Summary Part 2

• Location => Look out for Feature
• Object creation from Adaptor or from API object
• No allele strings for structural variations, only coordinates
Variation consequences

- Define consequence of a variation in relation to the transcript structure it overlaps
- Sequence ontology provides controlled vocabulary to describe consequences

7660 Variations for SRGAP2
Variation consequences

Regulatory region

Coding region
missense
stop gained

Splice site region
Variation consequences

http://www.ensembl.org/info/docs/variations/predicted_data.html
<table>
<thead>
<tr>
<th>Variant Type</th>
<th>Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Missense variant</td>
<td>A sequence variant, that changes one or more bases, resulting in a different amino acid sequence but where the length is preserved (SO:0001583)</td>
<td></td>
</tr>
<tr>
<td>Splice region variant</td>
<td>A sequence variant in which a change has occurred within the region of the splice site, either within 1-3 bases of the exon or 3-8 bases of the intron (SO:0001630)</td>
<td></td>
</tr>
<tr>
<td>Incomplete terminal codon variant</td>
<td>A sequence variant where at least one base of the final codon of an incompletely annotated transcript is changed (SO:0001626)</td>
<td></td>
</tr>
<tr>
<td>Synonymous variant</td>
<td>A sequence variant where there is no resulting change to the encoded amino acid (SO:0001619)</td>
<td></td>
</tr>
<tr>
<td>Stop retained variant</td>
<td>A sequence variant where at least one base in the terminator codon is changed, but the terminator remains (SO:0001567)</td>
<td></td>
</tr>
<tr>
<td>Coding sequence variant</td>
<td>A sequence variant that changes the coding sequence (SO:0001580)</td>
<td></td>
</tr>
<tr>
<td>Mature miRNA variant</td>
<td>A transcript variant located with the sequence of the mature miRNA (SO:0001620)</td>
<td></td>
</tr>
<tr>
<td>5 prime UTR variant</td>
<td>A UTR variant of the 5' UTR (SO:0001623)</td>
<td></td>
</tr>
<tr>
<td>3 prime UTR variant</td>
<td>A UTR variant of the 3' UTR (SO:0001624)</td>
<td></td>
</tr>
<tr>
<td>Intron variant</td>
<td>A transcript variant occurring within an intron (SO:0001627)</td>
<td></td>
</tr>
<tr>
<td>NMD transcript variant</td>
<td>A variant in a transcript that is the target of NMD (SO:0001621)</td>
<td></td>
</tr>
<tr>
<td>Non coding exon variant</td>
<td>A sequence variant that changes non-coding exon sequence (SO:0001792)</td>
<td></td>
</tr>
<tr>
<td>NC transcript variant</td>
<td>A transcript variant of a non coding RNA (SO:0001619)</td>
<td></td>
</tr>
<tr>
<td>Upstream gene variant</td>
<td>A sequence variant located 5' of a gene (SO:0001631)</td>
<td></td>
</tr>
<tr>
<td>Downstream gene variant</td>
<td>A sequence variant located 3' of a gene (SO:0001632)</td>
<td></td>
</tr>
<tr>
<td>ALL</td>
<td>All variations (WARNING: table may not load for this number of variants!)</td>
<td></td>
</tr>
</tbody>
</table>
TranscriptVariation TV

- Can be retrieved from VariationFeature or TranscriptVariationAdaptor
- TranscriptVariation: VariationFeature overlapping or near a Transcript
- VariationFeature can overlap more than one transcript (alternative splicing)
- Most ‘severe’ consequence is stored in VariationFeature
rs2289361, 1:206566172, G/A

ENST00000295713

get_all_TranscriptVariations

transcript
cdna_coords
cds_coords
translation_coords

<table>
<thead>
<tr>
<th>Transcript (strand)</th>
<th>Allele (transcript allele)</th>
<th>Type</th>
<th>Position in transcript</th>
<th>Position in CDS</th>
<th>Position in protein</th>
<th>Amino acid</th>
<th>Codons</th>
<th>SIFT</th>
<th>PolyPhen</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENST00000295713 (+)</td>
<td>Missense variant</td>
<td>353</td>
<td>355</td>
<td>119</td>
<td>R/C</td>
<td>CGC/TGC</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
# TranscriptVariation TV

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Return type</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transcript</td>
<td>Transcript object</td>
<td>$tv-&gt;transcript()</td>
</tr>
<tr>
<td>Consequence type</td>
<td>String Listref of string</td>
<td>$tv-&gt;display_consequence() *most severe $tv-&gt;consequence_type() *all</td>
</tr>
<tr>
<td>Coordinates</td>
<td>int</td>
<td>$tv-&gt;cdna_start(), $tv-&gt;cdna_end(), $tv-&gt;cds_start(), $tv-&gt;translation_start()</td>
</tr>
<tr>
<td>Amino acid</td>
<td>string (F/L)</td>
<td>$tv-&gt;pepallele_string()</td>
</tr>
</tbody>
</table>
Example Variation consequence

```perl
my $va = $registry->get_adaptor('human', 'variation', 'variation');
my $vfa = $registry->get_adaptor('human', 'variation', 'variationfeature');

my $variation = $va->fetch_by_name('rs201168060');
my $vfs = $vfa->fetch_all_by_Variation($variation);

print 'Variation rs201168060 has ', scalar @$vfs, ' VariationFeature(s)', "\n";

my $vf = $vfs->[0];
print $vf->variation_name, ' Most severe consequence: ', $vf->display_consequence(), "\n";
print 'All consequences: ', join('', @{$vf->consequence_type()}), "\n\n";

my $tvs = $vf->get_all_TranscriptVariations();

print 'Variation has ', scalar @$tvs, ' TranscriptVariations. (Overlaps that many Transcripts)', "\n";
foreach my $tv (@$tvs) {
    print $tv->transcript->stable_id, ' Most severe consequence: ',
    $tv->display_consequence(), "\n";
    print 'All consequences: ', join('', @{$tv->consequence_type()}), "\n";
}
```
Variation rs201168060 has 1 VariationFeature(s).
rs201168060 Most severe consequence: missense_variant
All consequences: missense_variant, 5_prime_UTR_variant, splice_region_variant, upstream_gene_variant

Variation has 5 TranscriptVariations. (Overlaps that many Transcripts)
ENST00000419187 Most severe consequence: splice_region_variant
All consequences: splice_region_variant, 5_prime_UTR_variant

ENST00000414007 Most severe consequence: missense_variant
All consequences: missense_variant, splice_region_variant

ENST00000450872 Most severe consequence: upstream_gene_variant
All consequences: upstream_gene_variant

ENST00000295713 Most severe consequence: missense_variant
All consequences: missense_variant, splice_region_variant

ENST00000414359 Most severe consequence: missense_variant
All consequences: missense_variant, splice_region_variant
Exercise 4a – Variation consequence

- Fetch all transcript variations (germline and somatic) in transcript ENST00000001008 in human and retrieve the following information:
  - Variation name
  - Consequence type (most severe)
  - Amino acid change*
  - Position in cDNA* and position in translation*
- In a second attempt filter for transcript variations of consequence type: ‘missense_variant’

**HINT:** fetch_all_by_Transcripts method requires a listref of objects; so use e.g. [$transcript] instead of $transcript

* if information exists