1000 Genomes Browser Quick start guide

http://browser.1000genomes.org

6 January 2008
Overview

• Based on version 50 ("old version") of Ensembl code
• Contains all of the gene information normally present in Ensembl
  – Gene and transcript annotation, external references, sequence data
• There are things that don’t work and we have not transferred to "production" web hardware
  – Please send questions, problems or apparent errors to flicek@ebi.ac.uk
1000 Genomes Browser Home Page
Main view

- Built on Ensembl
- Navigation is on the left hand side
- Options are drop down menus on the tops of the windows
- Includes only human data in current release
  - Comparative genomics information will be available in a future release
  - All appropriate pages have links to current versions of Ensembl and UCSC
Individual-specific SNPs

- The 1KG individuals can be viewed on the graphical view (contig view) pages
- These are selected from the “Features” menu and appear as tracks near the bottom of the display
- Tracks for all SNPs and the SNPs on selected Affy arrays are also available
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 (mostly imported from dbSNP) about population frequencies, identifiers, individual genotypes and other information.
SNPView

- Temporary rsName
- Identifiers
- Flanking sequence
- Population frequency
- Individual genotypes
- Change display options (individuals, SNP type, etc.)
- Location information
Resequencing alignment

- View any region of the genome in alignment with reference, 4 1KG individuals, Watson, Venter
- Assumption made that if there is sequence coverage and not a SNP called, the base is the same as the reference
- Use “Resequencing alignment” link on the left side of pages to access view
Resequencing alignment options

<table>
<thead>
<tr>
<th>Genomic Location and Markup options</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chromosome Name: 9</td>
</tr>
<tr>
<td>Start: 21944015</td>
</tr>
<tr>
<td>End: 21964014</td>
</tr>
<tr>
<td>Strand: Forward</td>
</tr>
<tr>
<td>Exons to highlight: None</td>
</tr>
<tr>
<td>Highlight variations: Yes</td>
</tr>
<tr>
<td>Line numbering: None</td>
</tr>
<tr>
<td>Alignment width: 60</td>
</tr>
<tr>
<td>Matching basepairs: Show all</td>
</tr>
<tr>
<td>Codons: Do not show codons</td>
</tr>
<tr>
<td>Title display: None</td>
</tr>
</tbody>
</table>

Reference individual: NCBI36
Resequenced Human individuals:
- 1KG NA12878
- 1KG NA12891
- 1KG NA12892
- 1KG NA19240
- Venter
- Watson

Deselect all individuals
Select all individuals
Update

Fields marked with * are required
### Resequencing alignment output

<table>
<thead>
<tr>
<th>Marked up sequence</th>
<th>~ No resequencing coverage at this position</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Homo_sapiens</strong></td>
<td>chromosome:NCBI36::219944015:21964014:1</td>
</tr>
<tr>
<td>NCB136</td>
<td>AAAACCTGATCTGCTTACATATTGACAAAGACATCTACAAACCTCAGAGATGACGA</td>
</tr>
<tr>
<td>JKG_NA12873</td>
<td>AAAACCTGATCTGCTTACATATTGACAAAGACATCTACAAACCTCAGAGATGACGA</td>
</tr>
<tr>
<td>JKG_NA12891</td>
<td>AAAACCTGATCTGCTTACATATTGACAAAGACATCTACAAACCTCAGAGATGACGA</td>
</tr>
<tr>
<td>JKG_NA12922</td>
<td>AAAACCTGATCTGCTTACATATTGACAAAGACATCTACAAACCTCAGAGATGACGA</td>
</tr>
<tr>
<td>Venetor</td>
<td>AAAACCTGATCTGCTTACATATTGACAAAGACATCTACAAACCTCAGAGATGACGA</td>
</tr>
<tr>
<td>Watson</td>
<td>AAAACCTGATCTGCTTACATATTGACAAAGACATCTACAAACCTCAGAGATGACGA</td>
</tr>
<tr>
<td>NCB136</td>
<td>TTTAAAAAATACCTGGACCTTCTAGACTATAGGAGTTCCAAAGGACGGAGACAC</td>
</tr>
<tr>
<td>JKG_NA12873</td>
<td>TTTAAAAAATACCTGGACCTTCTAGACTATAGGAGTTCCAAAGGACGGAGACAC</td>
</tr>
<tr>
<td>JKG_NA12891</td>
<td>TTTAAAAAATACCTGGACCTTCTAGACTATAGGAGTTCCAAAGGACGGAGACAC</td>
</tr>
<tr>
<td>JKG_NA12922</td>
<td>TTTAAAAAATACCTGGACCTTCTAGACTATAGGAGTTCCAAAGGACGGAGACAC</td>
</tr>
<tr>
<td>Venetor</td>
<td>TTTAAAAAATACCTGGACCTTCTAGACTATAGGAGTTCCAAAGGACGGAGACAC</td>
</tr>
<tr>
<td>Watson</td>
<td>TTTAAAAAATACCTGGACCTTCTAGACTATAGGAGTTCCAAAGGACGGAGACAC</td>
</tr>
<tr>
<td>NCB136</td>
<td>GGAAGGCTGGTTCTACGAGGAGAACAGCAGACTATAGGAGTTCCAAAGGACGGAGACAC</td>
</tr>
<tr>
<td>JKG_NA12873</td>
<td>GGAAGGCTGGTTCTACGAGGAGAACAGCAGACTATAGGAGTTCCAAAGGACGGAGACAC</td>
</tr>
<tr>
<td>JKG_NA12891</td>
<td>GGAAGGCTGGTTCTACGAGGAGAACAGCAGACTATAGGAGTTCCAAAGGACGGAGACAC</td>
</tr>
<tr>
<td>JKG_NA12922</td>
<td>GGAAGGCTGGTTCTACGAGGAGAACAGCAGACTATAGGAGTTCCAAAGGACGGAGACAC</td>
</tr>
<tr>
<td>Venetor</td>
<td>GGAAGGCTGGTTCTACGAGGAGAACAGCAGACTATAGGAGTTCCAAAGGACGGAGACAC</td>
</tr>
<tr>
<td>Watson</td>
<td>GGAAGGCTGGTTCTACGAGGAGAACAGCAGACTATAGGAGTTCCAAAGGACGGAGACAC</td>
</tr>
<tr>
<td>NCB136</td>
<td>ATGCCTGAGTGGTCTTGGTGGCGAAGGAGAATACCAGCGAAGCTGCTTTGGGCTAG</td>
</tr>
<tr>
<td>JKG_NA12873</td>
<td>ATGCCTGAGTGGTCTTGGTGGCGAAGGAGAATACCAGCGAAGCTGCTTTGGGCTAG</td>
</tr>
<tr>
<td>JKG_NA12891</td>
<td>ATGCCTGAGTGGTCTTGGTGGCGAAGGAGAATACCAGCGAAGCTGCTTTGGGCTAG</td>
</tr>
<tr>
<td>JKG_NA12922</td>
<td>ATGCCTGAGTGGTCTTGGTGGCGAAGGAGAATACCAGCGAAGCTGCTTTGGGCTAG</td>
</tr>
<tr>
<td>Venetor</td>
<td>ATGCCTGAGTGGTCTTGGTGGCGAAGGAGAATACCAGCGAAGCTGCTTTGGGCTAG</td>
</tr>
<tr>
<td>Watson</td>
<td>ATGCCTGAGTGGTCTTGGTGGCGAAGGAGAATACCAGCGAAGCTGCTTTGGGCTAG</td>
</tr>
</tbody>
</table>

- Hets
- SNPs
- No coverage
Individual SNP consequences

- TranscriptSNPView
  - Linked off all Ensembl Genes
  - From transcript pages in left hand menu
  - Color-coded display of how SNPs affect transcripts
- Individual
- Coverage
- SNPs
- Legend
- SNP summary
  - Green
    - same as reference
  - Purple:
    - different
  - Hatched
    - heterozygous
LDView

- Based on data from HapMap and Perlegen populations
- Populations selectable from drop down tab
More SNPs displays

GeneSeqView

GeneView

GeneSNPView

TransView

ProtView

Genomic sequence

Gene variation info.

Protein information

Transcript information
SNP types

Non-synonymous  In coding sequence, resulting in an aa change
Synonymous     In coding sequence, not resulting in an aa change
Frameshift     In coding sequence, resulting in a frameshift
Stop lost      In coding sequence, resulting in the loss of a stop codon
Stop gained    In coding sequence, resulting in the gain of a stop codon

Essential splice site In the first 2 or the last 2 basepairs of an intron
Splice site     1-3 bps into an exon or 3-8 bps into an intron
Upstream       Within 5 kb upstream of the 5'-end of a transcript
Regulatory region In regulatory region annotated by Ensembl
5' UTR          In 5' UTR
Intronic        In intron
3' UTR          In 3' UTR
Downstream     Within 5 kb downstream of the 3'-end of a transcript
Intergenic     More than 5 kb away from a transcript
Credits

• Eugene Kulesha, Stephen Keenan
• Yuan Chen, Fiona Cunningham
• Laura Clarke, Zam Iqbal

• 1000 Genome Data providers
• Entire Ensembl Team