1000 Genomes Browser Orientation and Pilot Data

http://browser.1000genomes.org

Based on Pilot Project Data

Paul Flicek

European Bioinformatics Institute

3 November 2010

Overview

- Based on version 54 of Ensembl code and NCBI36 version of the human genome assembly
- Contains all of the gene information normally present in Ensembl
  - Gene and transcript annotation, external references, sequence data
- Incorporates essentially all of the 1000 Genomes Pilot data
  - Some details and additional configuration options will be added over the coming weeks
- Please send questions to info@1000genomes.org

- 1000 Genomes full project data will be released in a new browser as the data becomes available

http://browser.1000genomes.org

1000 Genomes

A Deep Catalog of Human Genetic Variation

Home

Search 1000Genomes

Go

e.g. **gene BRCA2** or **AL032821.2.1.143563**

Start Browsing 1000 Genomes data

Browse Human →
NCBI 36

Transcript SNP view →
View the consequences of sequence variation at the level of each transcript in the genome.

SeqAlignView →
Shows read-depth data alongside SNPs

Other sites using Ensembl software...

Browser update November 2010

based on the full pilot project data described in A map of human genome variation from population-scale sequencing , Nature 467 , 1061.1073.

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 1000genomes data by the whole scientific community, this browser (based on Ensembl) integrates the SNP calls from the March 2010 release. All of 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 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.

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

1000 Genomes release 3 - June 2010 © EBI empowered

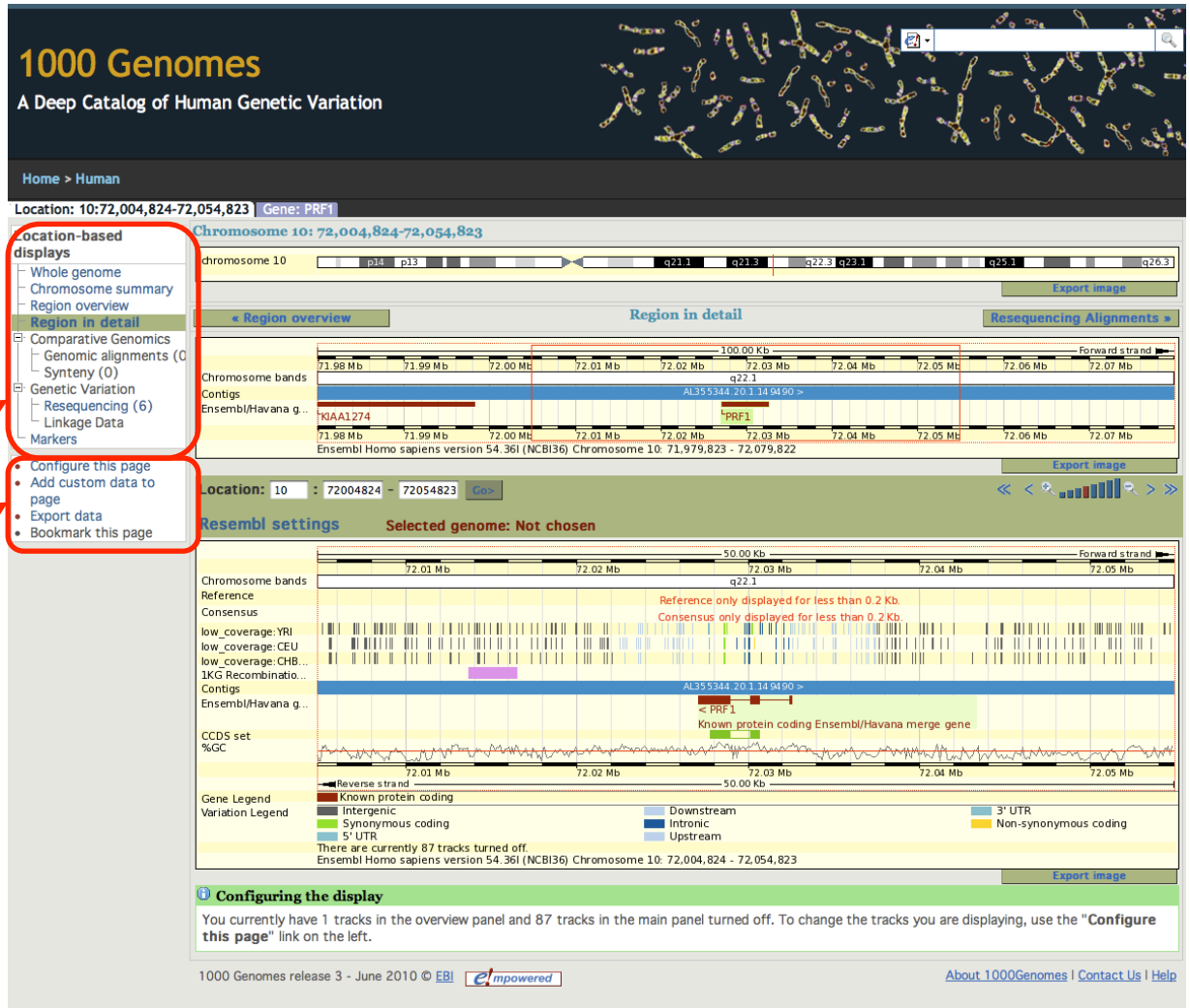
About 1000Genomes | Contact Us | Help

1000 Genomes Browser Home Page

http://browser.1000genomes.org

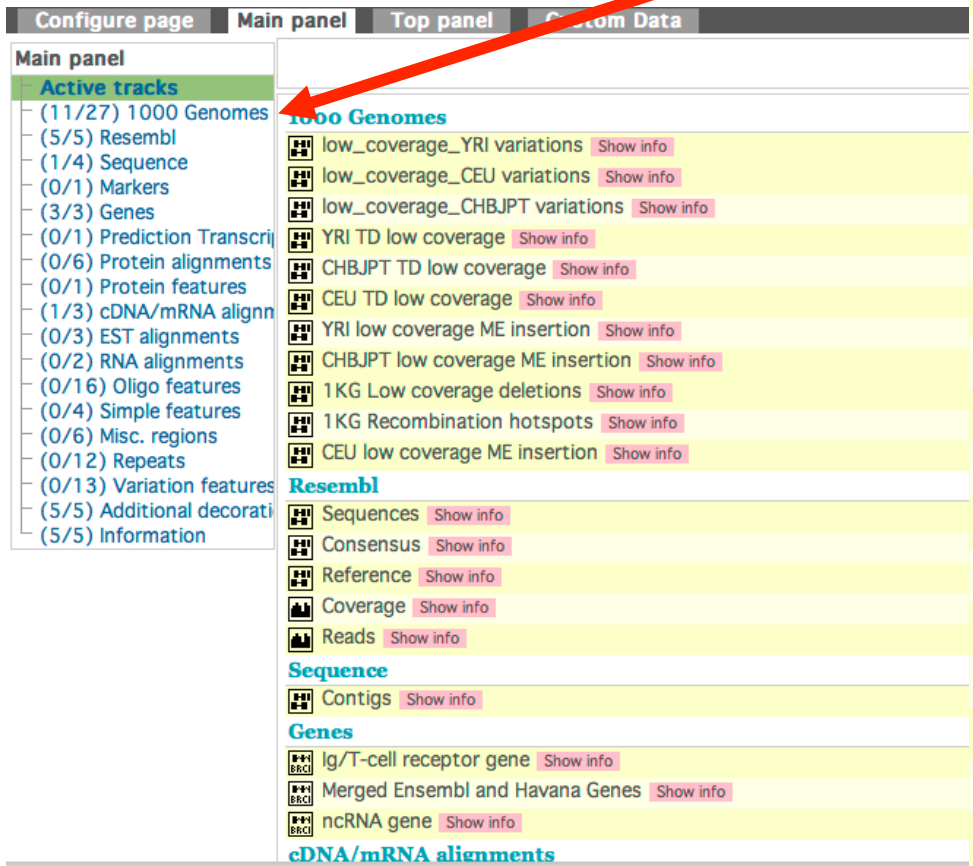
Main location view

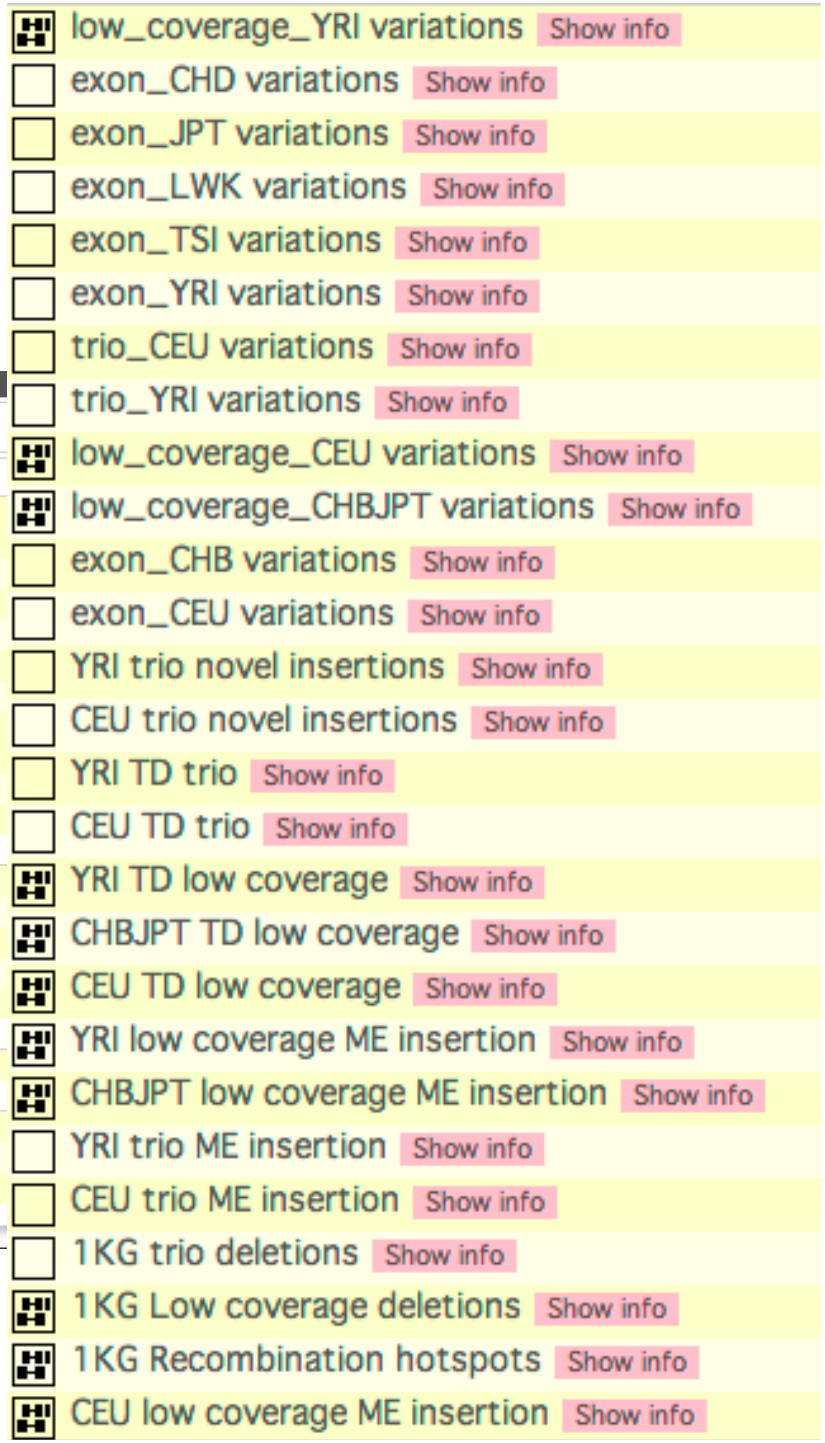
- Built on Ensembl
- Navigation is on the left hand side
- Page configuration and data expert links
- Includes only human data in current release



http://browser.1000genomes.org

Page configuration

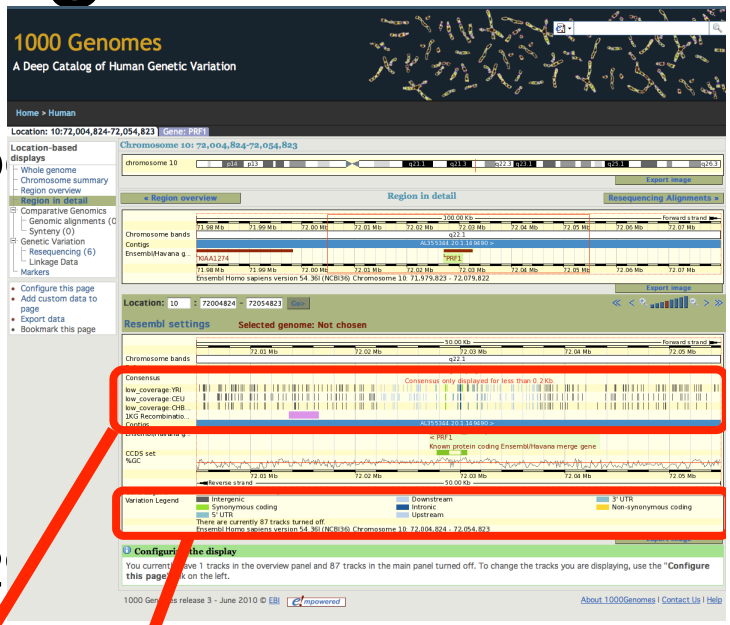


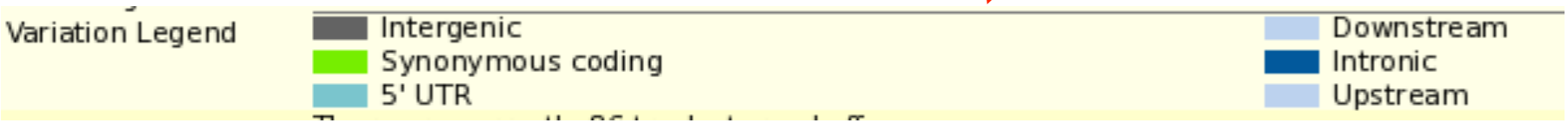


http://browser.1000genomes.org

1000 Genomes SNPs

- The high coverage trios, low coverage populations and exon capture data can b viewed in specific tracks on location pages
- These are selected from the “Configure this page” menu and appear as tracks near the middle of the display
- Tracks for all SNPs from dbSNP build 12 are also available





http://browser.1000genomes.org

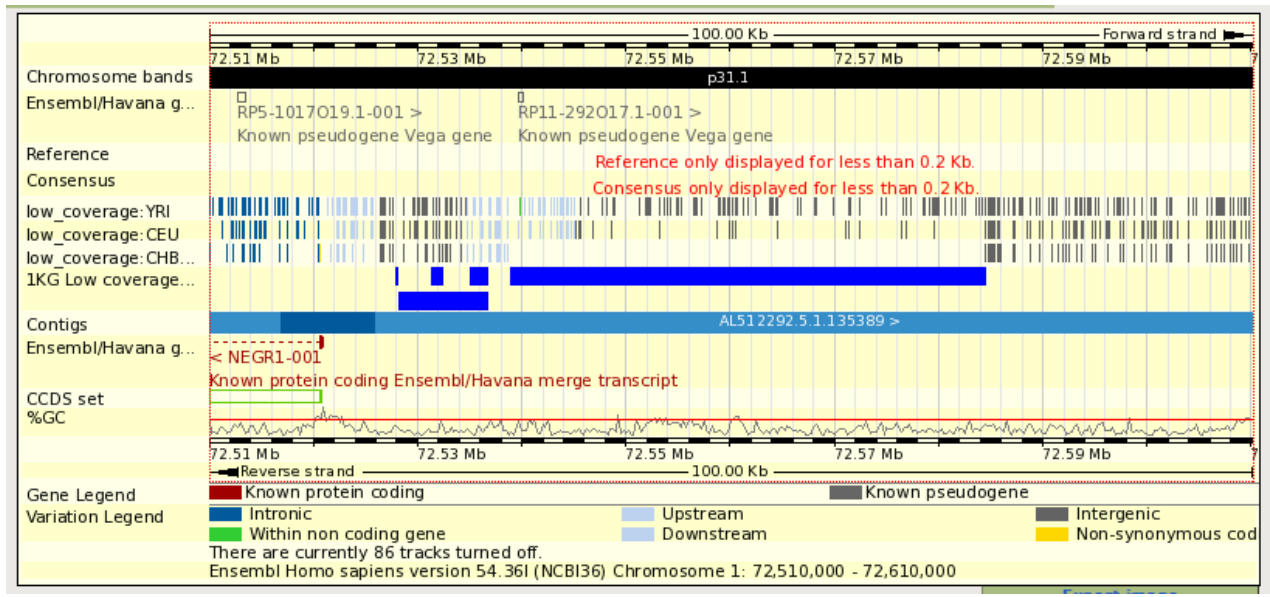
SNP Information

- SNPs are clickable which brings up a small window with basic information
- The “SNP properties” link leads to a dedicated page for the SNP with detailed information (including information imported from dbSNP)

http://browser.1000genomes.org

SV Information

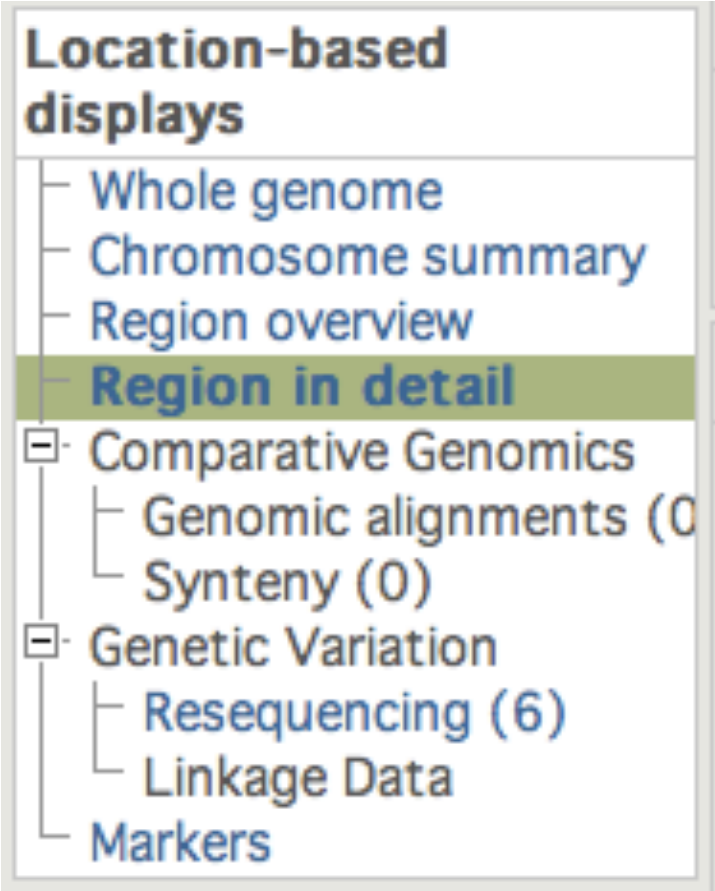
- Structural variants of specific types are selectable via the “configure this page” link
- Pop-up menus from clicking on the variant list whether the variant has been validated



http://browser.1000genomes.org

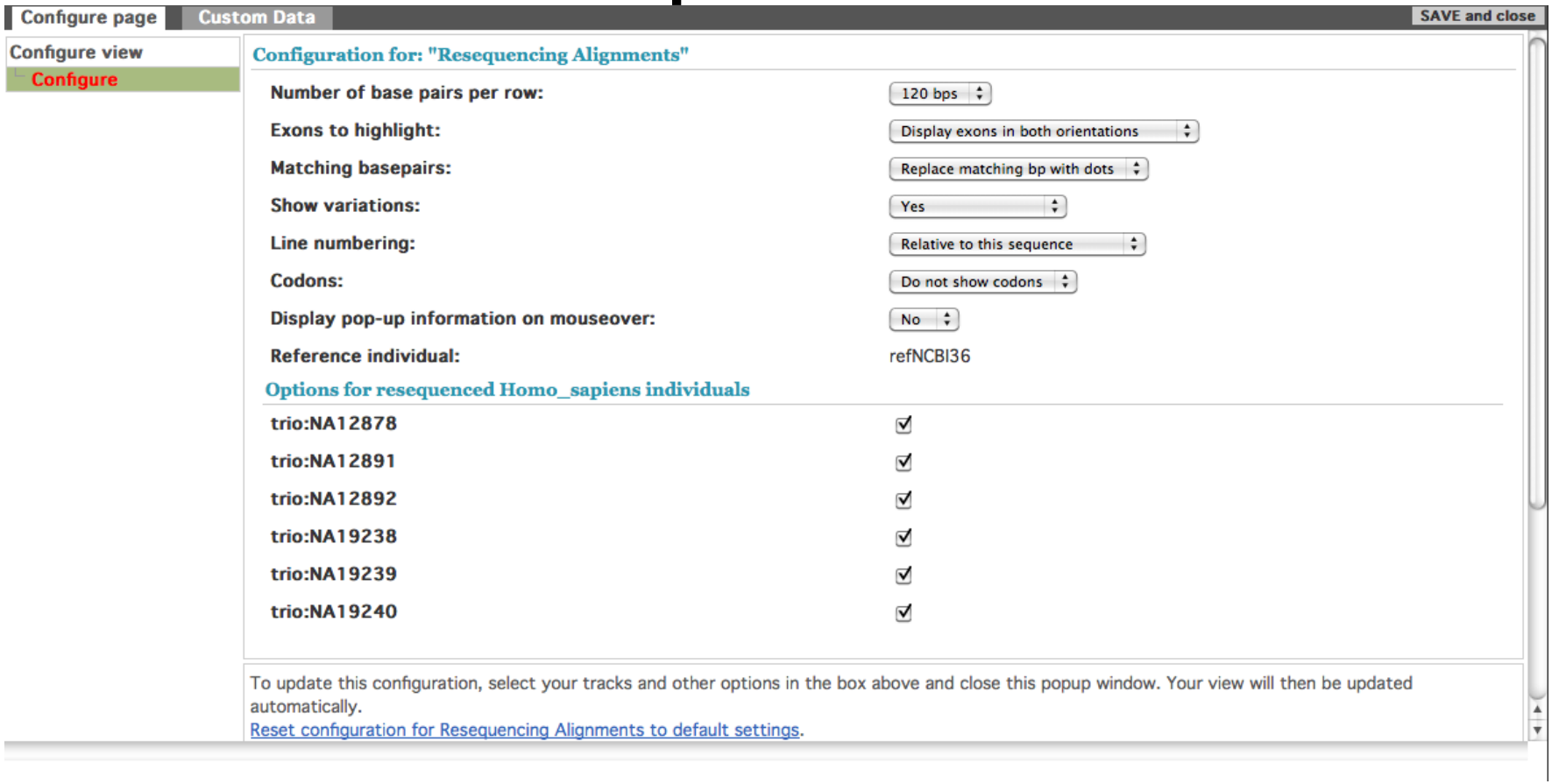
Resequencing alignment

- View any region of the genome in alignment with reference for the 6 high coverage trio individuals
- Assumption made that if there is sequence coverage and not a SNP called, the base is the same as the reference
  - Not a de novo assembly
- Use “Resequencing” link on the left side of pages to access view



http://browser.1000genomes.org

Resequencing alignment options



http://browser.1000genomes.org

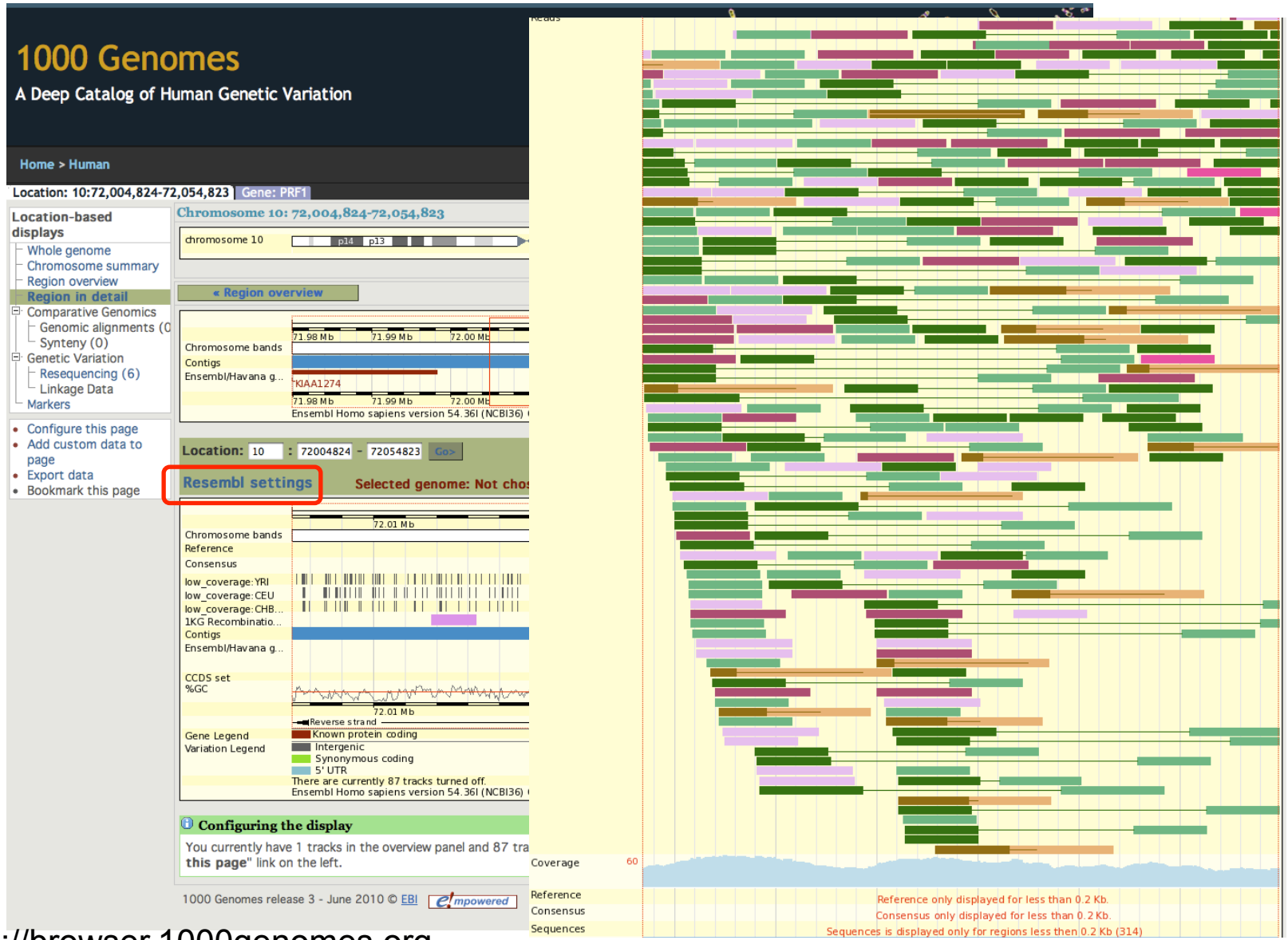
Resequencing alignment output

```
THIS STYLE: Location of selected exons
THIS STYLE: Location of SNPs
THIS STYLE: Resequencing coverage
·    Basepairs in secondary individuals matching the reference strain are replaced with dots
~    No resequencing coverage at this position
Homo_sapiens > chromosome:NCBI36:1:1084501:1130500:1

refNCBI36          1 TGGTGGCTGCAGTGAGCCGAGATCGCACCACTGCACTCCAGCCTGGGCAACAGAGCCAGACTCCATTTAAAAAAAGAAAAGCAGGTGAGGACGTGTGAGCAAGTCTGGGCCATGCTGCCA   120
trio:NA12878       1 ........................................................................................G...................   120
trio:NA12891       1 ........................................................................................R...................   120
trio:NA12892       1 ........................................................................................G...................   120
trio:NA19238       1 ........................................................................................G...................   120
trio:NA19239       1 ............................................................................................................   120
trio:NA19240       1 ........................................................................................R...................   120

refNCBI36        121 AGCTCCCGCTTCTCCGAGCAGCTCCACACAAGGAGCAGAGGCAGCTCCAGTTCCACAGCCAGACACAGTCATTTTCCCTCTACTCAGGAATTAGGCAGGATGGTGTGGGGCCTGAGTGTC   240
trio:NA12878     121 ........................................................................................................   240
trio:NA12891     121 ........................................................................................................   240
trio:NA12892     121 ........................................................................................................   240
trio:NA19238     121 ........................................................................................................   240
trio:NA19239     121 ........................................................................................................   240
trio:NA19240     121 ........................................................................................................   240

refNCBI36        241 ACCATGAAGCCGGGAAGCAGGTGTTCAGCCAGACGCAAATATTCCTCCTCAAAGCCAGGGGAGCGGCCACAGTGGATTTTATTATGGGGCGACGGGGGCACGCCCAGTGCCCCGCCACTC   360
trio:NA12878     241 ........................................................................................A...................   360
trio:NA12891     241 ........................................................................................R...................   360
trio:NA12892     241 ........................................................................................A...................   360
trio:NA19238     241 ........................................................................................A...................   360
trio:NA19239     241 ............................................................................................................   360
trio:NA19240     241 ........................................................................................R...................   360
```

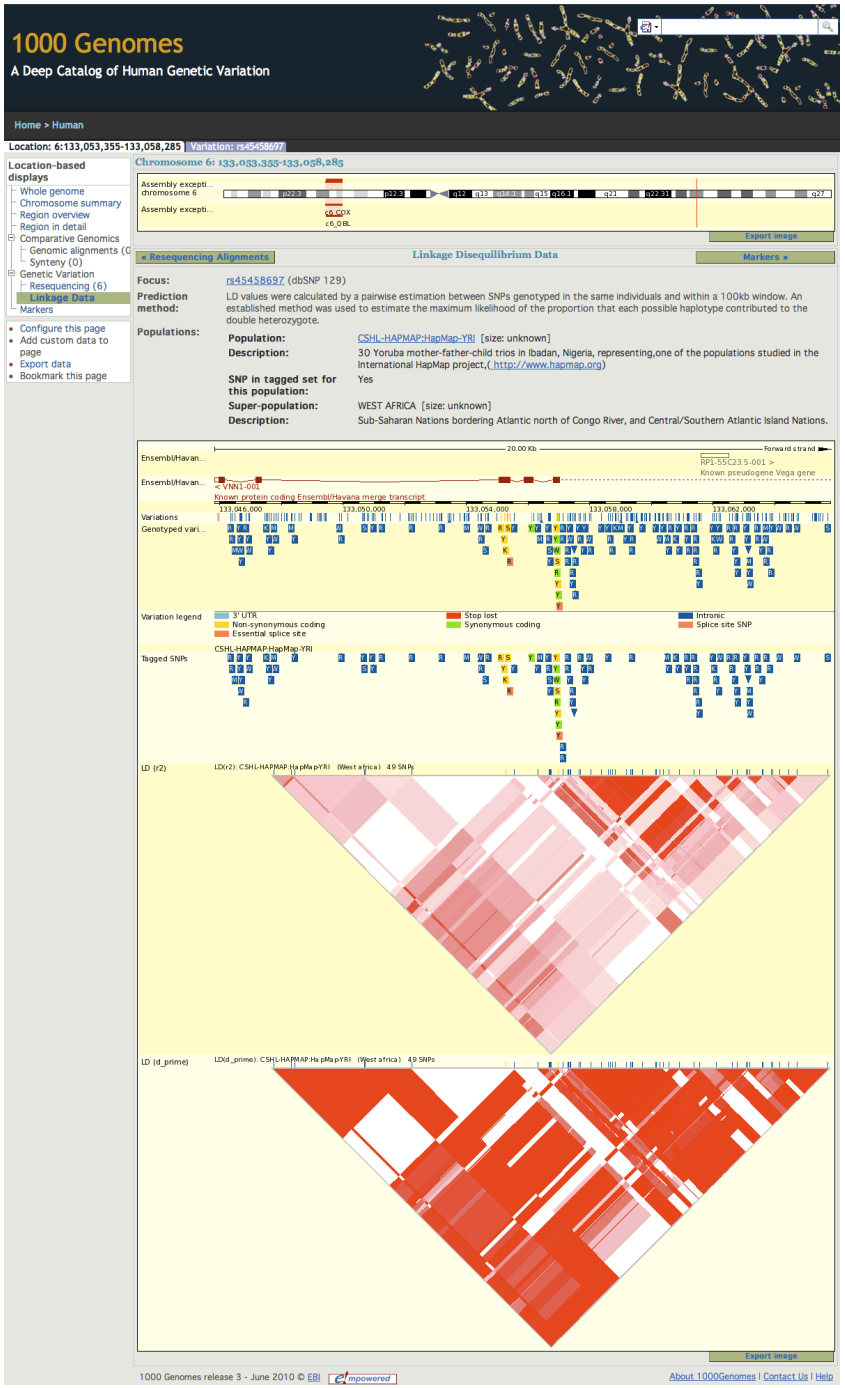
- SNPs are annotated
  - Heterozygous SNPs with ambiguity codes
- "~" means no coverage

http://browser.1000genomes.org

Sequence level data (trios only)



http://browser.1000genomes.org



http://browser.1000genomes.org

LD Information

- Currently based on data from HapMap and Perlegen populations
- Populations selectable from drop down tab

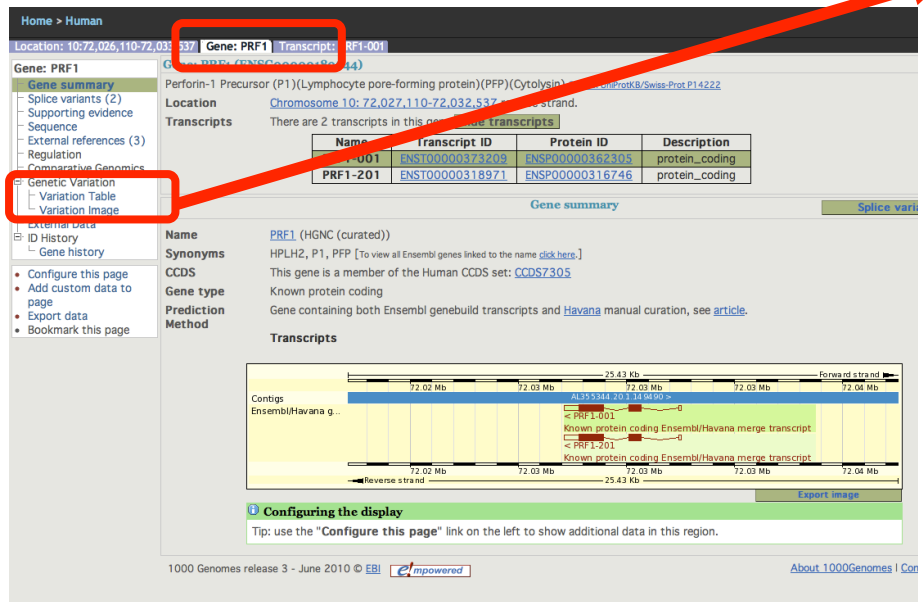
Data Export

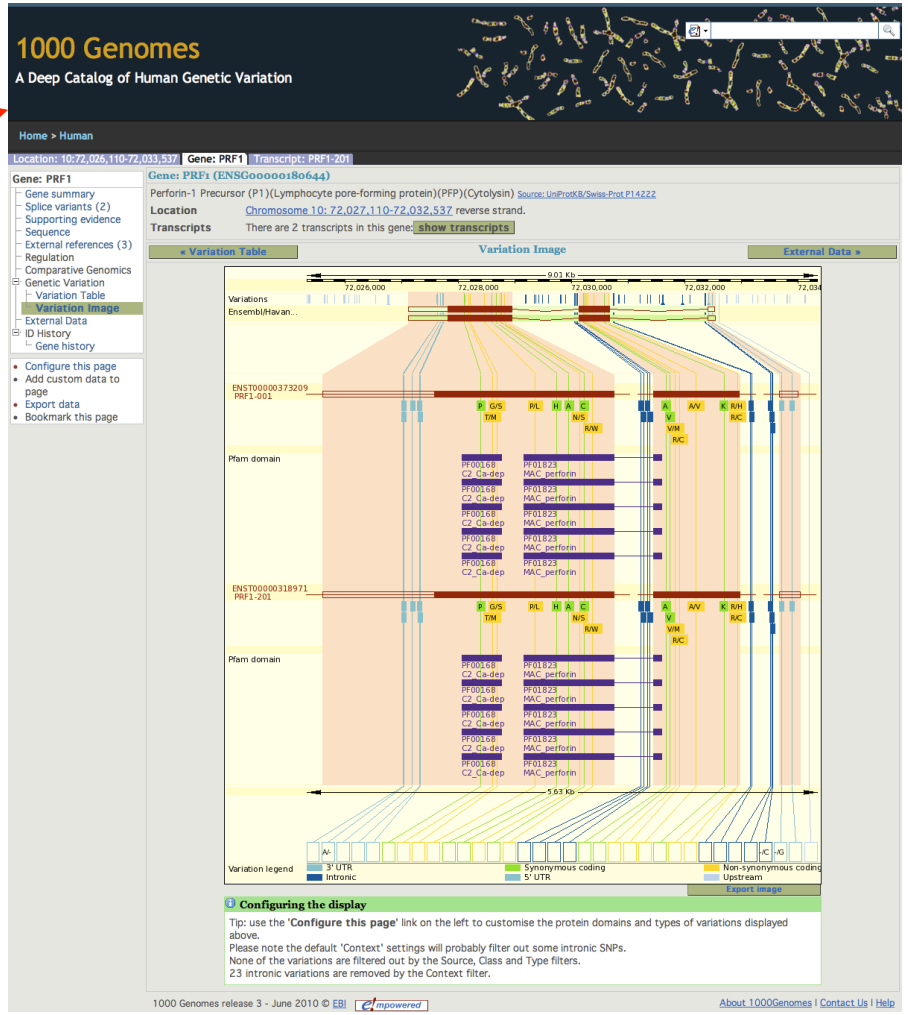
- Summary data from the region being viewed can be exported
- Additional export configuration options coming soon

http://browser.1000genomes.org

More variation displays

Gene Page

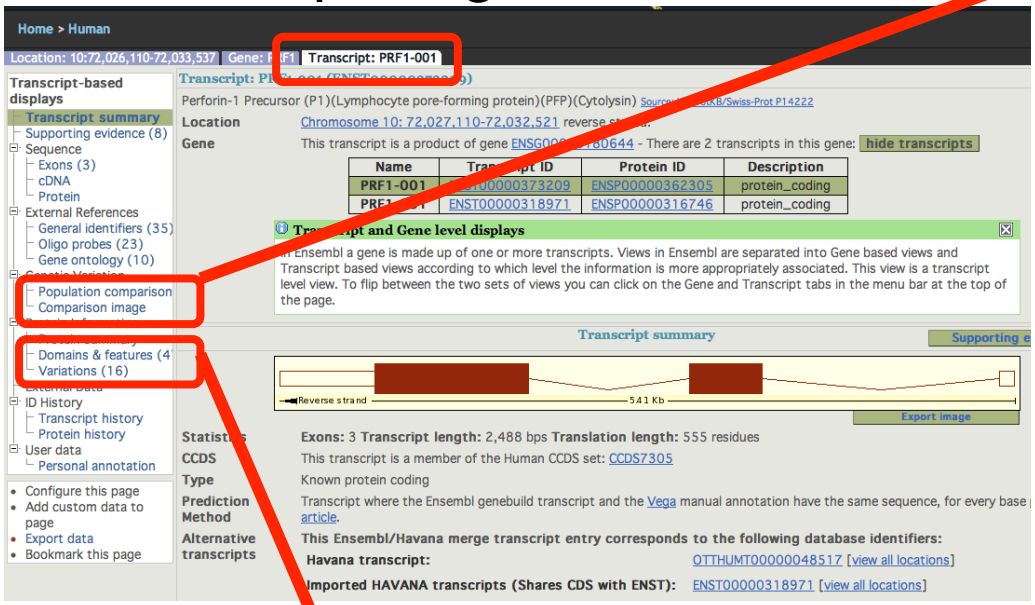


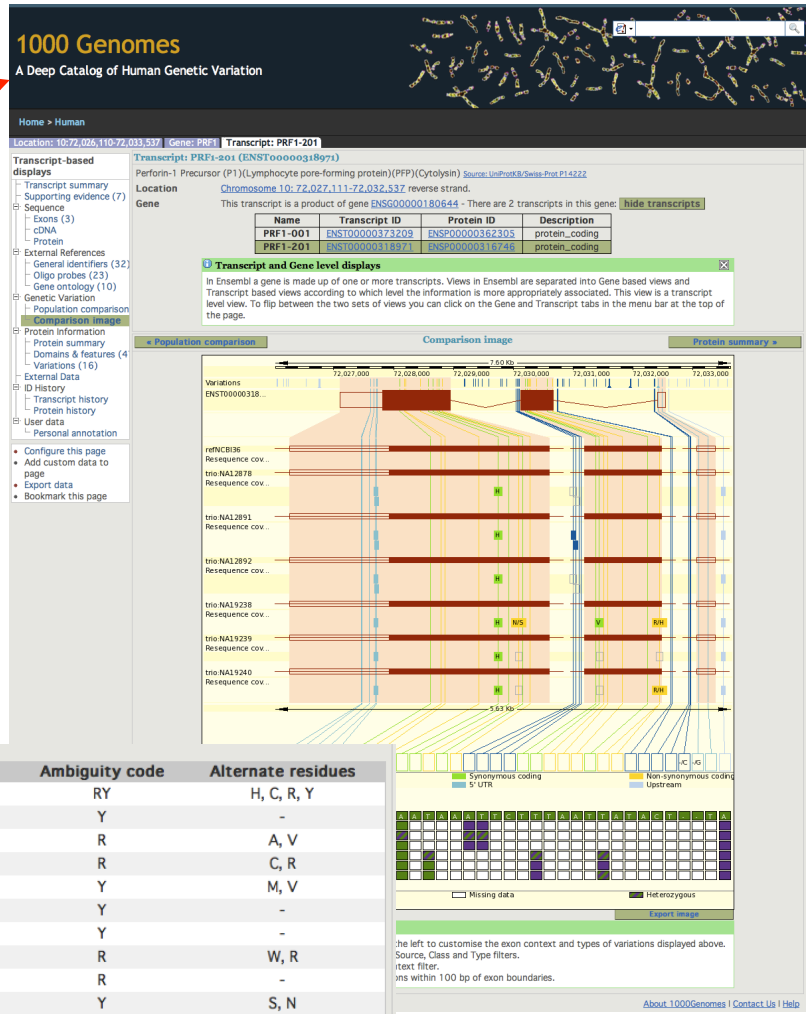


http://browser.1000genomes.org

More variation displays

Transcript Page





Residue	SNP ID	SNP type	Alleles	Ambiguity code	Alternate residues
4	rs12161733	Non-synonymous	A/G	RY	H, C, R, Y
32	rs2228018	Synonymous	T/C	Y	-
91	rs35947132	Non-synonymous	A/G	R	A, V
126	rs34279237	Non-synonymous	A/G	R	C, R
135	rs12263572	Non-synonymous	T/C	Y	M, V
145	trio:YRI:10:72030230	Synonymous	T/C	Y	-
154	low_coverage:YRI:10:72030203	Synonymous	T/C	Y	-
225	rs28933973	Non-synonymous	A/G	R	W, R
242	rs35329429	Synonymous	A/G	R	-
252	rs28933375	Non-synonymous	T/C	Y	S, N
274	rs885821	Synonymous	A/G	R	-
300	rs885822	Synonymous	A/G	R	-
345	rs28933374	Non-synonymous	A/G	R	P, L
426	rs1042652	Non-synonymous	T/C	Y	S, G
435	rs28933376	Non-synonymous	A/G	R	T, M
459	rs2228019	Synonymous	A/G	R	-

http://browser.1000genomes.org

Credits

- Eugene Kulesha, Glenn Proctor
- Natassa Spiridou
- Will McLaren, Fiona Cunningham
- Laura Clarke, Holly Zheng-Bradley, Rick Smith

- info@1000genomes.org