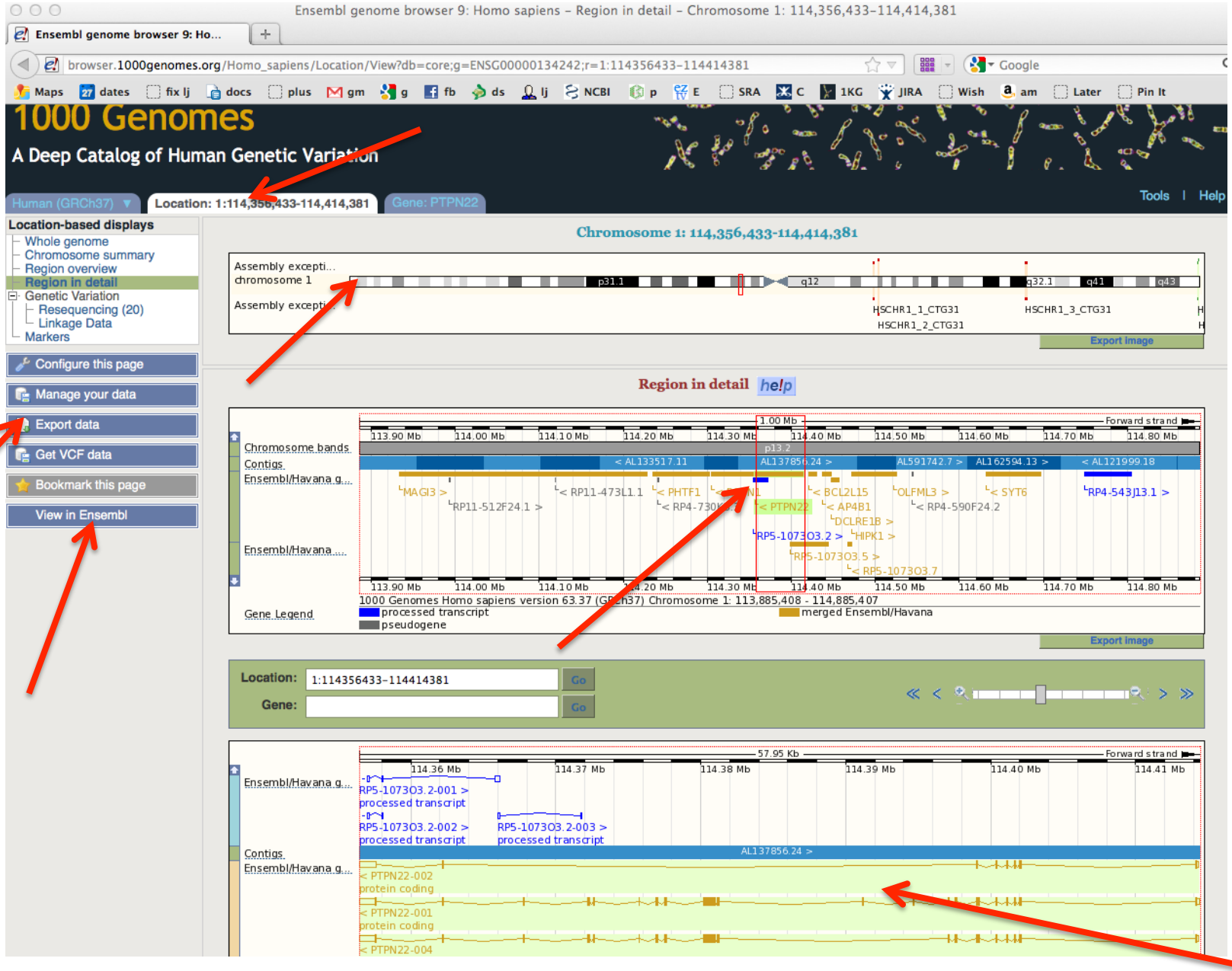
- Search for PTPN22
- Click ‘Region in Detail’



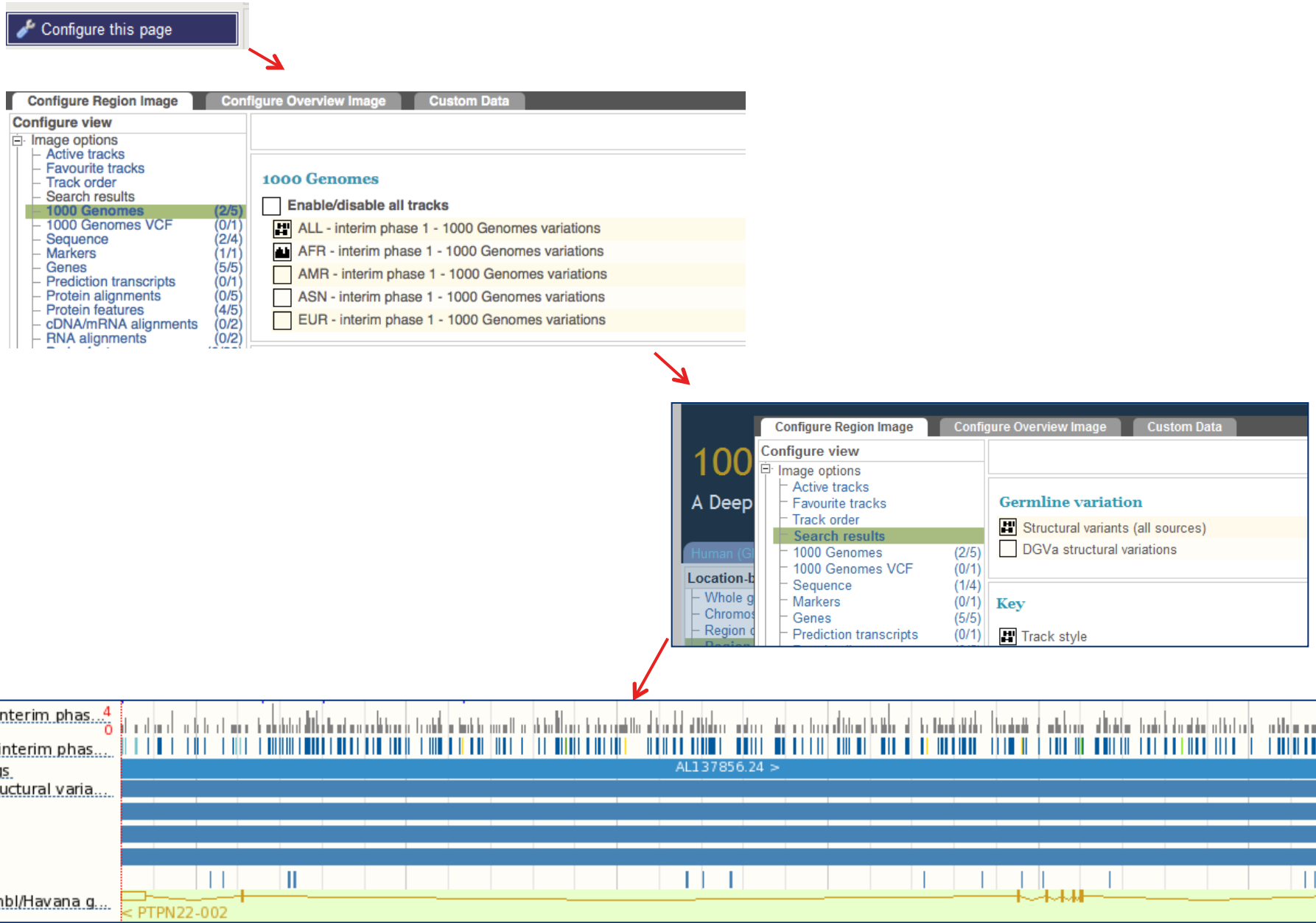
A THOUSAND GENOMES

EMBL-EBI

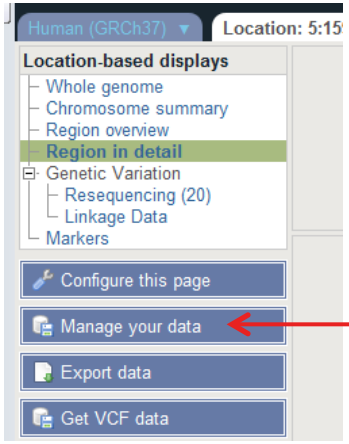
Region in Detail



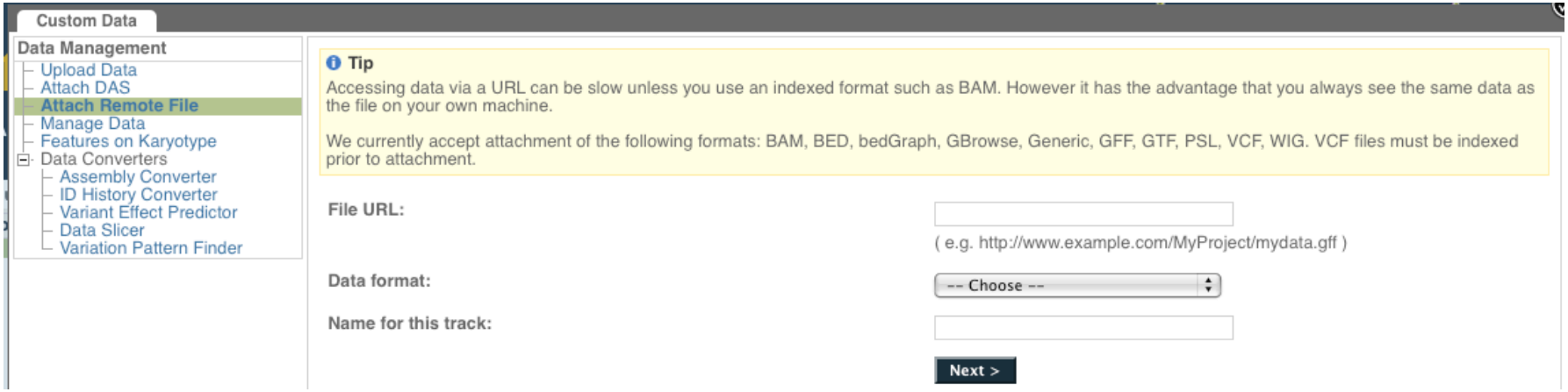
Turning on Tracks



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

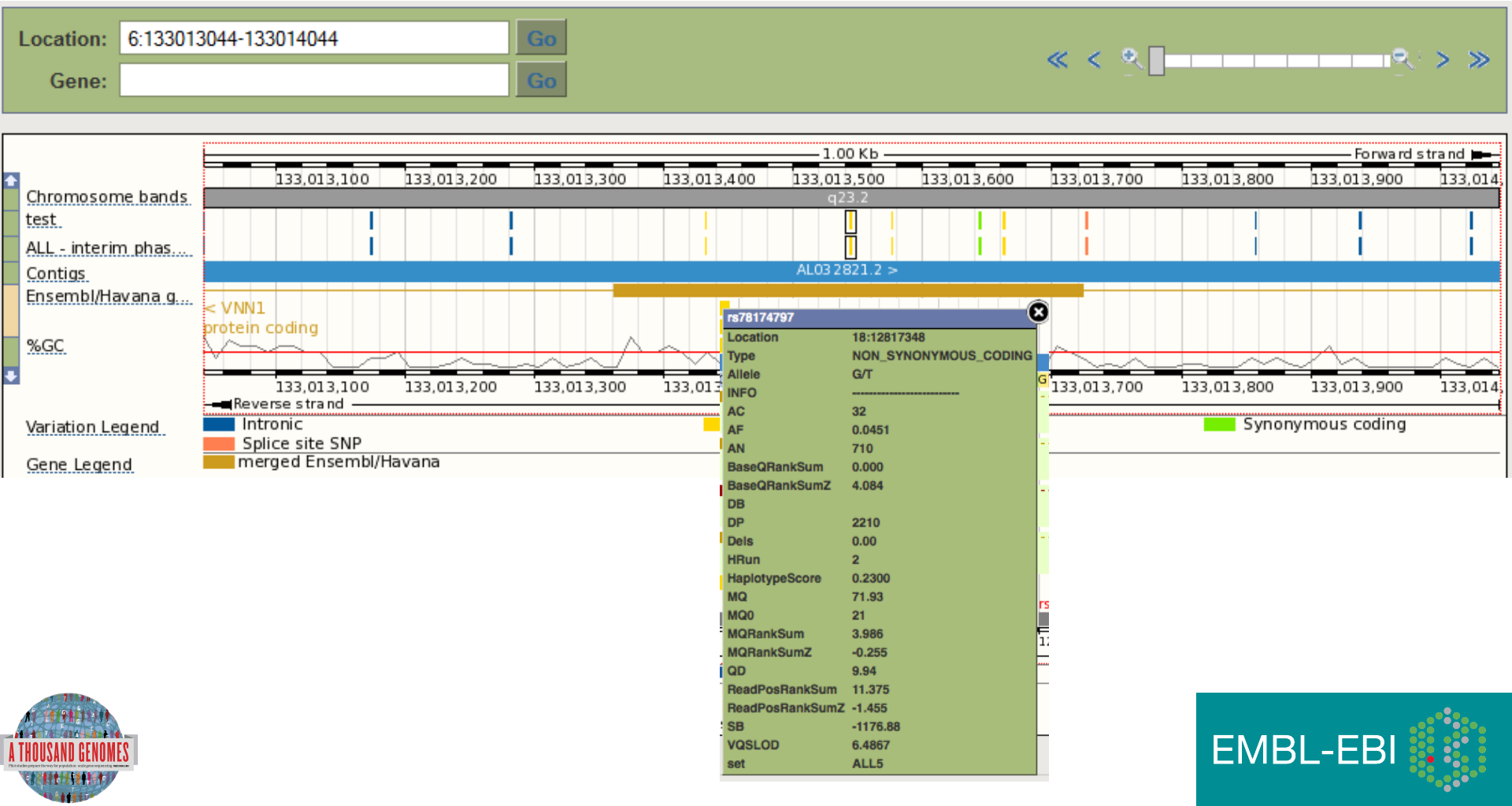
A THOUSAND GENOMES

EMBL-EBI

Uploaded VCF

Example:

ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20110521/ALL.wgs.phase1_release_v2.20101123.snps_indels_sv.sites.vcf.gz



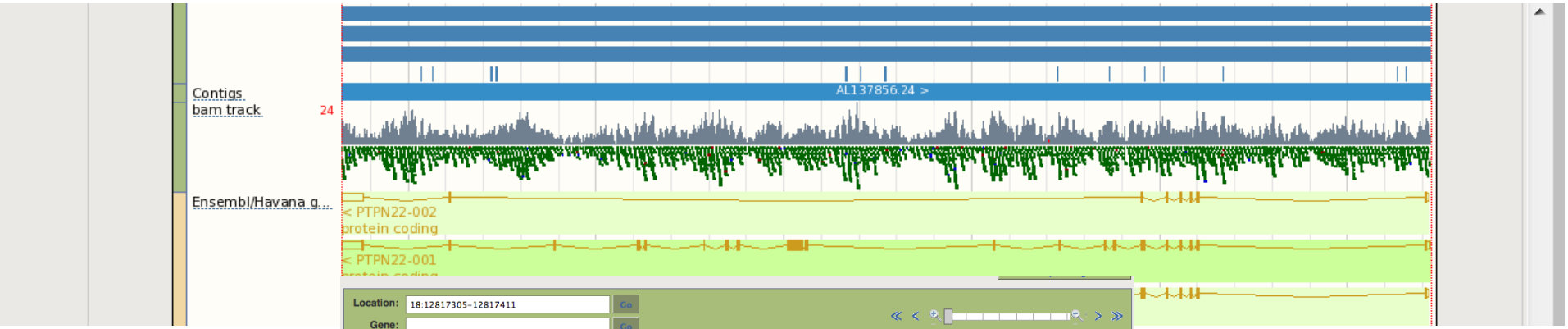
A THOUSAND GENOMES

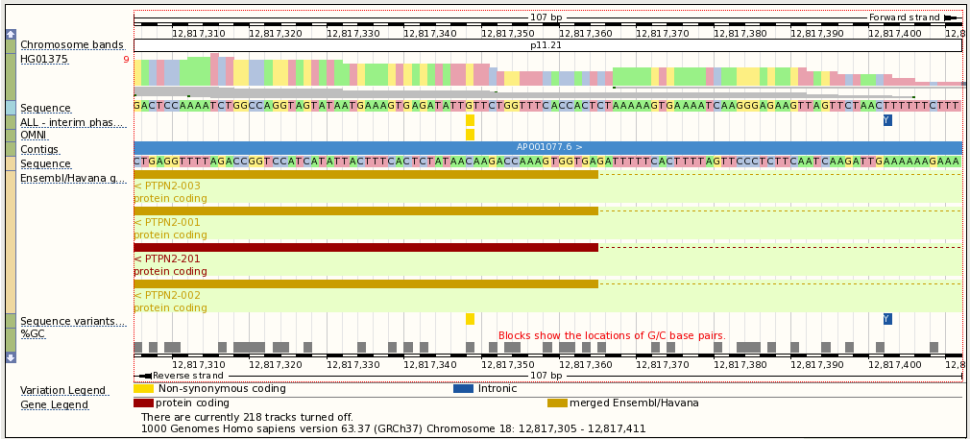
EMBL-EBI

Uploaded BAM

Example:

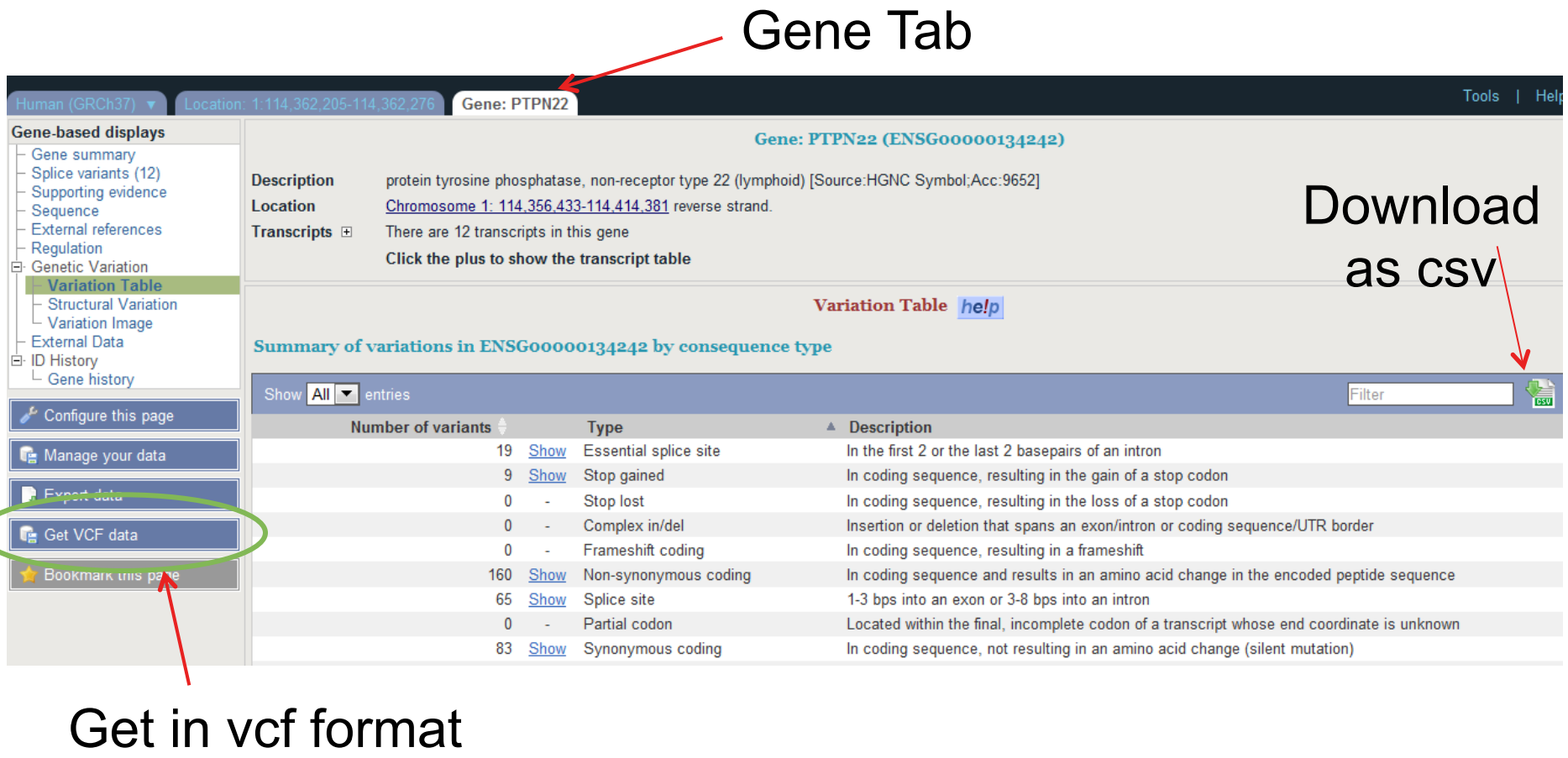
http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data/HG01375/alignment/HG01375.mapped.ILLUMINA.bwa.CLM.low_coverage.20111114.bam





Back to browsing…

Click the Gene tab, then ‘Variation Table’ or ‘Variation Image’

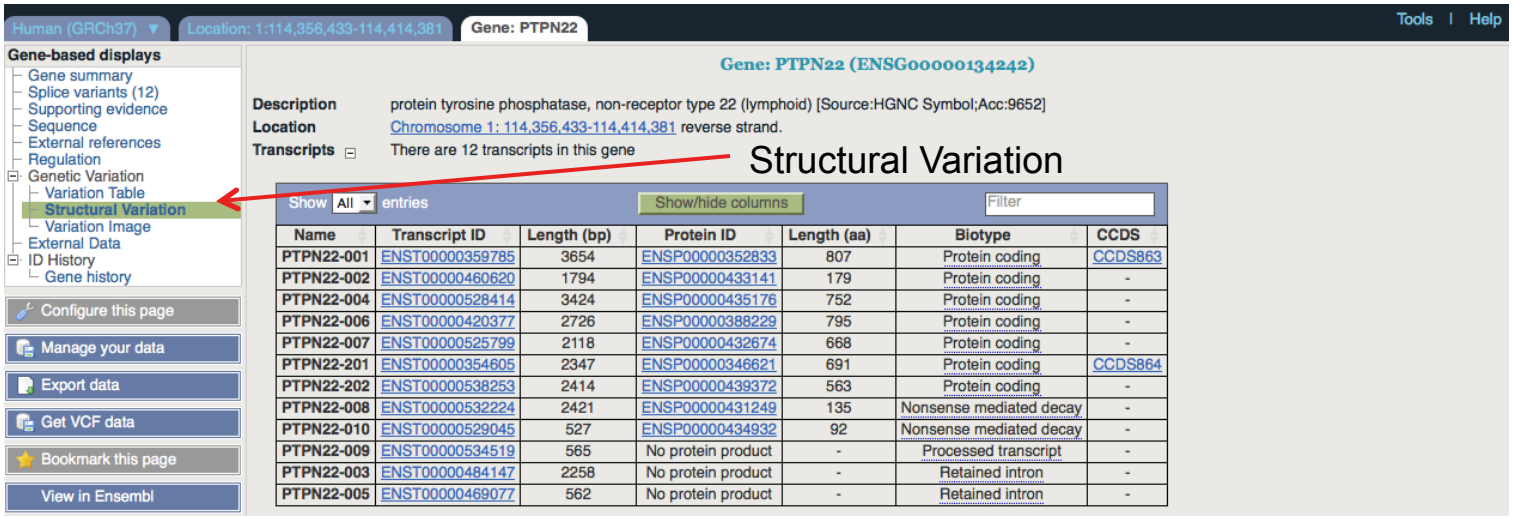


Get in vcf format

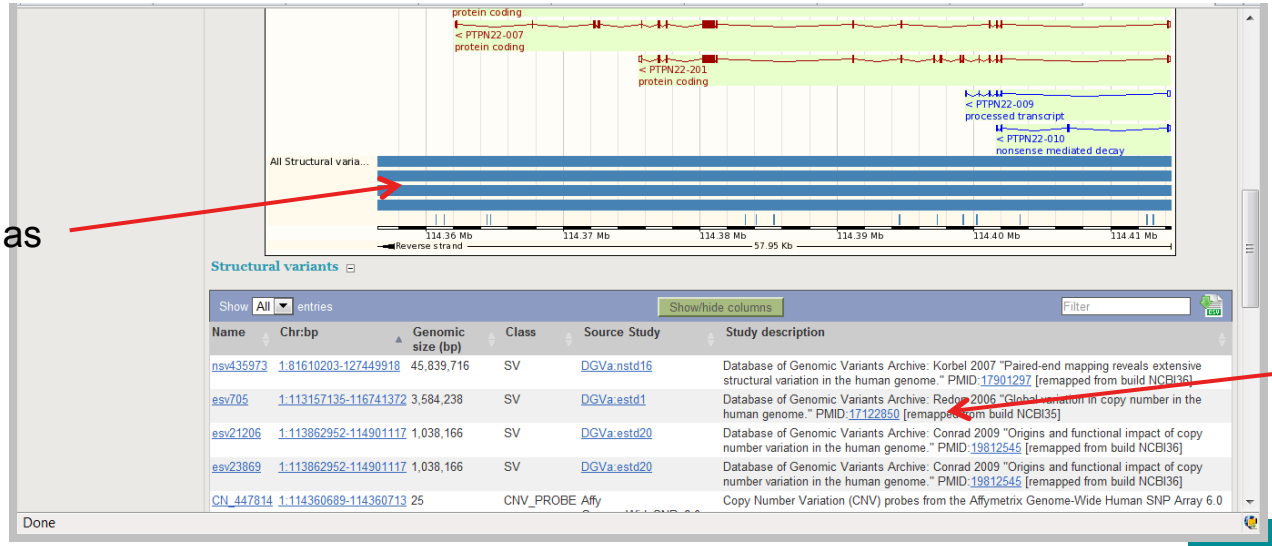
A THOUSAND GENOMES

EMBL-EBI

Structural variation (in the Gene tab)



Structural Variants as boxes



Table

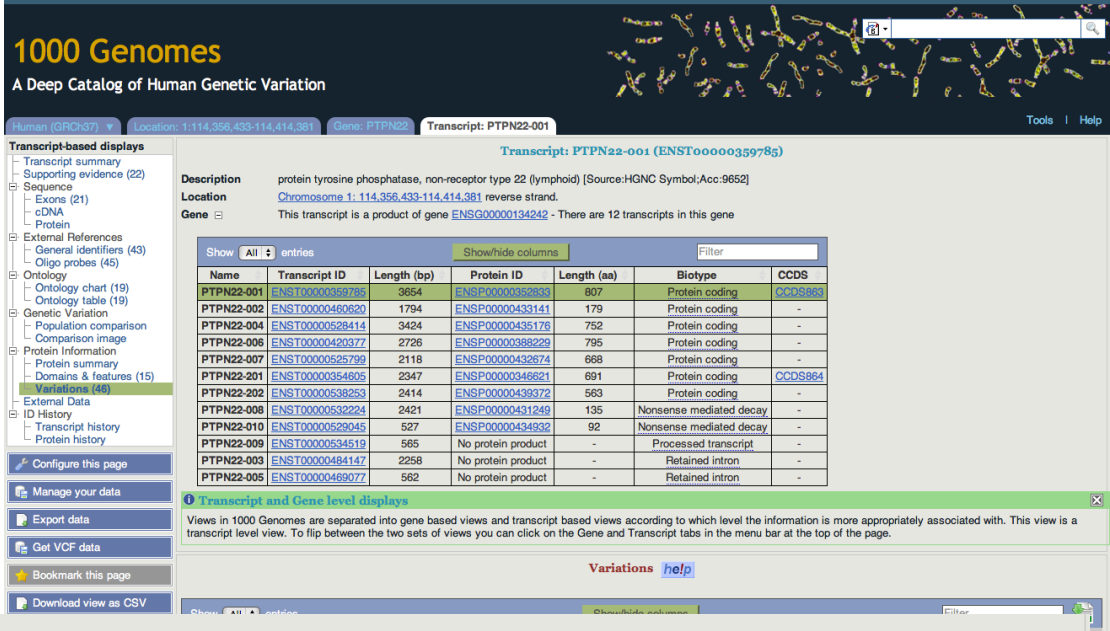
Variation Image

- Gene variation zoom

Transcript Tab: Variations

Effect on Protein:

- SIFT
- PolyPhen

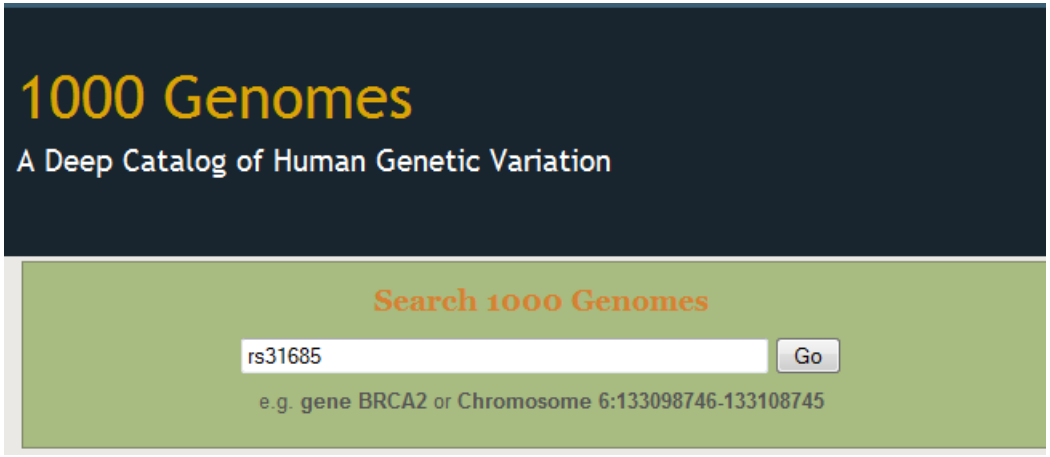


Download view as CSV

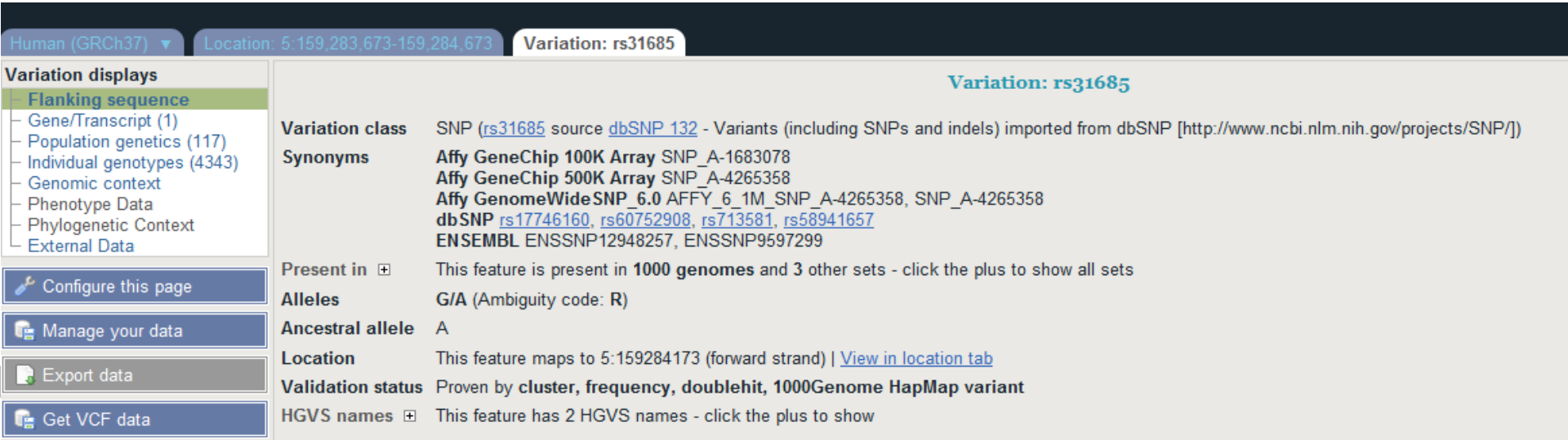
Show All entries | Show/hide columns | Filter

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
778	rs41313296	Non-synonymous coding	T/A	W	N, I	AAT, ATT	deleterious	probably damaging

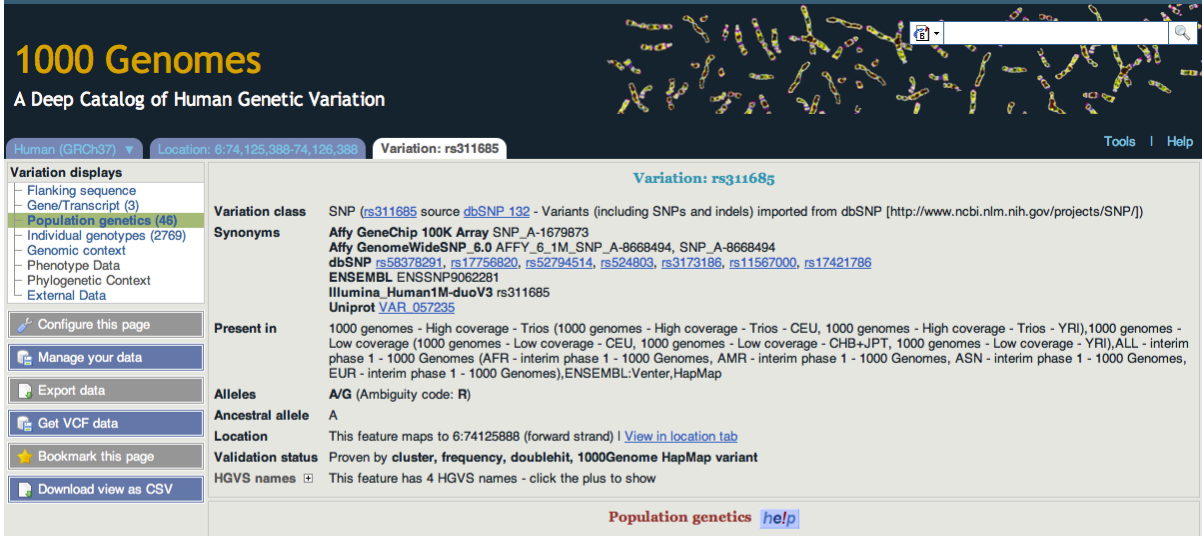
Start again- search for a variation (rs31685)



- The Variation tab- left hand links take you to more information



- Population



1000 genomes alleles frequencies

AFR	ALL	AMR	ASN	EUR
A: 69% G: 31%	A: 51% G: 49%	A: 54% G: 46%	A: 45% G: 55%	A: 42% G: 58%

1000 genomes

Show/hide columns | Filter | CSV

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 | CSV

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	

The Browser: Coming Soon

A THOUSAND GENOMES

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

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

Questions

Please send any questions about this presentation and any other material on our website to info@1000genomes.org

EMBL-EBI

Thanks

- The 1000 Genomes Project Consortium
- Paul Flicek
- Richard Smith
- Holly Zheng Bradley
- Ian Streeter
- Giulietta Spudich