

Module 3

Genome Browsing

Using Web Browsers to View
Genome Annotation

Kerstin Howe
Wellcome Trust Sanger Institute
zfish-help@sanger.ac.uk



Introduction

- Genome browsing
- The Ensembl gene set
- Guided examples
- Make your own data visible: BED files

Genome Browsers

- Ensembl Genome Browser <http://www.ensembl.org/>
- NCBI Map Viewer <http://www.ncbi.nlm.nih.gov/mapview/>
- UCSC Genome Browser <http://genome.ucsc.edu/>

NCBI Map Viewer

PubMed
Entrez
BLAST
OMIM
Taxonomy
Structure

Search

[Danio rerio \(zebrafish\) Zv9 \(Current\)](#) [BLAST zebrafish sequences](#)

Chromosome: [1] 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 MT

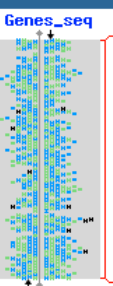
Master Map: Genes On Sequence [Summary of Maps](#) [Maps & Options](#)

Region Displayed: 0-60M bp [Download/View Sequence/Evidence](#)

Contig	Dr	UniG	Genes_seq	Symbol	O	Links	E	Description
NW_003039	Dr. 75999			grk1a	+	ZFIN ug sv pr dl ev hm sts		best RefSeq G protein-coupled receptor kinase 1 a
NW_003334	Dr. 1310			zgc:154001	+	ZFIN ug sv pr dl ev hm sts		best RefSeq
NW_003039	Dr. 75698			mbnl2	+	ZFIN ug sv pr dl ev hm sts		best RefSeq muscleblind-like 2 (Drosophila)
NW_003334	Dr. 75307			zgc:162208	+	ZFIN ug sv pr dl ev hm		best RefSeq AIG2-like domain-containing protein 1-B
NW_001878	Dr. 24233			tuba8l2	+	ZFIN ug sv pr dl ev hm sts		best RefSeq tubulin, alpha 8 like 2
NW_001878	Dr. 31421			zgc:165621	+	ZFIN ug sv pr dl ev hm sts		best RefSeq
NW_001878	Dr. 35143			zgc:103678	+	ZFIN ug sv pr dl ev hm		best RefSeq
NW_001878	Dr. 85105			crmp1	+	ZFIN ug sv pr dl ev hm		best RefSeq collapsin response mediator protein 1
NW_001878	Dr. 33271			pax5	+	ZFIN ug sv pr dl ev hm sts		best RefSeq paired box gene 5
NW_001878	Dr. 65130			zgc:77713	+	ZFIN ug sv pr dl ev hm sts		best RefSeq
NW_003039	Dr. 83488			ube2k	+	ZFIN ug sv pr dl ev hm sts		best RefSeq ubiquitin-conjugating enzyme E2K (UBC1 homolog, yeast)
NW_003334	Dr. 29079			plr1	+	ZFIN ug sv pr dl ev hm sts		best RefSeq pleiotropic regulator 1
NW_001878	Dr. 907			as3mt	+	ZFIN ug sv pr dl ev hm sts		best RefSeq arsenic (+3 oxidation state) methyltransferase
NW_001878	Dr. 113580			naa15b	+	ZFIN ug sv pr dl ev hm sts		best RefSeq N(alpha)-acetyltransferase 15, NatA auxiliary subunit b
NW_001878	Dr. 77175			itsn1	+	ZFIN ug sv pr dl ev hm sts		best RefSeq intersectin 1 (SH3 domain protein)
NW_001878	Dr. 75723			col17a1a	+	ZFIN ug sv pr dl ev		best RefSeq collagen, type XVII, alpha 1a
NW_003039	Dr. 75606			acy3.2	+	ZFIN ug sv pr dl ev hm		best RefSeq aspartoacylase (aminocyclase) 3.2
NW_001878	Dr. 47214			golga7ba	+	ZFIN ug sv pr dl ev		best RefSeq golgin A7 family, member Ba
NW_001878	Dr. 116170			si:ch211-114113.9	+	ZFIN ug sv pr dl ev hm		best RefSeq
NW_001878	Dr. 75674			dnm2l	+	ZFIN ug sv pr dl ev hm sts		best RefSeq dynamin2-like
NW_003334	Dr. 104633							
NW_001878	Dr. 76185							
NW_001878	Dr. 76262							
NW_001878	Dr. 32392							
NW_001878	Dr. 4740							
NW_001878	Dr. 1111							
NW_001878	Dr. 87556							
NW_001884	Dr. 4960							
NW_001884	Dr. 5694							
NW_001878	Dr. 77108							
NW_001878	Dr. 72266							
NW_003039	Dr. 76034							
NW_003334	Dr. 77654							

Region Shown:

out
 zoom
 in

You are here: 

default
 master

UCSC Genome Browser

The screenshot displays the UCSC Genome Browser interface for the Zebrafish genome (Zv9/danRer7) assembly. The main track area shows a 20 kb region of chromosome 13, with coordinates ranging from 25,010,000 to 25,080,000. The interface includes a navigation bar at the top with links for Genomes, Genome Browser, Tools, Mirrors, Downloads, My Data, About Us, View, and Help. Below the navigation bar, there are controls for moving and zooming, including a search bar containing the coordinates chr13:25,004,825-25,082,273 (77,449 bp) and a 'go' button. The main track area displays various genomic features, including RefSeq Genes, Other RefSeq, Zebrafish Gene Collection Full ORF mRNAs, Ensembl Gene Predictions (Ensembl 71), Human Proteins Mapped by Chained tBLASTn, Zebrafish mRNAs from GenBank, Zebrafish ESTs That Have Been Spliced, Fish Basewise Conservation by PhyloP (5 species), Fish Conservation by PhastCons (5 species), Multiz Alignment & Conservation (8 Species), Multiz Alignments & Conservation (8 Species), and Repeating Elements by RepeatMasker. The bottom of the interface features a 'move start' and 'move end' section with a '2.0' zoom level and navigation arrows.

caveat: check the assembly version!

Ensembl Genome Browser

 [BLAST/BLAT](#) | [BioMart](#) | [Tools](#) | [Downloads](#) | [Help & Documentation](#) | [Blog](#) | [Mirrors](#) [Login/Register](#)

Search: for

e.g. [BRCA2](#) or [rat X:100000..200000](#) or [coronary heart disease](#)

What's New in Release 75 (February 2014)

- [New VEP interface](#)
- [New 'Age of base' track for human](#)
- [New GENCODE basic renderer for human and mouse](#)

[Full details of this release](#)
[More release news on our blog →](#)

Latest blog posts




- 15 May 2014: [GRC Assembly Workshop at Genome Informatics](#)
- 28 Apr 2014: [Sexual parasitism in the Amazon molly](#)
- 17 Apr 2014: [Ensembl Release Cycle and GRCh38](#)

[Go to Ensembl blog →](#)

Browse a Genome

The Ensembl project produces genome databases for vertebrates and other eukaryotic species, and makes this information freely available online.

Popular genomes

 Human GRCh37	 Mouse GRCm38
 Zebrafish Zv9	

★ [Log in to customize this list](#)

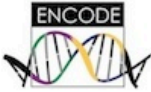
All genomes

-- Select a species --

[View full list of all Ensembl species](#)

Other species are available in [Ensembl Pre!](#) and [EnsemblGenomes](#)

ENCODE data in Ensembl



Variant Effect Predictor



Gene expression in different tissues



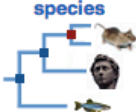
Find SNPs and other variants for my gene

```
GTATATACATTCC  
CRRRAAAGTCTT  
CTTCTAAATTCT  
GAAACATTTCC
```

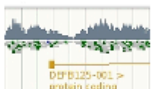
Retrieve gene sequence

```
GCCTGACTTCGGGTGG:  
GGGCTTGTGGCGGAGCC  
GGGCTCTGCTGGCCCT:  
AGGGGACAGATTTGTGA:  
CACCTCTGGAGCGGTTC:  
CCCACTCCAGCGTGGCG:
```

Compare genes across species




Use my own data in Ensembl



Learn about a disease or phenotype



Did you know...?



If you're using the [VEP script](#), please use [the cache](#). Our Variant Effect Predictor tool also has a [web interface](#).



Ensembl : What annotation is available?

- **Genes** protein coding genes
 gene/transcript/peptide models (coding and non-coding)
- **IDs** in other databases
- **aligned data**, like cDNAs, RNAseq, peptides, micro array probes, BAC clones, etc.
- Cytogenetic bands, **markers**, **repeats** etc.
- **Comparative data** orthologues and paralogues, gene trees
 protein families
 whole genome alignments
 syntenic regions
- **Variation data** Single Nucleotide Polymorphisms (SNPs), indels, phenotypes,
 population frequencies, variant effect prediction (VEP), etc.
- **Regulatory data** e.g. regulatory elements from ENCODE
- **External resources** e.g. GRC trackhub, mapped next-gen reads

The Ensembl gene set

- Bimonthly releases, updated gene set ~ every 6 months
- New 'Genebuild' with every new assembly

The following archives are available for this page:

- [Ensembl 74: Dec 2013](#) (Zv9)
- [Ensembl 73: Sep 2013](#) (Zv9) - patched/updated gene set Jul 2013
- [Ensembl 72: Jun 2013](#) (Zv9)
- [Ensembl 71: Apr 2013](#) (Zv9) - patched/updated gene set Nov 2012
- [Ensembl 70: Jan 2013](#) (Zv9)
- [Ensembl 69: Oct 2012](#) (Zv9) - patched/updated gene set Jun 2012
- [Ensembl 68: Jul 2012](#) (Zv9)
- [Ensembl 67: May 2012](#) (Zv9) - patched/updated gene set Mar 2012
- [Ensembl 66: Feb 2012](#) (Zv9)
- [Ensembl 65: Dec 2011](#) (Zv9) - patched/updated gene set Oct 2011
- [Ensembl 64: Sep 2011](#) (Zv9) - patched/updated gene set May 2011
- [Ensembl 63: Jun 2011](#) (Zv9)
- [Ensembl 62: Apr 2011](#) (Zv9)
- [Ensembl 59: Aug 2010](#) (Zv8)
- [Ensembl 54: May 2009](#) (Zv8) - gene set updated Apr 2009

[More information about the Ensembl archives](#)

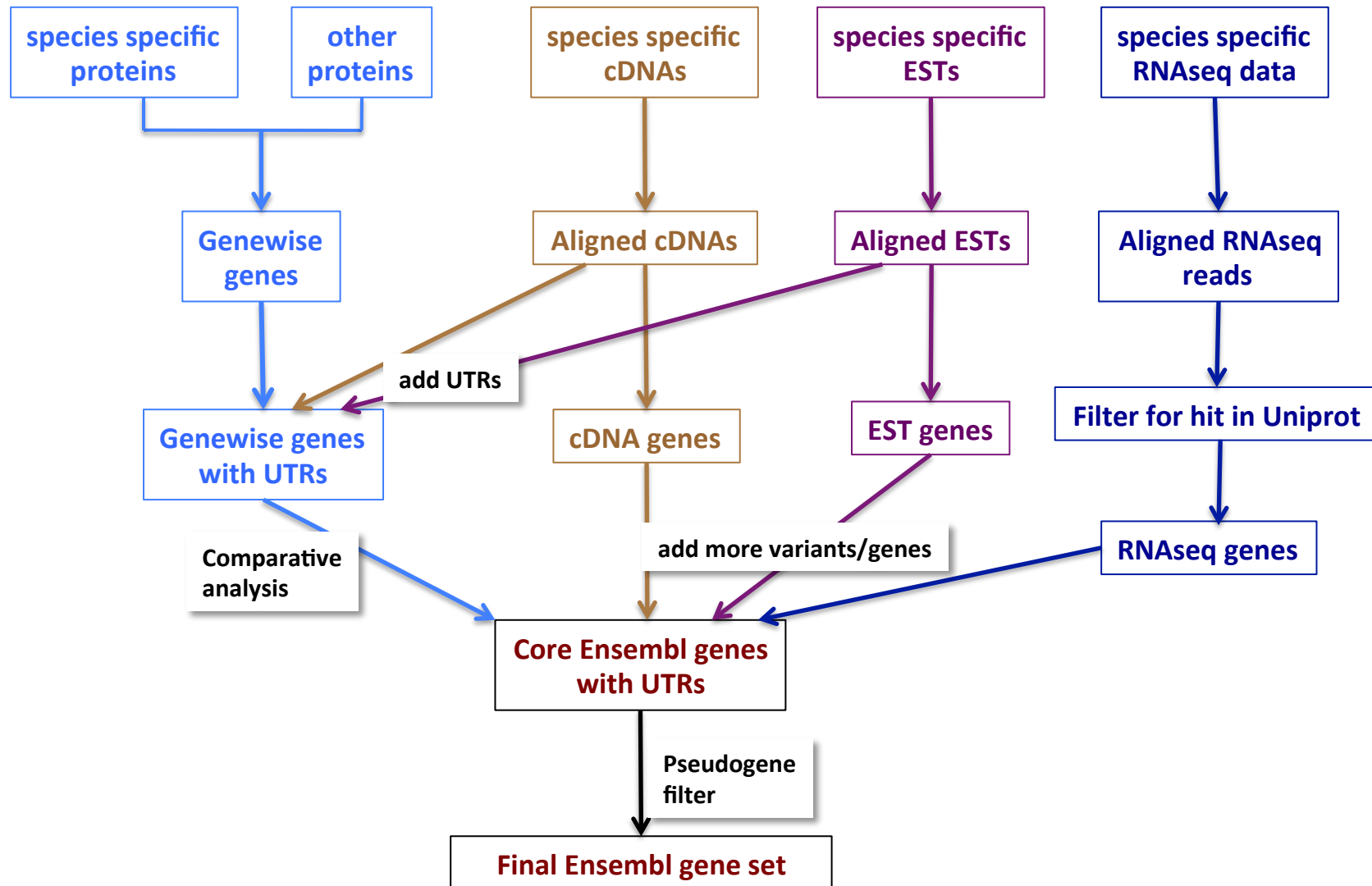
- **Genes are built on evidence, no gene is predicted on sequence alone!**

The Ensembl gene set

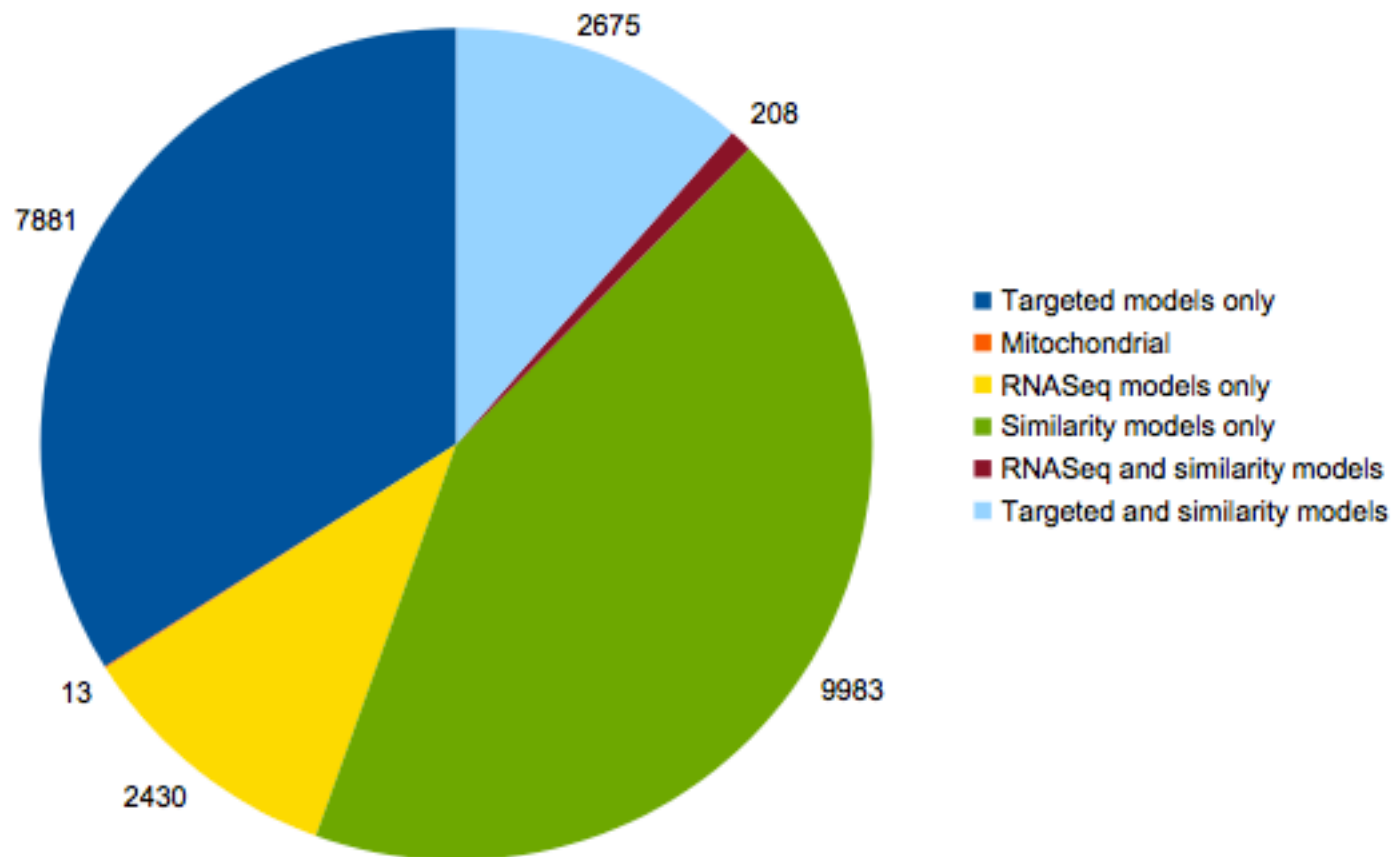
All Ensembl gene predictions are based on experimental evidence

- UniProt / Swiss-Prot
A manually curated database and therefore of highest accuracy
- NCBI RefSeq
A partially manually curated database
- UniProt / TrEMBL
Automatically annotated translations of ENA coding sequence (CDS) features
- ENA / GenBank / DDBJ
Primary nucleotide sequence repository

The Ensembl gene build

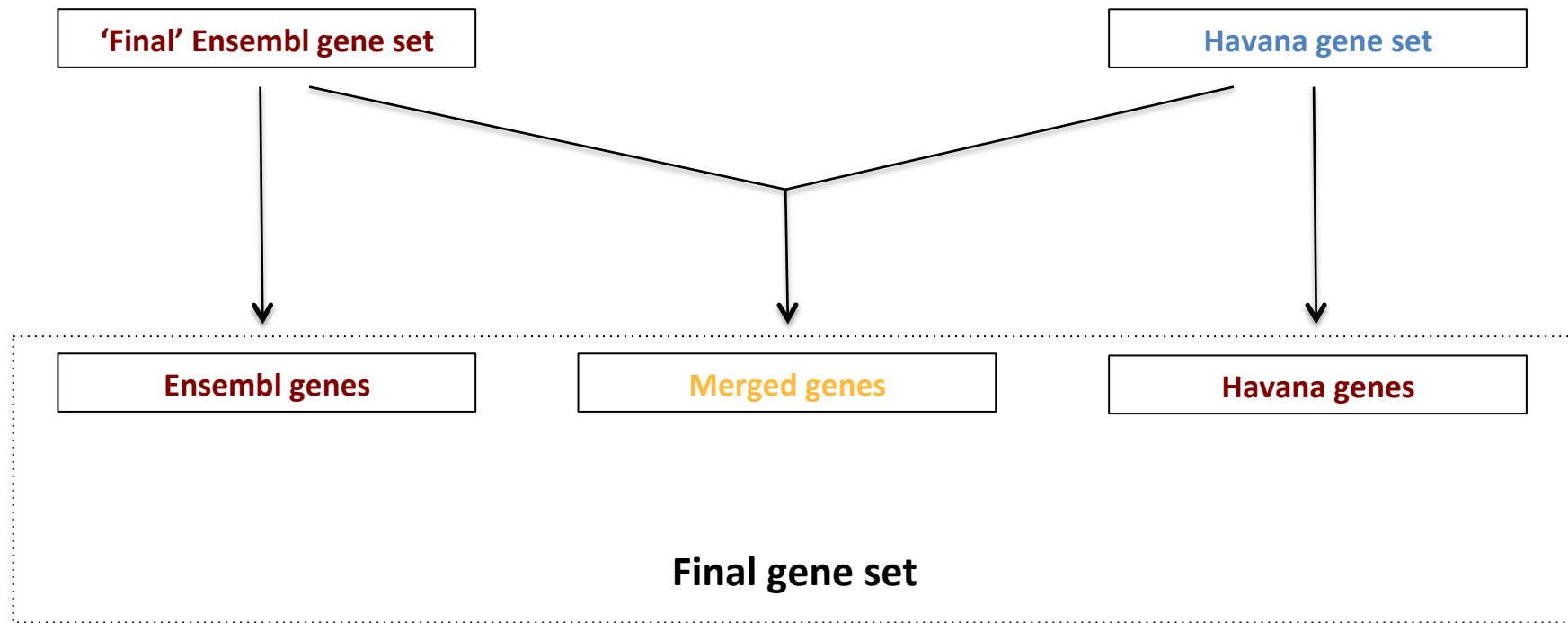


Supporting evidence for final Ensembl gene set



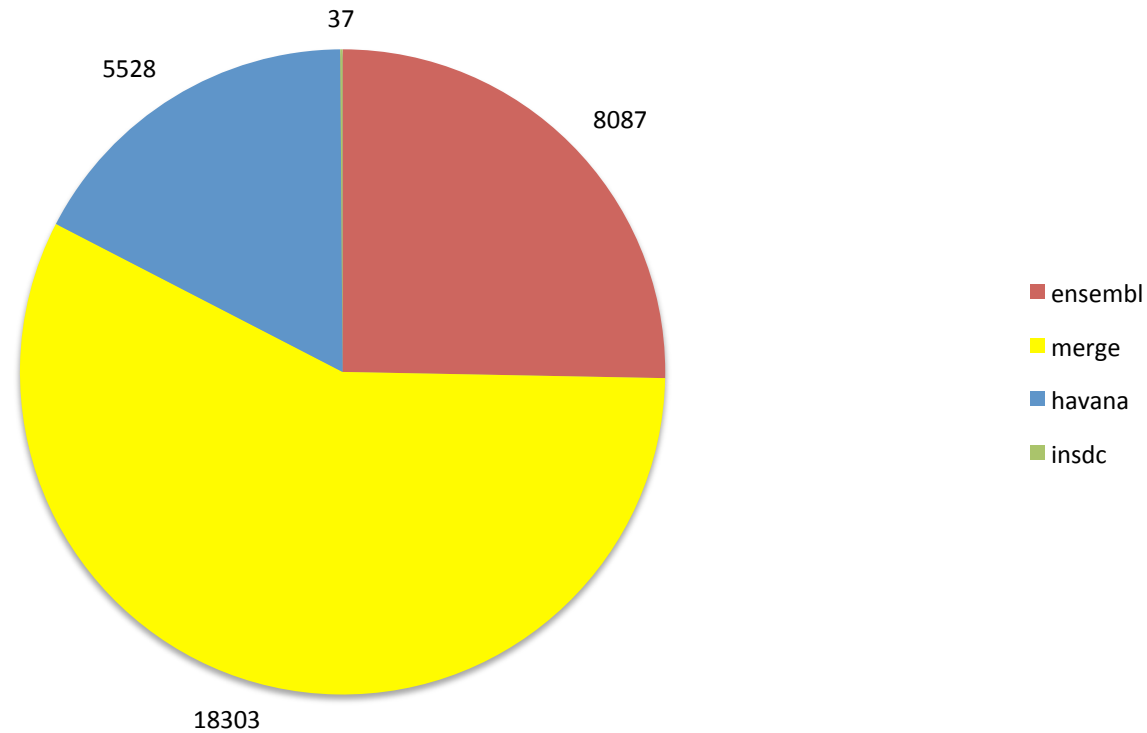
The Ensembl gene build : Gene merge

Getting the optimum gene set by combining automated and manual annotation

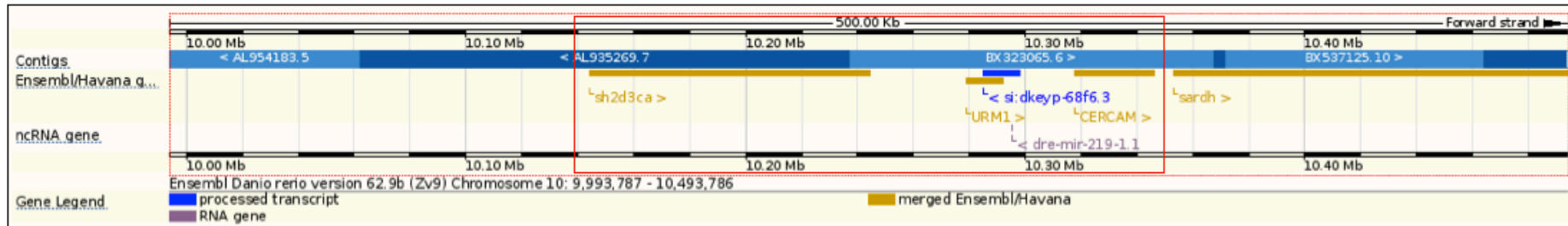


The Ensembl gene build : Gene merge

Getting the optimum gene set by combining automated and manual annotation



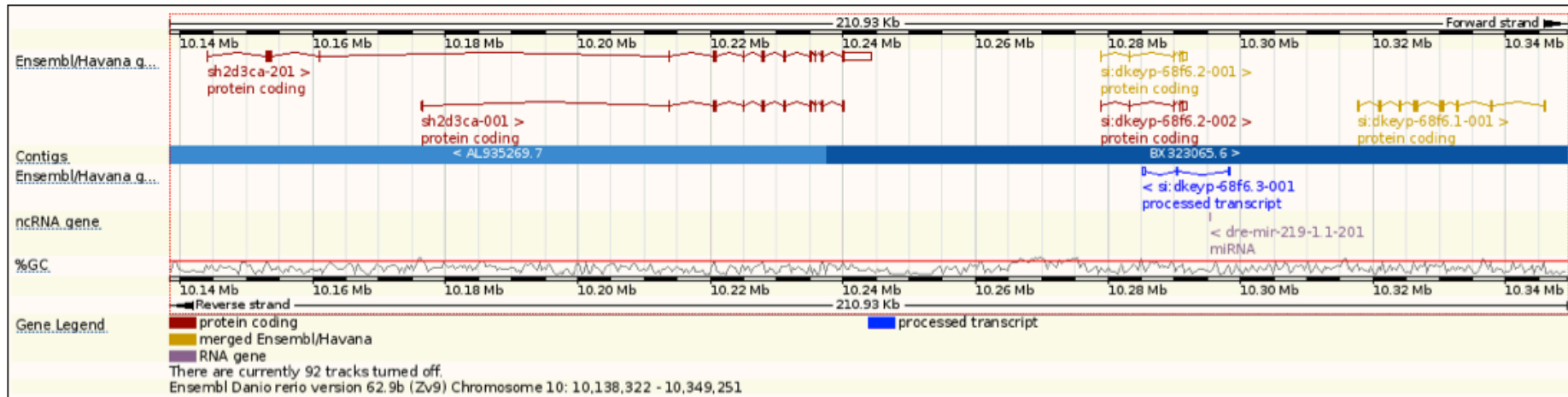
The Ensembl gene build : Gene merge



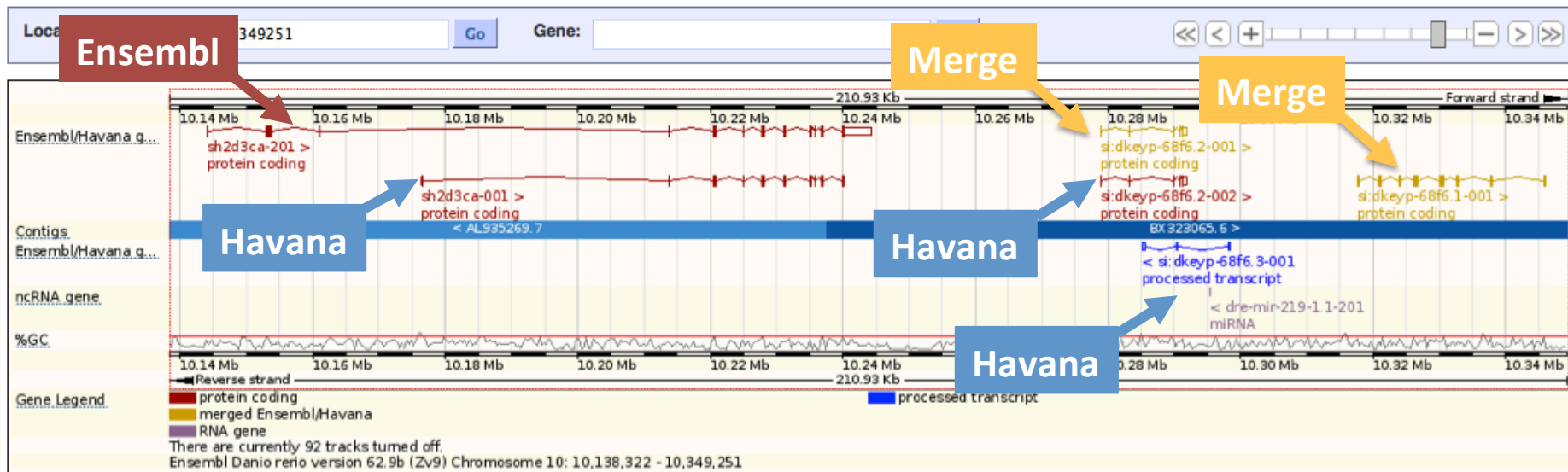
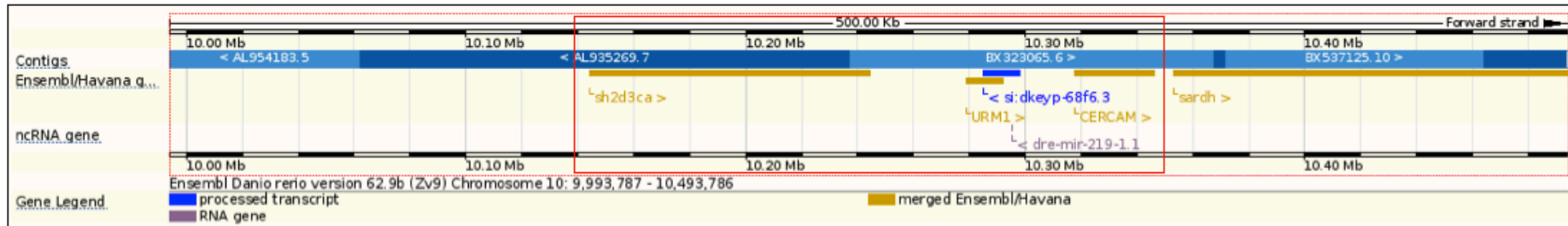
Export Image

Location: Go Gene: Go

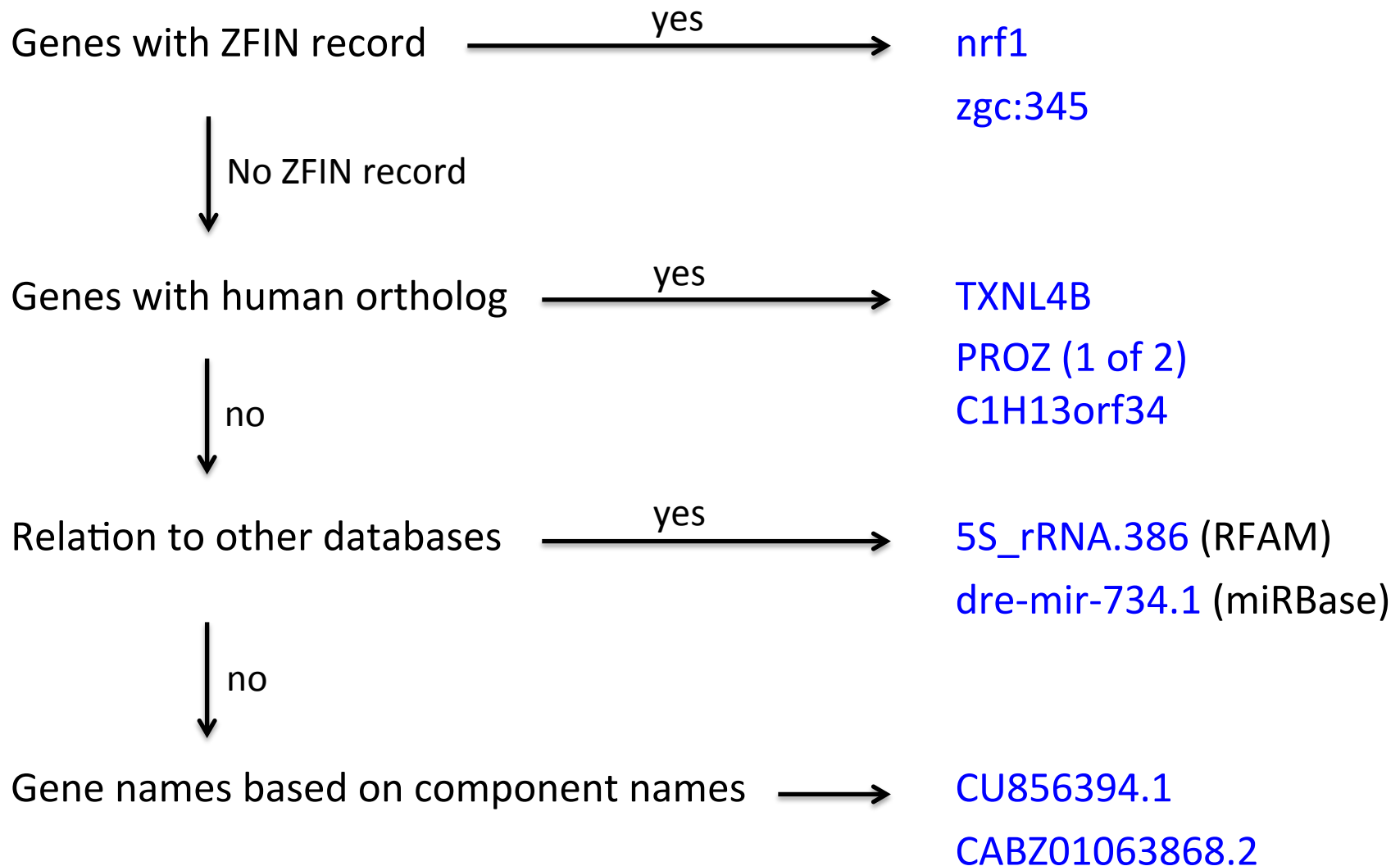
Navigation controls: << < + [slider] - > >>



The Ensembl gene build : Gene merge



Ensembl gene names



Ensembl stable identifiers

- ENSG### Ensembl Gene ID
- ENST### Ensembl Transcript ID
- ENSP### Ensembl Peptide ID
- ENSE### Ensembl Exon ID

- For other species than human a suffix is added, e.g.
for mouse (*Mus musculus*): ENSMUSG###
for zebrafish DAR (*Danio rerio*): ENSDARG###
note RNASEQDAR[G/T/P/E]###

- For imported genes Ensembl uses the original identifiers

Access to Genome Annotation

- Release web site <http://www.ensembl.org/>
- Pre-Release <http://pre.ensembl.org/>
- Archive <http://archive.ensembl.org>

- BioMart <http://www.ensembl.org/biomart/martview>
- Flat file repository <ftp://ftp.ensembl.org/>

- MySQL interface ensemldb.ensembl.org
- Perl API <http://www.ensembl.org/info/docs/api>
- REST API <http://rest.ensembl.org>

Help and Information

- Zebrafish specific help zfish-help@sanger.ac.uk
- Zebrafish project <http://www.sanger.ac.uk/resources/zebrafish/genomeproject.html>
- Zebrafish at the GRC <http://www.genomereference.org>

- Ensembl helpdesk helpdesk@ensembl.org
- View animated tutorials
www.ensembl.org/info/website/tutorials/index.html
- Mailing lists: announce@ensembl.org
ensembl-dev@ensembl.org

Guided examples

- A stroll through Ensembl
- Introduction to BioMart
- Making your own data visible