Outline

a. Introduction
b. Data objects & object adaptors
c. Ensembl documentation
d. The Registry & Ensembl API script design
e. Coordinate systems & slices
f. Features
g. Genes, transcripts, exons & translations
h. External references
Genes, transcripts, exons & translations

- DNA
  - Promoter
  - Enhancers
  - Transcription

- Pre-mRNA
  - Exon
  - Exon
  - Exon
  - Introns
  - Splicing

- mRNA
  - 5' UTR
  - Open reading frame
  - 3' UTR
  - Translation

- Protein
Genes, Transcript and Exons are objects that can be used just like any other Feature object.

A Gene is a set of alternatively spliced Transcripts.

A Transcript is a set of Exons.

Introns are not explicitly defined in the database.

[Diagram of gene, transcript, and exon structures]
# helper function: returns location and stable_id string for a feature
sub get_string {
    my $feature = shift;
    my $stable_id = $feature->stable_id;
    my $seq_region = $feature->slice->seq_region_name;
    my $start = $feature->start;
    my $end = $feature->end;
    my $strand = $feature->strand;
    return "$stable_id $seq_region:$start-$end($strand)";
}

# fetch a gene by its stable identifier
my $gene = $gene_adaptor->fetch_by_stable_id('ENSG00000123427');

# print out the gene, its transcripts, and its exons
print "Gene: ", get_string($gene), "\n";
foreach my $transcript (@{$gene->get_all_Transcripts}) {
    print "  Transcript: ", get_string($transcript), "\n";
    foreach my $exon (@{$transcript->get_all_Exons}) {
        print "    Exon: ", get_string($exon), "\n";
    }
}
OUTPUT:
Gene: ENSG00000123427  12:58165275-58176324(1)
   Transcript: ENST00000548256  12:58165275-58174213(1)
      Exon: ENSE000002360002  12:58165275-58165408(1)
      Exon: ENSE000003090890  12:58166800-58166911(1)
      Exon: ENSE000003030714  12:58168412-58168550(1)
      Exon: ENSE000002406112  12:58174038-58174213(1)
   Transcript: ENST00000551420  12:58165983-58174563(1)
      Exon: ENSE000002355737  12:58165983-58166180(1)
      Exon: ENSE000003223131  12:58166800-58166911(1)
      Exon: ENSE000002376752  12:58174038-58174563(1)
   Transcript: ENST00000300209  12:58166383-58176186(1)
      Exon: ENSE000002301479  12:58166383-58166684(1)
      Exon: ENSE000003090890  12:58166800-58166911(1)
      Exon: ENSE000002393444  12:58174038-58176186(1)
   etc.
Translations

Translations are not Features.

A Translation object defines the UTR and CDS of a Transcript.

Peptides are not stored in the database, they are computed on the fly using Transcript objects.

Not all transcripts have a translation (e.g. ncRNAs)
Translations – Code Example

my $transcript_adaptor = $registry->get_adaptor( 'Homo sapiens', 'Core', 'Transcript' );

# fetch a transcript from the database
my $transcript =
    $transcript_adaptor->fetch_by_stable_id('ENST00000333012');

# obtain the translation of the transcript
my $translation = $transcript->translation;

# print out the translation info
print "Translation: ", $translation->stable_id, "\n";
print "Start Exon: ",$translation->start_Exon->stable_id,"\n";
print "End Exon: " , $translation->end_Exon->stable_id, "\n";

# cDNA start and end (spliced sequence with UTR)
print "Start : ", $translation->cdna_start, "\n";
print "End : ", $translation->cdna_end, "\n";

# print the peptide which is the product of the translation
print "Peptide : ", $transcript->translate->seq, "\n";
OUTPUT:

Translation: ENSP00000327425
Start Exon: ENSE00002340145
End Exon: ENSE00002428887
Start : 48
End : 497
Peptide:
  MADPGPDPESESESVFPREVGLFADSYSEKSQFCFCGHLTVITQNFSGRLGVAARVDAAALSLCNYSQNVDFR
  GKKVIELGAGTGIVGILAALQGAYGLVRETEDDVIEQELWRGMRGACGHALSMSTMTPWESIKGSSVRGGCYHH
(a) Fetch gene ‘CSNK2A1’ and print the number of its transcripts and exons.
✓ hint: use GeneAdaptor method fetch_by_display_label; remember that not all transcripts have a translation

(a) For the above gene, get all the transcripts and list the number of exons in each and the translations.

(a) Why do the exon numbers not match?