

# Manual Annotation of the Mouse Genome: HAVANA and EUCOMM

Mark Thomas

# Overview

- HAVANA
  - Otterlace and Zmap
  - Annotation and Vector Design
- EUCOMM
  - Objectives
  - Conditional Targeting
- VEGA
- Results (current status)
  - Annotation
  - High Throughput Gene Targeting (HTGT) resource



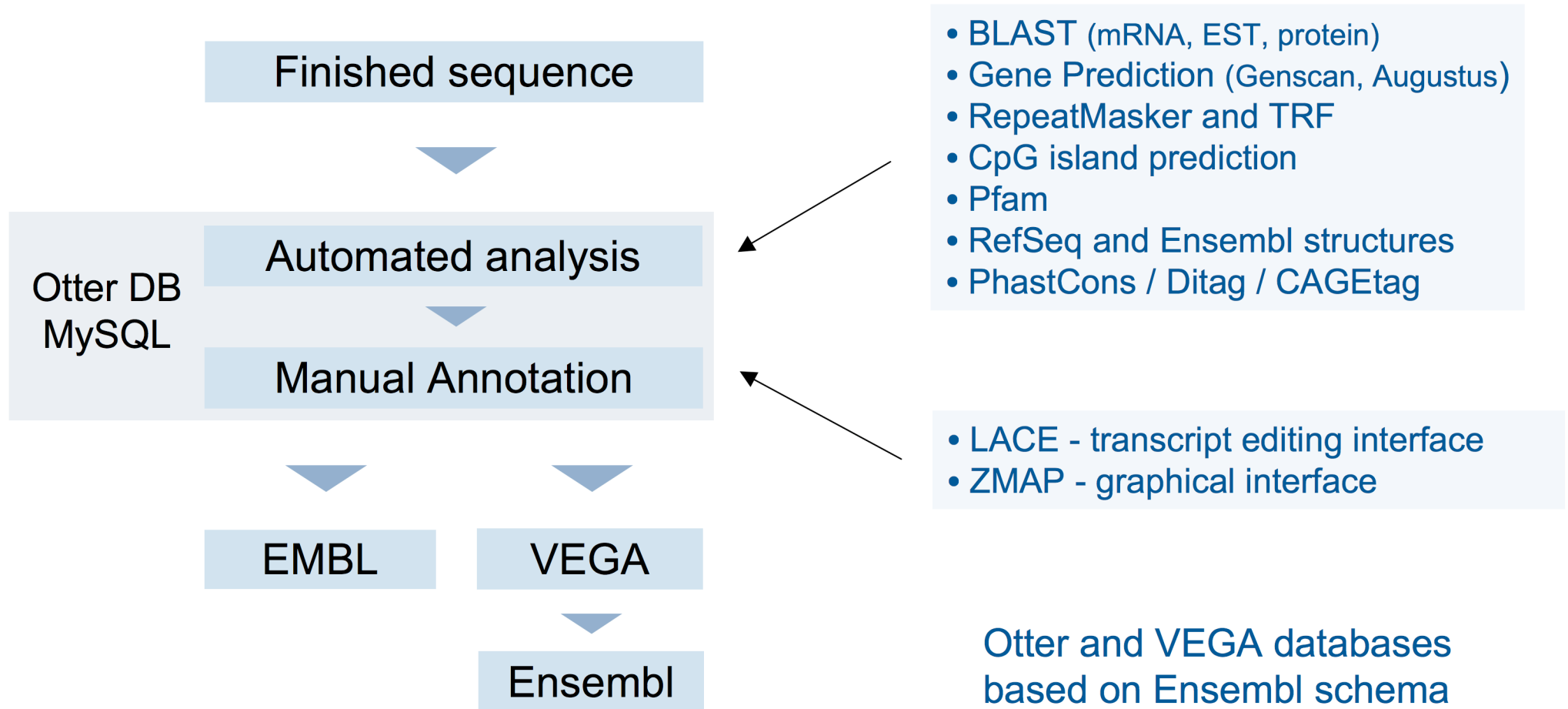


# HAVANA

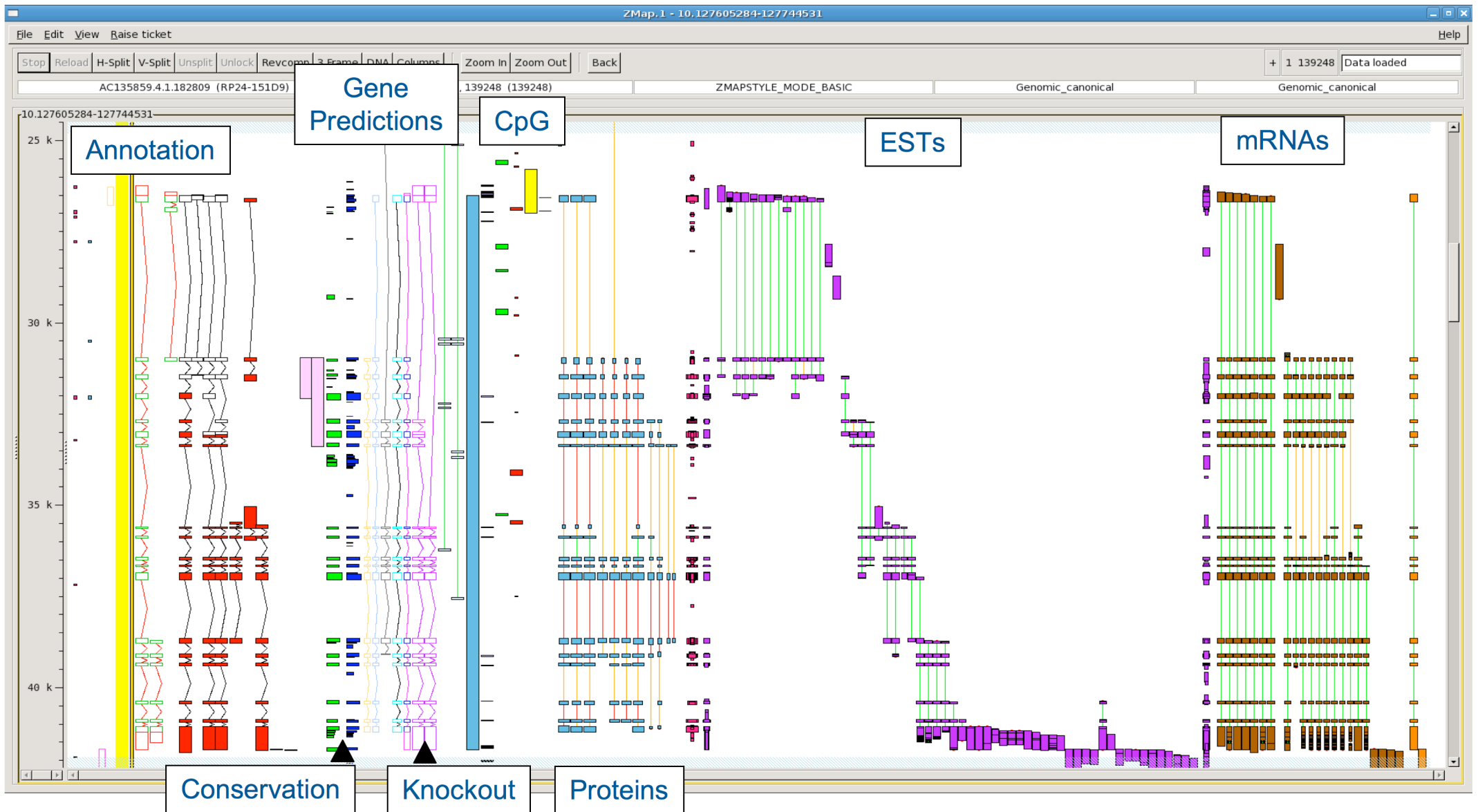
# HAVANA - manual annotation

- Otterlace used to manually annotate finished genomic sequence (primarily human, mouse and zebrafish) [Posters A3,C12 and C22 ]
- Annotation based on mRNA, EST and protein homology evidence only
- Annotation includes;
  - All transcripts including splice variants
  - Gene clusters
  - Pseudogenes
  - PolyA features

# HAVANA - annotation pipeline



# HAVANA - Otterlace and Zmap



Nature Precedings : doi:10.1038/npre.2009.3181.1 : Posted 27 Apr 2009

# HAVANA - Otterlace and Zmap

Nature Precedings : doi:10.1038/npre.2009.3181.1 : Posted 27 Apr 2009

The screenshot displays the HAVANA genome browser interface for chromosome 10, clone 949. The main window is titled "lace chr10-03, clone 949" and contains a list of transcripts. The transcript **RP24-151D9.3-001** is highlighted in yellow. Below the transcript list, there are three tabs: "Conservation", "Knockout", and "Proteins". To the right, a genomic map shows "ESTs" (Expressed Sequence Tags) and "mRNAs" (messenger RNAs) with their respective exon-intron structures and alignments. The map includes a scale bar at the top right showing "1 139248 Data loaded".

Transcript ID	Transcript ID	Transcript ID	Transcript ID
ESTT59605	<b>RP24-151D9.3-005</b>	PF05029.1	MUST83014
ESTT59607	MUST44776	<b>RP24-151D9.1-008</b>	genscan.3
MUST105247	ESTT59427	<b>RP24-151D9.1-015</b>	CCDS24270.1
MUST92033	<b>RP24-151D9.3-004</b>	<b>RP24-151D9.1-007</b>	augustus.5
MUST76233		<b>RP24-151D9.1-014</b>	MUST26449
CCDS24261.1	genscan.6	<b>RP24-151D9.1-013</b>	
	CCDS24264.1	<b>RP24-151D9.1-012</b>	
ESTT59424	augustus.7	<b>RP24-151D9.1-011</b>	CCDS24271.1
	MUST61995	PF04821.1	ESTT59510
MUST99139		augustus.6	ESTT59508
ESTT59604	ESTT59500	CCDS24266.1	ESTT59507
	ESTT59498	<b>RP24-151D9.1-006</b>	MUST105237
PF01199.1	ESTT59491	MUST105245	MUST105236
	PF02864.1	<b>RP24-151D9.1-002</b>	MUST5825
	PF02865.1	<b>RP24-151D9.1-005</b>	
<b>KO: RP24-151D9.3-002</b>	genscan.2	<b>RP24-151D9.1-010</b>	
<b>KO: RP24-151D9.3-001</b>	CCDS24269.1	MUST55539	
<b>RP24-151D9.3-001</b>	augustus.2	MUST105246	
PF04960.1	MUST105238	<b>RP24-151D9.1-009</b>	
	MUST85708	OTTMUST00000059536	
augustus.4	ESTT59486	<b>RP24-151D9.1-003</b>	
	ESTT59488	<b>RP24-151D9.1-004</b>	
<b>RP24-151D9.3-010</b>	ESTT59489	<b>KO: RP24-151D9.1-001</b>	
ESTT59440	ESTT59476	<b>RP24-151D9.1-001</b>	
<b>RP24-151D9.3-003</b>	genscan.4	ESTT59470	
<b>RP24-151D9.3-009</b>	augustus.1	PF01017.1	
<b>RP24-151D9.3-007</b>	CCDS24268.1	PF00230.1	
augustus.8	MUST50901	CCDS24265.1	
<b>RP24-151D9.3-008</b>	CCDS24267.1	augustus.3	
<b>RP24-151D9.3-002</b>	ESTT59429	MUST105239	ESTT59445
	MUST105239	MUST60782	MUST26455
<b>RP24-151D9.3-011</b>	<b>RP24-151D9.1-017</b>	<b>KO: RP24-151D9.2-001</b>	
genscan.1	<b>RP24-151D9.1-016</b>	<b>RP24-151D9.2-001</b>	
CCDS24263.1	ESTT59473	genscan.5	
<b>RP24-151D9.3-006</b>			

# HAVANA - Otterlace and Zmap

Nature Precedings : doi:10.1038/npre.2009.3181.1 : Posted 27 Apr 2009

The screenshot displays the HAVANA genome browser interface. The main window, titled 'lace chr10-03, clone 949', shows a list of gene models and their associated transcripts. The transcript 'RP24-151D9.3-001' is highlighted in yellow. Below the list, a genomic track shows conservation, knockout, and protein annotations. To the right, a detailed view of the transcript 'RP24-151D9.3-001' is shown, including its coordinates (26499 to 41217), type (Known\_CDS), and full name (glutaminase 2 (liver, mitochondrial)).

**Gene Models and Transcripts:**

Gene Model	Transcript
ESTT59605	RP24-151D9.3-005
ESTT59607	MUST44776
MUST105247	ESTT59427
MUST92033	RP24-151D9.3-004
MUST76233	genscan.6
CCDS24261.1	CCDS24264.1
ESTT59424	augustus.7
MUST99139	MUST61995
ESTT59604	ESTT59500
PF01199.1	ESTT59498
KO: RP24-151D9.3-002	ESTT59491
KO: RP24-151D9.3-001	ESTT59488
RP24-151D9.3-001	PF02864.1
PF04960.1	PF02865.1
augustus.4	genscan.2
RP24-151D9.3-010	CCDS24269.1
ESTT59440	augustus.2
RP24-151D9.3-003	MUST105238
RP24-151D9.3-009	MUST85708
RP24-151D9.3-007	ESTT59486
augustus.8	ESTT59488
RP24-151D9.3-008	ESTT59489
RP24-151D9.3-002	ESTT59476
ESTT59429	genscan.4
RP24-151D9.3-011	augustus.1
genscan.1	augustus.8
CCDS24263.1	CCDS24268.1
RP24-151D9.3-006	MUST50901
	CCDS24267.1
	MUST105239
	MUST60782
	RP24-151D9.1-017
	RP24-151D9.1-016
	ESTT59473
	PF05029.1
	RP24-151D9.1-008
	RP24-151D9.1-015
	RP24-151D9.1-007
	RP24-151D9.1-014
	RP24-151D9.1-013
	RP24-151D9.1-012
	RP24-151D9.1-011
	PF04821.1
	augustus.6
	CCDS24266.1
	RP24-151D9.1-006
	MUST105245
	RP24-151D9.1-002
	RP24-151D9.1-005
	RP24-151D9.1-010
	MUST55539
	MUST105246
	RP24-151D9.1-009
	RP24-151D9.1-003
	RP24-151D9.1-004
	KO: RP24-151D9.1-001
	RP24-151D9.1-001
	ESTT59470
	PF01017.1
	PF00230.1
	CCDS24265.1
	augustus.3
	ESTT59445
	MUST26455
	KO: RP24-151D9.2-001
	RP24-151D9.2-001
	genscan.5
	MUST83014
	genscan.3
	CCDS24270.1
	augustus.5
	MUST26449
	CCDS24271.1
	ESTT59510
	ESTT59508
	ESTT59507
	MUST105237
	MUST105236
	MUST5825

**Transcript Details (RP24-151D9.3-001):**

- Name: RP24-151D9.3-001
- Type: Known\_CDS
- Start: Found
- End: Found
- Symbol: Gls2
- Full name: glutaminase 2 (liver, mitochondrial)

**Genomic Tracks:** Conservation, Knockout, Proteins

# HAVANA - Otterlace and Zmap

Nature Precedings : doi:10.1038/npre.2009.3181.1 : Posted 27 Apr 2009

The screenshot displays the HAVANA genome browser interface. The main window, titled 'lace chr10-03, clone 949', shows a list of gene models and transcripts. The transcript 'RP24-151D9.3-001' is highlighted in yellow. Below the list, a genomic track shows conservation, knockout sites, and protein structures. To the right, a detailed view of the transcript 'RP24-151D9.3-001' is shown, including its coordinates (26499 to 41217) and a list of associated ESTs. A dropdown menu for 'Transcript' annotations is open, showing various categories like 'Known\_CDS', 'Novel\_CDS', and 'Pseudogene'.

**ESTs**

26499	26230	+	26680	41217
	30955	+	31054	
	31410	+	31531	
	31946	+	32075	
	32659	+	32738	
	32979	+	33142	
	33332	+	33390	
	35596	+	35628	
	35848	+	35906	
	36434	+	36500	
	36634	+	36684	
	36861	+	37037	
	38645	+	38776	
	39071	+	39163	
	39334	+	39395	
	40364	+	40440	
	40879	+	40942	
	41061	+	41712	

**Transcript**

Name: RP24-151D9.3-001

Type: Known\_CDS

Start: Known\_CDS

Remark: Nonsense\_mediated\_decay

Annotation: Pseudogene

Conservation Knockout Proteins



# HAVANA - Otterlace and Zmap

Nature Precedings : doi:10.1038/npre.2009.3181.1 : Posted 27 Apr 2009

The screenshot displays the HAVANA software interface with three main windows:

- Gene Browser (lace chr10-03, clone 949):** Shows a list of gene models and transcripts. The transcript **RP24-151D9.3-001** is highlighted in yellow. Below the list are tracks for Conservation, Knockout, and Proteins.
- ESTs Window:** Displays a list of ESTs associated with the selected transcript, including accession numbers like ESTT59605, MUST105247, and CCDS24261.1.
- Transcript Detail Window (RP24-151D9.3-001):** Shows the transcript's name, type (Known\_CDS), start and end coordinates (Found), and full name: glutaminase 2 (liver, mitochondrial).



# HAVANA - Otterlace and Zmap

Nature Precedings : doi:10.1038/npre.2009.3181.1 : Posted 27 Apr 2009

The screenshot displays the HAVANA genome browser interface. The main window, titled 'lace chr10-03, clone 949', shows a list of gene models and transcripts. The transcript 'RP24-151D9.3-001' is highlighted in yellow. Below the list, a genomic track shows conservation, knockout, and protein annotations. To the right, a detailed view of the transcript 'RP24-151D9.3-001' is shown, including its name, type (Known\_CDS), start and end coordinates, and remarks.

**Transcript Details:**

- Name: RP24-151D9.3-001
- Type: Known\_CDS
- Start: Found
- End: Found
- Remarks:
  - CDS not found - 1
  - CDS not found - 2
  - CDS not found - 3
- Locus: UTR incomplete
- Symbol: G1s2
- Full name: glutaminase 2 (liver, mitochondrial)
- Alias(es):
- Remarks:

**Gene Models (Left Panel):**

Gene Model	Transcript	Transcript	Transcript
ESTT59605	RP24-151D9.3-005	PF05029.1	
ESTT59607	MUST44776	RP24-151D9.1-008	MUST83014
MUST105247	ESTT59427	RP24-151D9.1-015	gscan.3
MUST92033	RP24-151D9.3-004	RP24-151D9.1-007	CCDS24270.1
MUST76233		RP24-151D9.1-014	augustus.5
CCDS24261.1	gscan.6	RP24-151D9.1-013	MUST26449
	CCDS24264.1	RP24-151D9.1-012	
ESTT59424	augustus.7	RP24-151D9.1-011	CCDS24271.1
	MUST61995	PF04821.1	ESTT59510
MUST99139		augustus.6	ESTT59508
ESTT59604	ESTT59500	CCDS24266.1	ESTT59507
	ESTT59498	RP24-151D9.1-006	MUST105237
PF01199.1	ESTT59491	MUST105245	MUST105236
	PF02864.1	RP24-151D9.1-002	MUST5825
	PF02865.1	RP24-151D9.1-005	
KO: RP24-151D9.3-002	gscan.2	RP24-151D9.1-010	
KO: RP24-151D9.3-001	CCDS24269.1	MUST55539	
RP24-151D9.3-001	augustus.2	MUST105246	
PF04960.1	MUST105238	RP24-151D9.1-009	
	MUST85708	OTTMUST00000059536	
augustus.4	ESTT59486	RP24-151D9.1-003	
	ESTT59488	RP24-151D9.1-004	
RP24-151D9.3-010	ESTT59489	KO: RP24-151D9.1-001	
ESTT59440	ESTT59476	RP24-151D9.1-001	
RP24-151D9.3-003	gscan.4	ESTT59470	
RP24-151D9.3-009	augustus.1	PF01017.1	
RP24-151D9.3-007	augustus.8	CCDS24268.1	
augustus.8	RP24-151D9.3-008	CCDS24265.1	
RP24-151D9.3-008	MUST50901	augustus.3	
RP24-151D9.3-002	CCDS24267.1	ESTT59445	
ESTT59429	MUST105239	MUST26455	
RP24-151D9.3-011	MUST60782	KO: RP24-151D9.2-001	
gscan.1	RP24-151D9.1-017	RP24-151D9.2-001	
CCDS24263.1	RP24-151D9.1-016		
RP24-151D9.3-006	ESTT59473	gscan.5	

# HAVANA - Otterlace and Zmap

Nature Precedings : doi:10.1038/npre.2009.3181.1 : Posted 27 Apr 2009

The screenshot displays the HAVANA genome browser interface. The main window shows a genomic track for 'lace chr10-03, clone 949'. The track includes various annotations such as ESTs (Exon Splicing Tables), Conservation, Knockout, and Proteins. A specific transcript, RP24-151D9.3-001, is highlighted in yellow. Below the main track, there are three buttons labeled 'Conservation', 'Knockout', and 'Proteins'. To the right, a detailed view of the transcript RP24-151D9.3-001 is shown, including its name, type (Known\_CDS), start and end coordinates, and full name (glutaminase 2 (liver, mitochondrial)).

**ESTs**

ESTT59605	RP24-151D9.3-005	PF05029.1	MUST83014
ESTT59607	MUST44776	RP24-151D9.1-008	gencan.3
MUST105247	ESTT59427	RP24-151D9.1-015	CCDS24270.1
MUST92033	RP24-151D9.3-004	RP24-151D9.1-007	augustus.5
MUST76233	gencan.6	RP24-151D9.1-014	MUST26449
CCDS24261.1	CCDS24264.1	RP24-151D9.1-013	
ESTT59424	augustus.7	RP24-151D9.1-012	CCDS24271.1
MUST99139	MUST61995	RP24-151D9.1-011	ESTT59510
ESTT59604	ESTT59500	PF04821.1	ESTT59508
PF01199.1	ESTT59498	augustus.6	ESTT59507
KO: RP24-151D9.3-002	ESTT59491	CCDS24266.1	MUST105237
KO: RP24-151D9.3-001	ESTT59491	MUST105245	MUST105236
RP24-151D9.3-001	PF02864.1	RP24-151D9.1-002	MUST5825
PF04960.1	PF02865.1	RP24-151D9.1-005	
augustus.4	gencan.2	RP24-151D9.1-010	
RP24-151D9.3-010	CCDS24269.1	MUST55539	
ESTT59440	augustus.2	MUST105246	
RP24-151D9.3-003	MUST105238	RP24-151D9.1-009	
RP24-151D9.3-009	MUST85708	RP24-151D9.1-003	
RP24-151D9.3-007	ESTT59486	RP24-151D9.1-004	
augustus.8	ESTT59488	KO: RP24-151D9.1-001	
RP24-151D9.3-008	ESTT59489	RP24-151D9.1-001	
RP24-151D9.3-002	ESTT59476	ESTT59470	
ESTT59429	gencan.4	PF01017.1	
RP24-151D9.3-011	augustus.1	PF00230.1	
gencan.1	CCDS24268.1	CCDS24265.1	
CCDS24263.1	MUST50901	augustus.3	
RP24-151D9.3-006	CCDS24267.1	ESTT59445	
	MUST105239	MUST26455	
	MUST60782	KO: RP24-151D9.2-001	
	RP24-151D9.1-017	RP24-151D9.2-001	
	RP24-151D9.1-016	RP24-151D9.2-001	
	ESTT59473	gencan.5	

**Transcript**

Name: RP24-151D9.3-001

Type: Known\_CDS

Start: Found End: Found

Remarks:

Annotation

**Locus**

Symbol: G1s2 Known

Full name: glutaminase 2 (liver, mitochondrial)

Alias(es):

Remarks:

Annotation

**Conservation** **Knockout** **Proteins**

# HAVANA - Otterlace and Zmap

Nature Precedings : doi:10.1038/npre.2009.3181.1 : Posted 27 Apr 2009

The screenshot displays the HAVANA software interface with three main windows:

- Gene Browser (lace chr10-03, clone 949):** Shows a list of gene models and transcripts. The transcript **RP24-151D9.3-001** is highlighted in yellow. Below the list are tracks for Conservation, Knockout, and Proteins.
- ESTs Window:** Displays a list of ESTs associated with the selected transcript, including identifiers like ESTT59605, MUST105247, and CCDS24261.1.
- Transcript Detail Window (RP24-151D9.3-001):** Shows the transcript's name, type (Known\_CDS), start and end coordinates (Found), and locus information (Symbol: G1s2, Full name: glutaminase 2 (liver, mitochondrial)).

# HAVANA - Otterlace and Zmap

Nature Precedings : doi:10.1038/npre.2009.3181.1 : Posted 27 Apr 2009

The screenshot displays the HAVANA software interface with three main windows:

- Gene Browser (lace chr10-03, clone 949):** Shows a list of gene models and transcripts. The transcript **RP24-151D9.3-001** is highlighted in yellow. Below the list are tracks for Conservation, Knockout, and Proteins.
- ESTs Window:** Displays a list of ESTs associated with the selected transcript, including accession numbers like ESTT59605, MUST105247, and CCDS24261.1.
- Transcript Detail Window (RP24-151D9.3-001):** Shows the transcript's name, type (Known\_CDS), start and end coordinates (Found), and full name: glutaminase 2 (liver, mitochondrial).



# HAVANA - Otterlace and Zmap

Nature Precedings : doi:10.1038/npre.2009.3181.1 : Posted 27 Apr 2009

The screenshot displays the HAVANA software interface for analyzing gene annotations. The main window, titled "lace chr10-03, clone 949", shows a list of transcripts and their associated evidence (ESTs, cDNAs, proteins). The transcript **RP24-151D9.3-001** is highlighted in yellow. A detailed view of this transcript is shown on the right, including its coordinates (26499 to 41217) and associated evidence (ESTs, cDNAs, proteins). A transcript details window is also open, showing the protein and cDNA sequences. At the bottom, there are buttons for "Conservation", "Knockout", and "Proteins".

Transcript	Evidence
ESTT59605	RP24-151D9.3-005
ESTT59607	MUST44776
MUST105247	ESTT59427
MUST92033	RP24-151D9.3-004
MUST76233	RP24-151D9.1-007
CCDS24261.1	RP24-151D9.1-014
	RP24-151D9.1-013
	RP24-151D9.1-012
	RP24-151D9.1-011
	PF04821.1
	augustus.6
	CCDS24264.1
	augustus.7
	MUST61995
	ESTT59500
	ESTT59498
	ESTT59491
	PF02864.1
	PF02865.1
	genscan.2
	CCDS24269.1
	augustus.2
	MUST105238
	MUST85708
	ESTT59486
	ESTT59488
	ESTT59489
	ESTT59476
	genscan.4
	augustus.1
	CCDS24268.1
	MUST50901
	CCDS24267.1
	augustus.3
	ESTT59429
	MUST105239
	MUST60782
	RP24-151D9.1-017
	RP24-151D9.1-016
	ESTT59473
	PF05029.1
	RP24-151D9.1-008
	RP24-151D9.1-015
	RP24-151D9.1-007
	RP24-151D9.1-014
	RP24-151D9.1-013
	RP24-151D9.1-012
	RP24-151D9.1-011
	PF04821.1
	augustus.6
	CCDS24266.1
	MUST105245
	RP24-151D9.1-002
	RP24-151D9.1-005
	RP24-151D9.1-010
	MUST55539
	MUST105246
	RP24-151D9.1-009
	OTTMUST00000059536
	RP24-151D9.1-003
	RP24-151D9.1-004
	KO: RP24-151D9.1-001
	RP24-151D9.1-001
	ESTT59470
	PF01017.1
	PF00230.1
	CCDS24265.1
	augustus.3
	ESTT59445
	MUST26455
	KO: RP24-151D9.2-001
	RP24-151D9.2-001
	genscan.5

# HAVANA - Otterlace and Zmap

Nature Precedings : doi:10.1038/npre.2009.3181.1 : Posted 27 Apr 2009

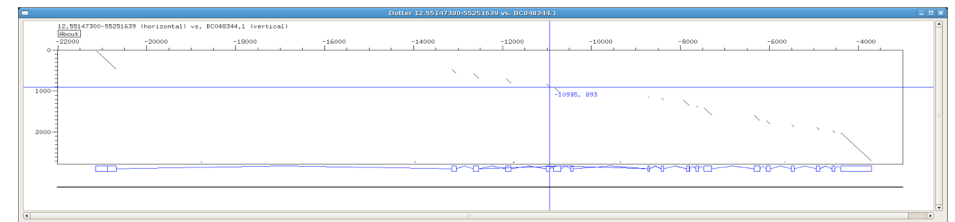
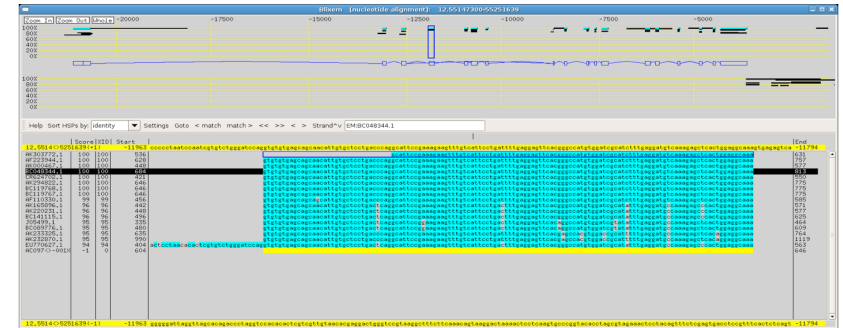
The screenshot displays the HAVANA software interface for genome annotation. The main window, titled 'lace chr10-03, clone 949', shows a list of transcripts and their associated evidence (ESTs, cDNAs, proteins). A transcript, **RP24-151D9.3-001**, is highlighted in yellow. A pop-up window titled 'Evi: RP24-151D9.3-001' shows the evidence for this transcript, including protein (Sw: P28492.3, Sw: Q571F8.2), cDNA (Em: AK165896.1, Em: AK220231.1, Em: BC141115.1), and EST (Em: AI195532.1, Em: BI148066.1, Em: CJ064043.1) records. Another window titled 'RP24-151D9.3-001 translation' shows the amino acid sequence of the protein:   
 >RP24-151D9.3-001  
 MRSMRALQNALSRAGSHGRRGGWGHPSRGPLLGRGVRYLLGAAAQGRGTPHSHQPQHS  
 HDASHSGMLPRLGDLDFYIAEQGERIPIHKFTTALKATGLQTSDFRLQDCMSKMQRMVQ  
 ESSSGLLDRELFQKCVSSNIVLLTQAFRKKFVIPDFEFTGHVDRIFEDAKEPTGGKVA  
 AYIPLHAKSNPDLWGVSLCTVDGQRHVSVGHTKIPFCLQSCVKPLTYAISVSTLGTDYVHK  
 FVGKEPSGLRYNKLNLNEEGIPHNPMVNAGAIIVSSLIKMDCNKAEKFDVFLQYLNMAG  
 NEFMGFSNATFQSEKETGDRNYAIGYYLKEKKCFPKGVDMMMAALDLYFQLCSVEVTCESG  
 SVMAAATLANGGICPITGESVLSAEAVRNTLSLHSCGMYDFSGQFAPFHVGLPAKSAVSGA  
 ILLVVPNVMGMMCLSPPLDKLNSQRGINFCQKLVSLFNFNHNYDNLRHCAKLDPRREGG  
 EVRNKTVVNLLFAAYSQDVSAALRRFALSAMDMEQKDYDSRTALHVAAAEGHIEVVKFLIE  
 ACKVNPVFKDRWGNIPLDNAVQFNHLEVVKLLQDYHDSYLLSETQAEAAAETLSKENLES  
 MV\*

A 'Transcript' window shows details for the selected transcript: Name: RP24-151D9.3-001, Type: Known\_CDS, Start: Found, End: Found. The Locus section shows Symbol: G1s2, Full name: glutaminase 2 (liver, mitochondrial), and Alias(es):. At the bottom, three tracks labeled 'Conservation', 'Knockout', and 'Proteins' are visible.

# HAVANA - Otterlace and Zmap

## Additional Tools

- Blixem - multiple alignment
- Dotter - pairwise alignment
- Cross-Species



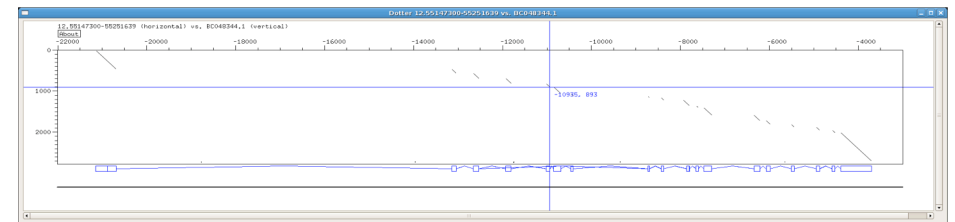
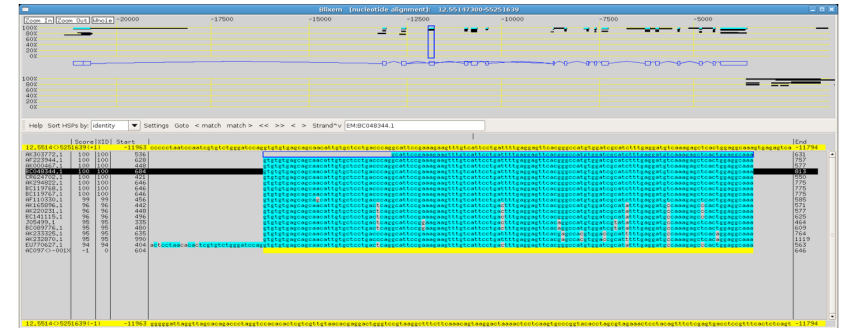




# HAVANA - Otterlace and Zmap

## Additional Tools

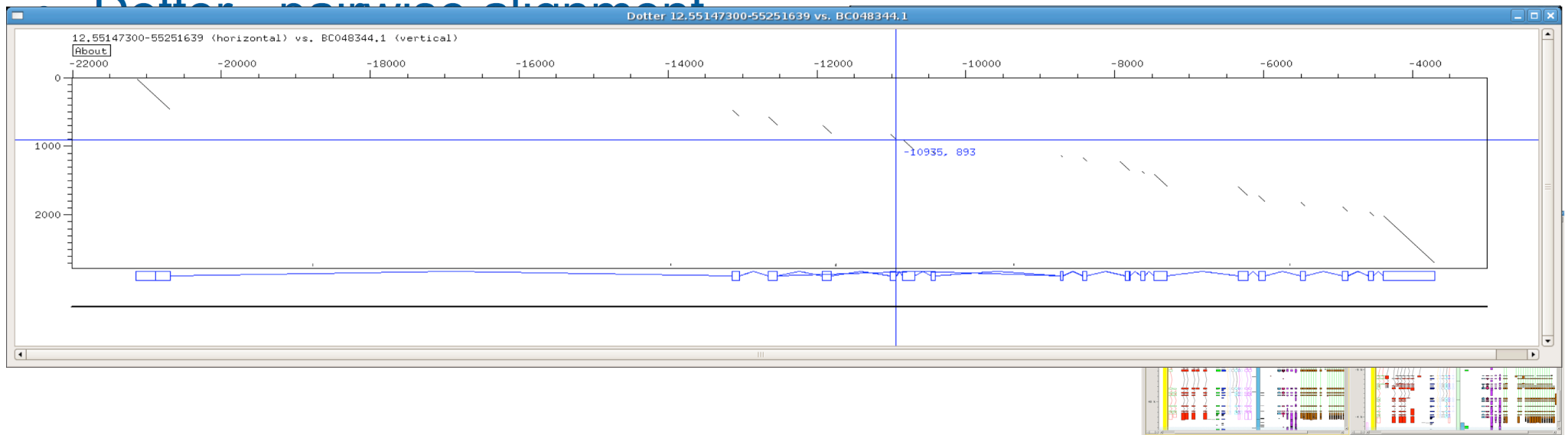
- Blixem - multiple alignment
- Dotter - pairwise alignment
- Cross-Species



# HAVANA - Otterlace and Zmap

## Additional Tools

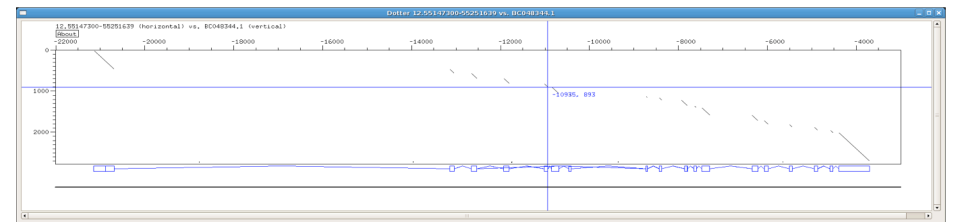
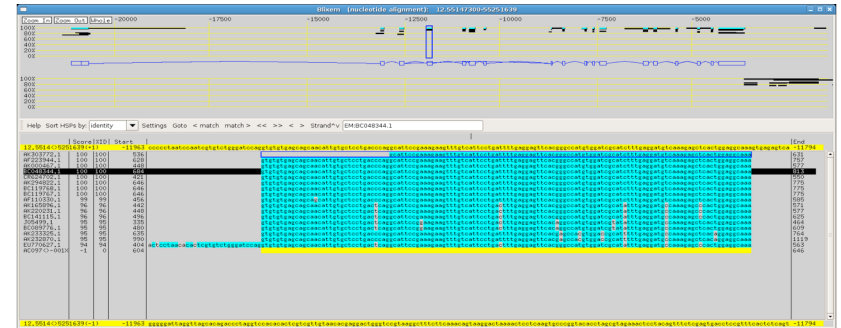
- Blix



# HAVANA - Otterlace and Zmap

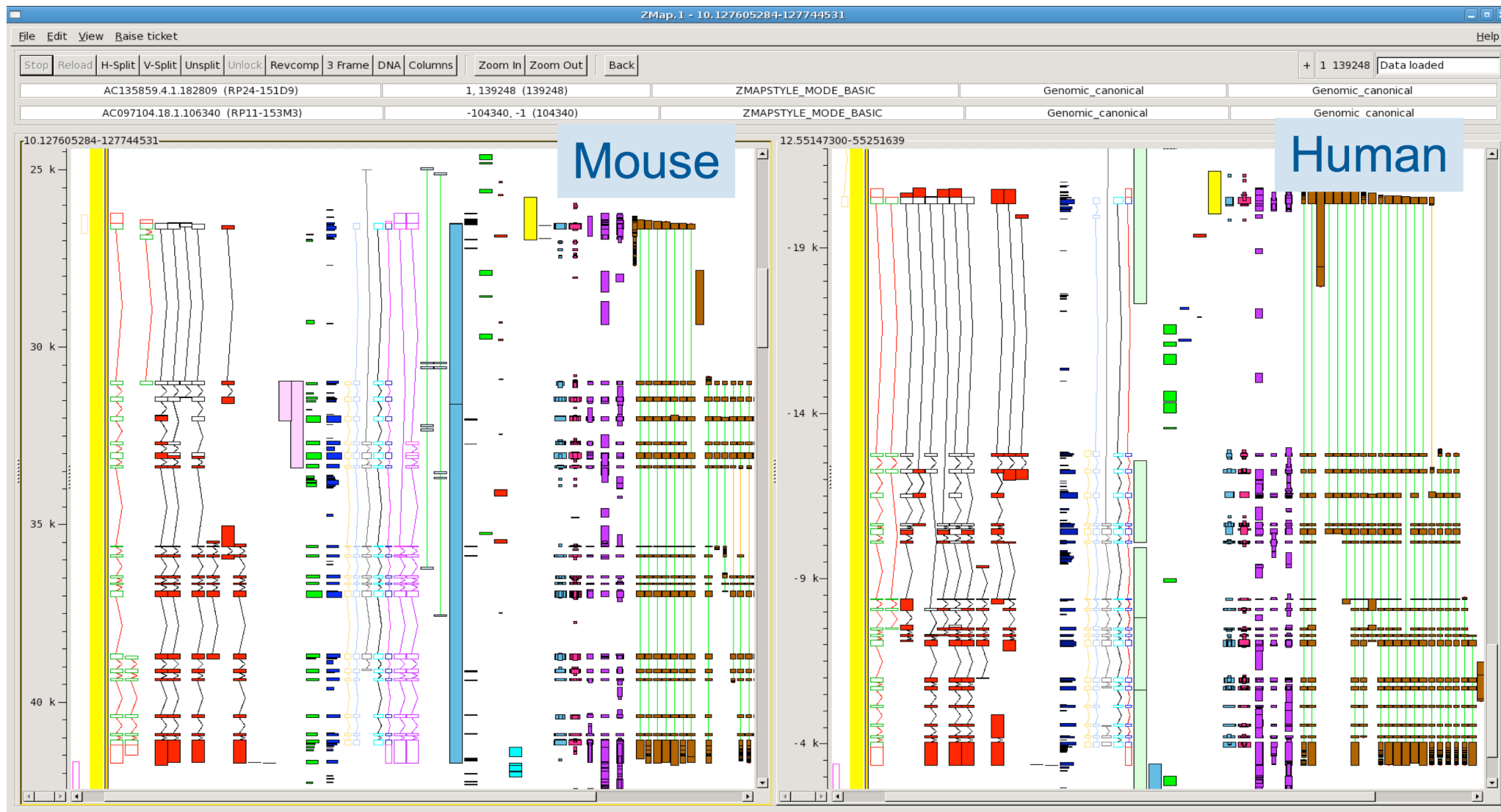
## Additional Tools

- Blixem - multiple alignment
- Dotter - pairwise alignment
- Cross-Species



# HAVANA - Otterlace and Zmap

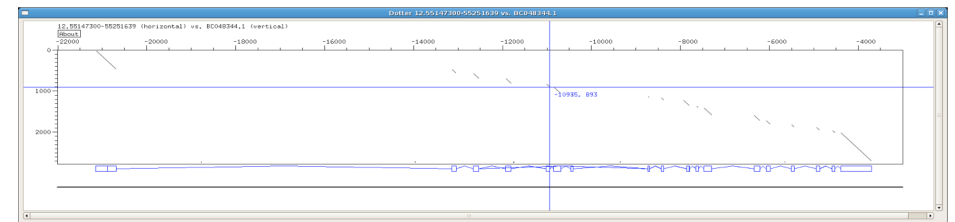
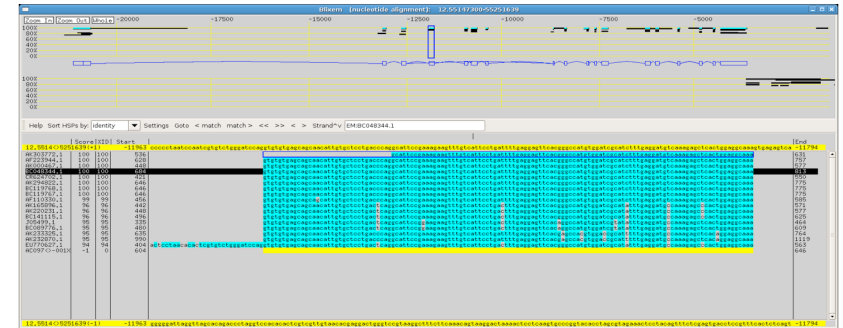
Nature Precedings : doi:10.1038/npre.2009.3181.1 : Posted 27 Apr 2009



# HAVANA - Otterlace and Zmap

## Additional Tools

- Blixem - multiple alignment
- Dotter - pairwise alignment
- Cross-Species



# EUCOMM

# EUCOMM (European Conditional Mouse Mutagenesis)

- Integrated Project funded by the EU Framework 6 programme
- 9 participating members in 4 countries
  - UK
    - Wellcome Trust Sanger Institute, Hinxton
    - MRC, Harwell
  - Germany
    - Helmholtz Zentrum (formerly GSF), Munich
    - CCR-Charité, Berlin
    - University of Frankfurt
    - University of Technology, Dresden
  - France
    - Institut Clinique de la Souris, Strasbourg
  - Italy
    - EMBL, Monterotondo
    - NRC, Monterotondo

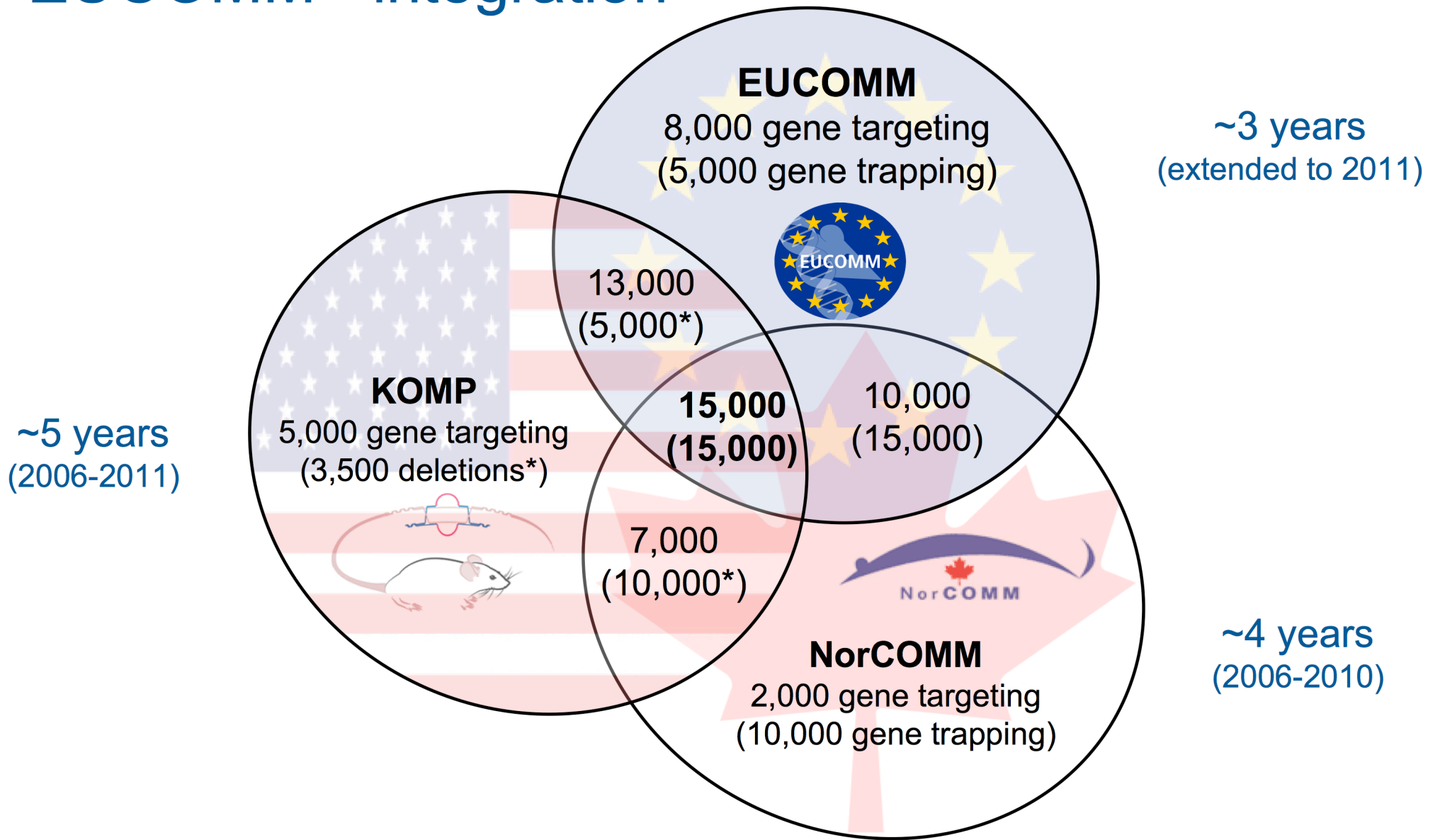


# EUCOMM - objectives

- Generation, archiving and distribution of up to 13,000 conditional mutations in mouse embryonic stem (ES) cells
  - 8,000 conditional gene targeting mutations
  - 5,000 conditional gene trapping mutations
- Production of up to 320 mouse mutant lines
- Creation of up to 14 mouse lines expressing the Cre recombinase in a tissue-specific manner

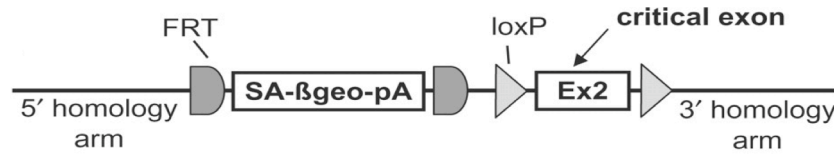


# EUCOMM - integration

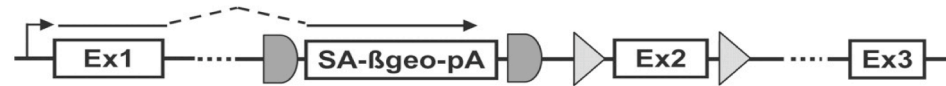


# EUCOMM - conditional targeting

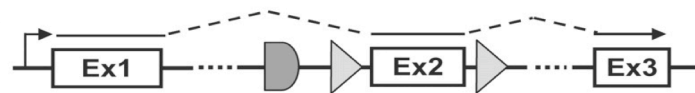
## A Targeting vector



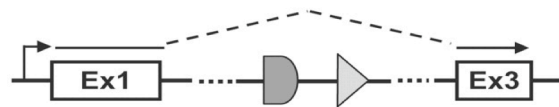
## B Targeted (knockout-first) allele



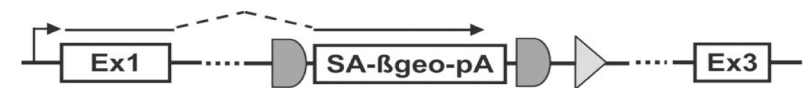
## C Conditional allele (wild-type)



## D Null allele (Δexon, frameshift)



## E lacZ-tagged null allele (Δexon)

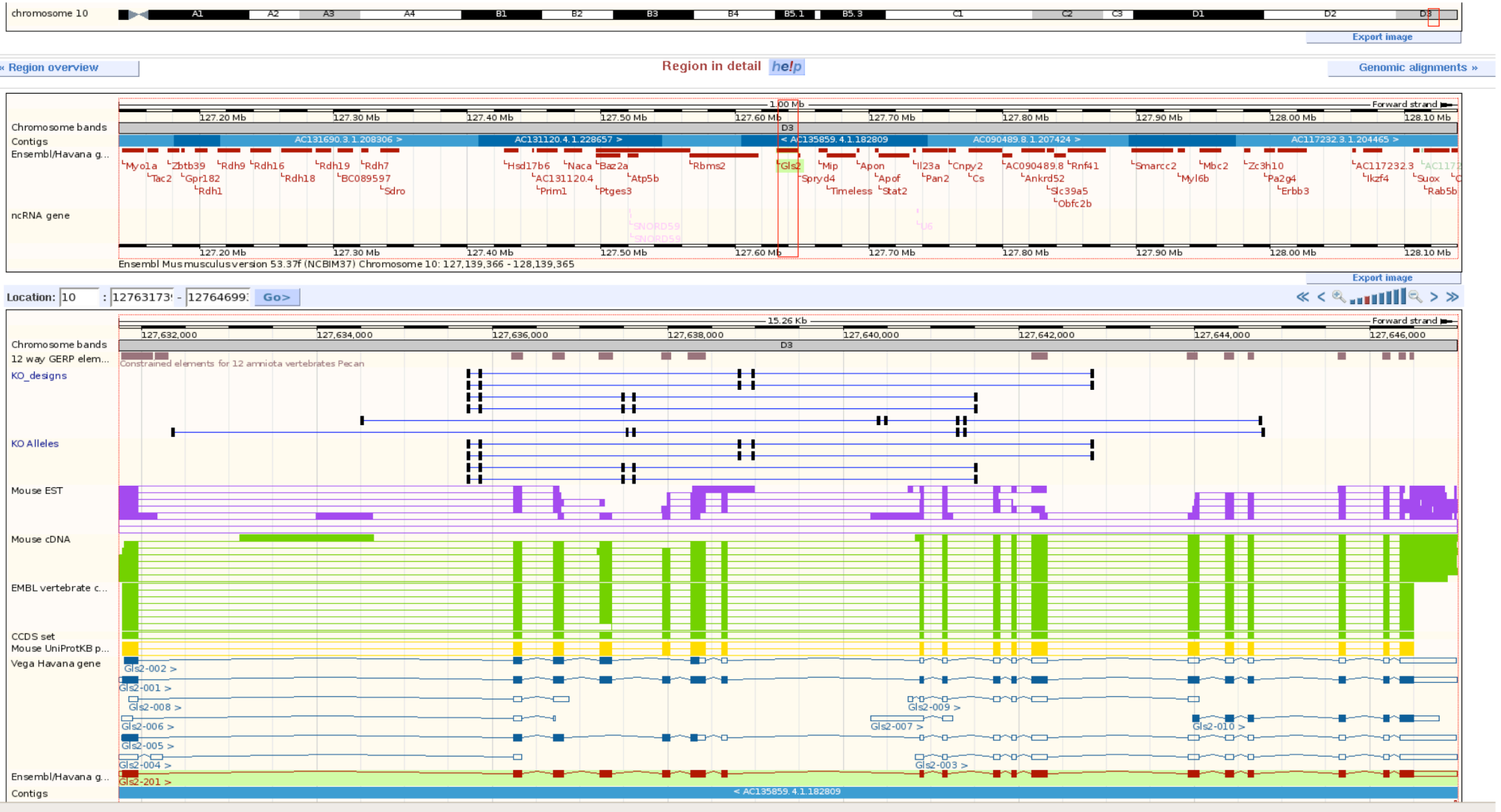


Method based on Testa et al. (2004)

Figure from Friedel et al. (2007)

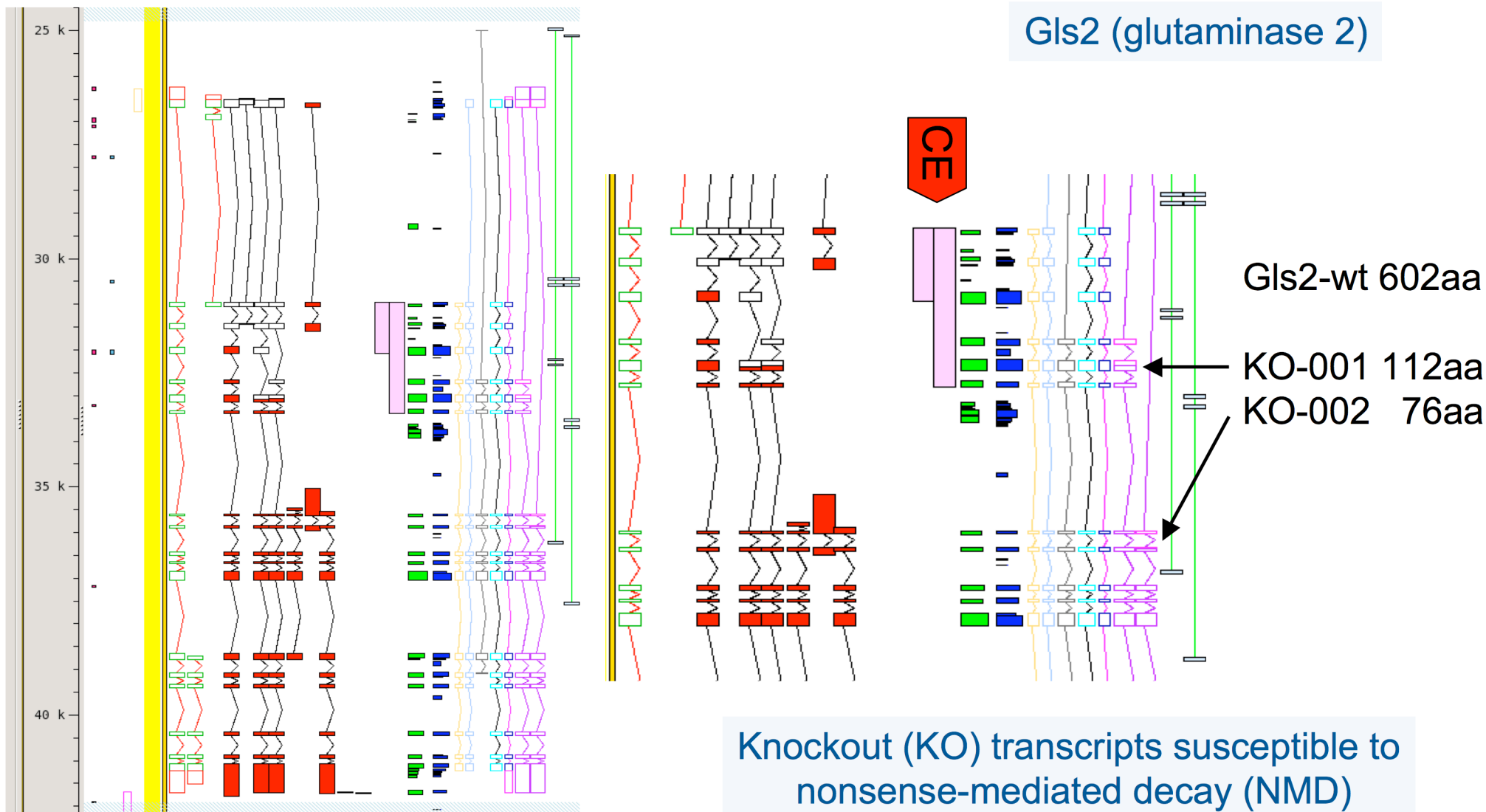
# EUCOMM - design selection (Ensembl)

Nature Precedings : doi:10.1038/npre.2009.3181.1 : Posted 27 Apr 2009

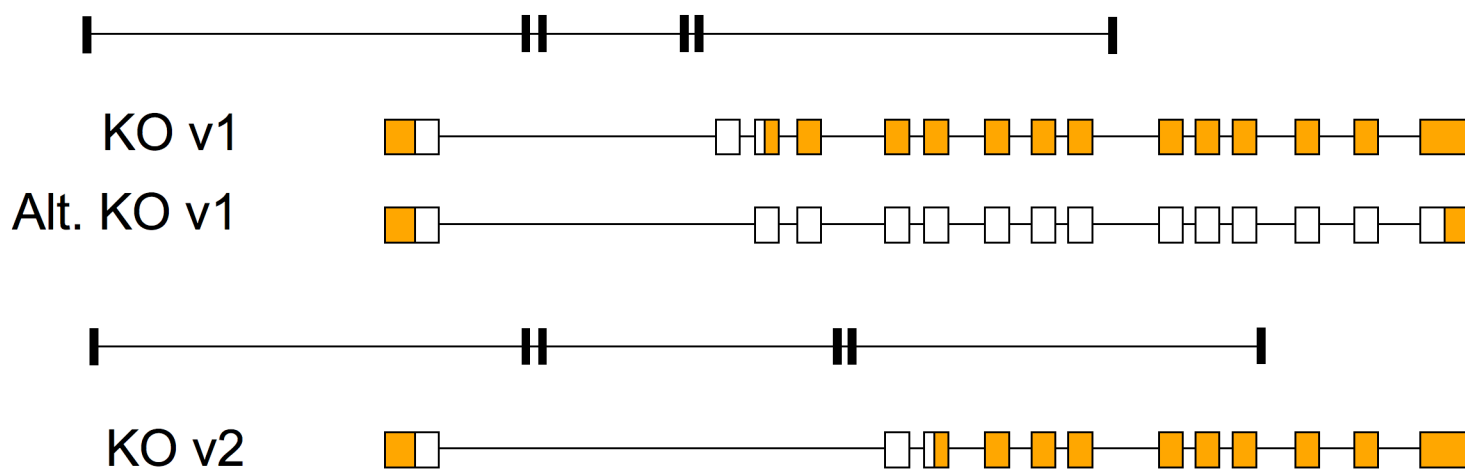
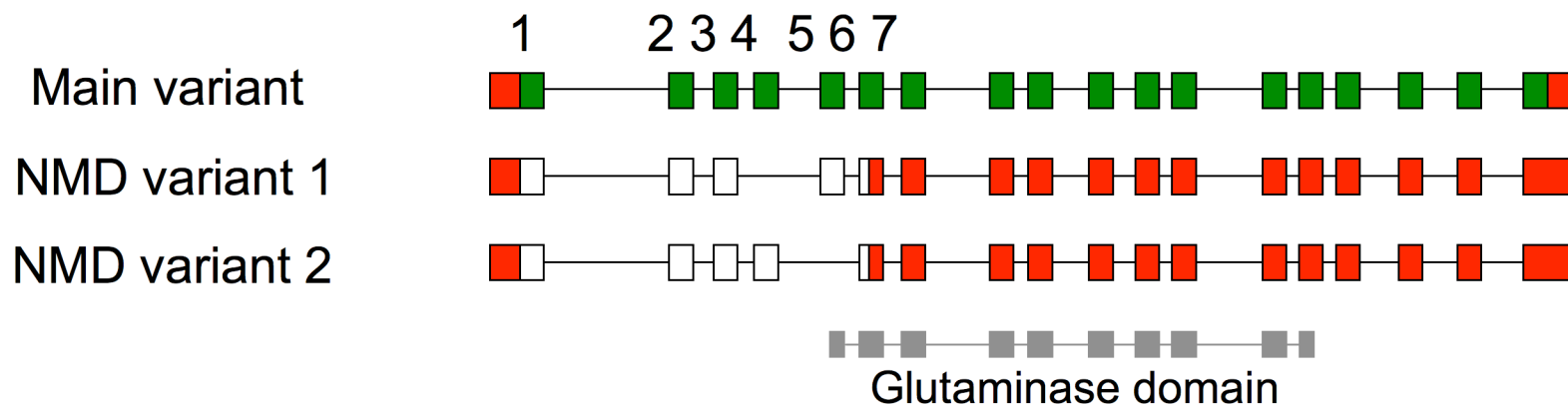


# EUCOMM - vector design

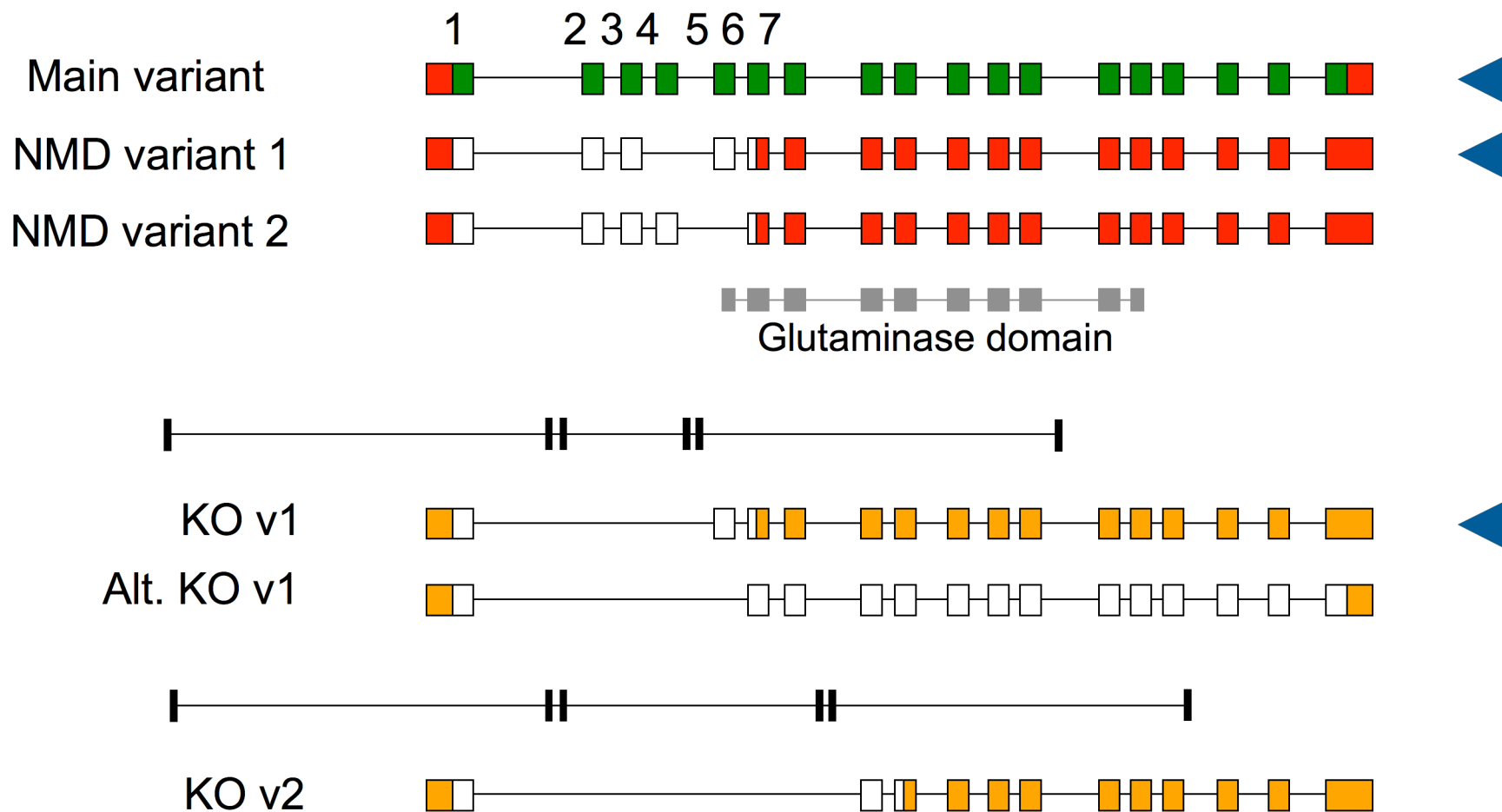
Gls2 (glutaminase 2)



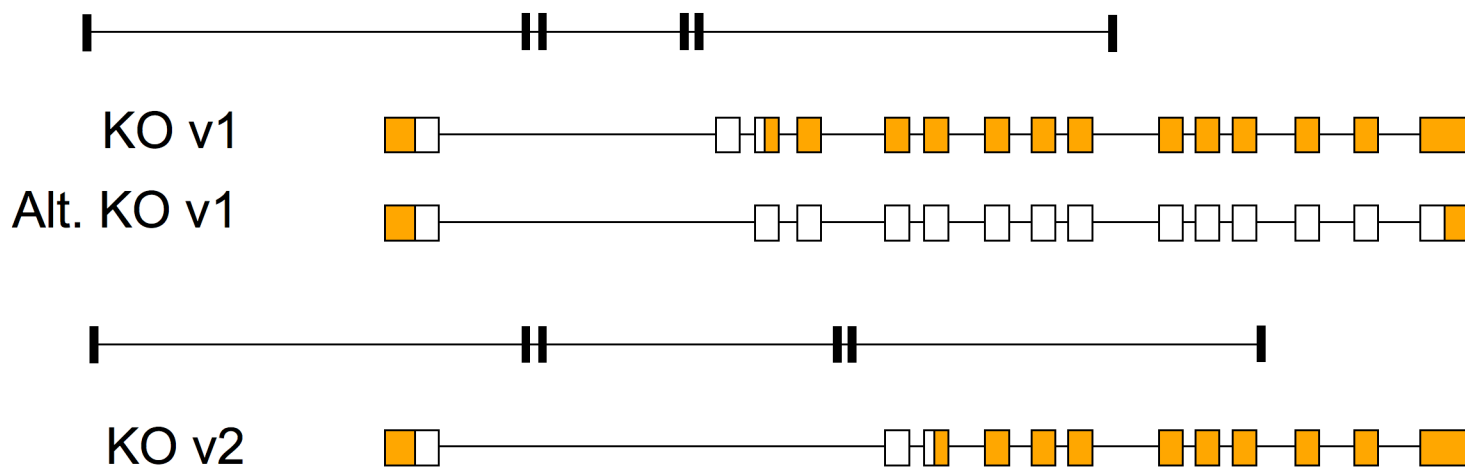
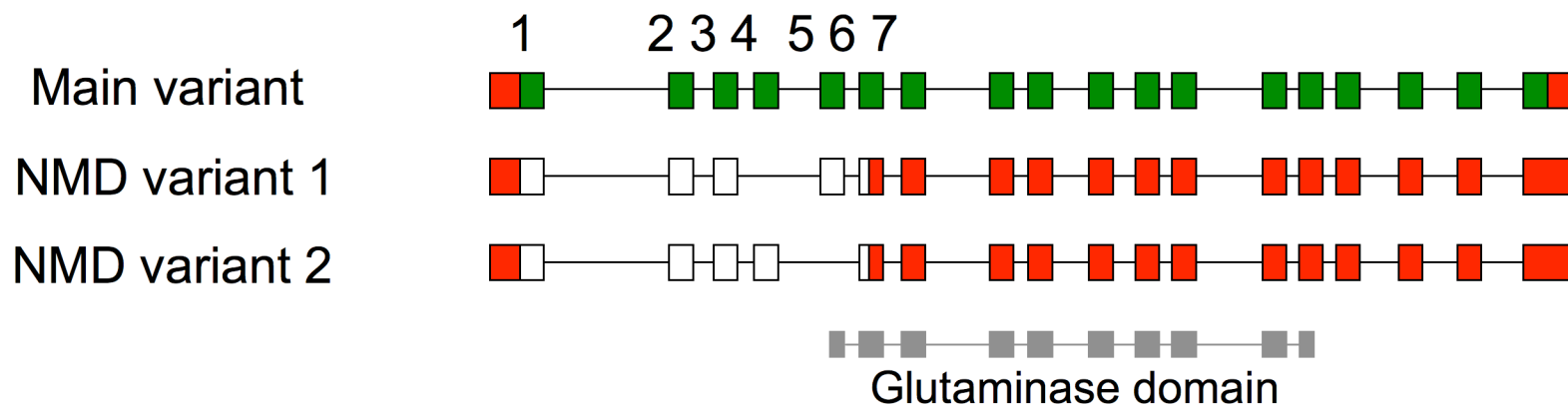
# EUCOMM - vector design



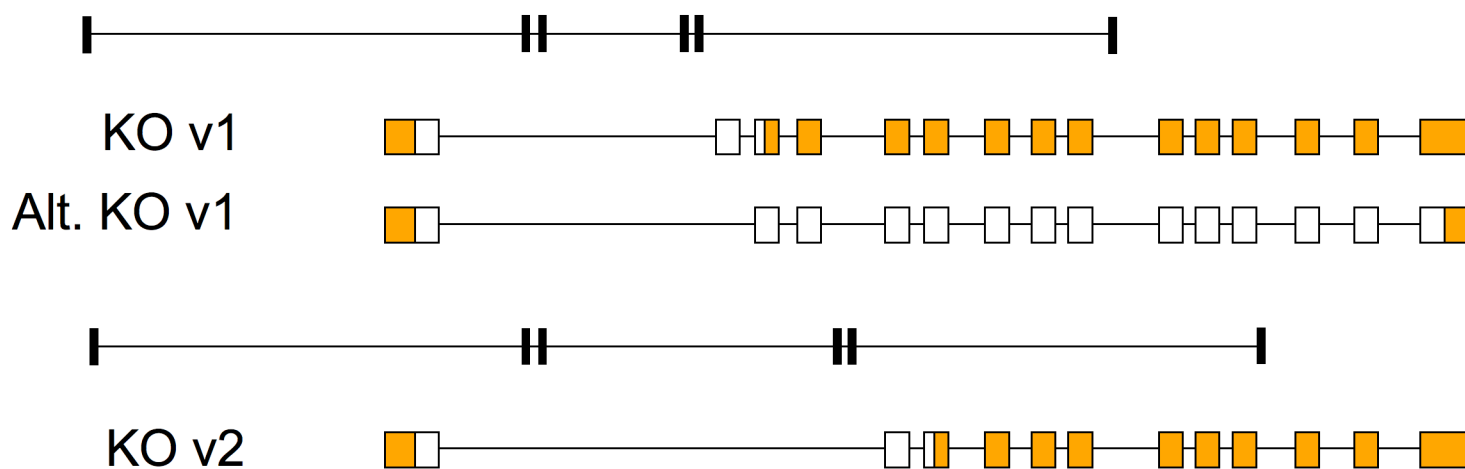
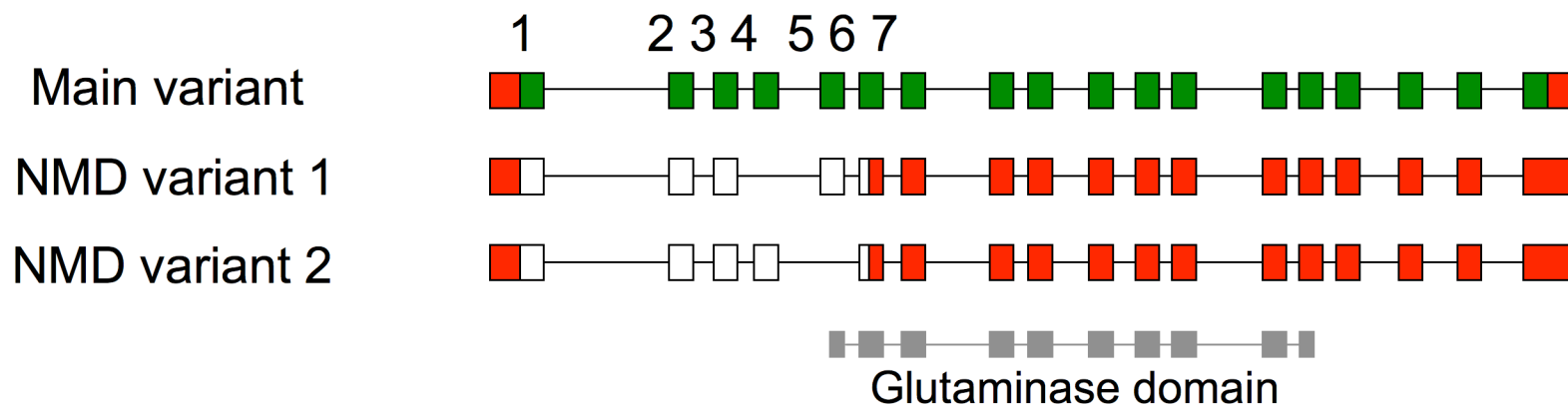
# EUCOMM - vector design



# EUCOMM - vector design



# EUCOMM - vector design





# VEGA

# VEGA (Vertebrate Genome Annotation)

- Based on Ensembl schema
- Contains only manual annotation (eg. HAVANA, WU, MIT, etc)
- Updated releases every 2 months
- Displays knockout (KO) transcripts for EUCOMM/KOMP

## Browse a genome



***Homo sapiens*** [30-03-2009]  
browse | Ensembl



***Mus musculus*** [30-03-2009]  
browse | Ensembl



***Danio rerio*** [30-03-2009]  
browse | Ensembl



***Gorilla gorilla*** [30-03-2009]  
browse | Ensembl



***Macropus eugenii*** [30-03-2009]  
browse



***Sus scrofa*** [16-05-2007]  
browse | Ensembl



***Canis familiaris*** [14-02-2005]  
browse | Ensembl

<http://vega.sanger.ac.uk>

# VEGA

Vega Mouse ContigView

Search Vega Mouse:    
 e.g. AL590386.4, AL450321.15.1.20917

Vega v35 Mar 2009

[HOME](#) [SITEMAP](#) [HELP](#)

Your Vega

- [Login or Register](#)
- [About User Accounts](#)

Chromosome 10  
12,912,012 - 13,605,550

- [View of Chromosome 10](#)
- [Graphical view](#)
- [Graphical overview](#)

Export data

- [Export from region...](#)

Use Vega to...

- [BLAST / SSAHA](#)
- [Text search](#)
- [Export data](#)
- [Download data](#)

Information

- [Site Map \(Mouse\)](#)
- [What's New](#)
- [Information](#)
- [Acknowledgements](#)

Other links

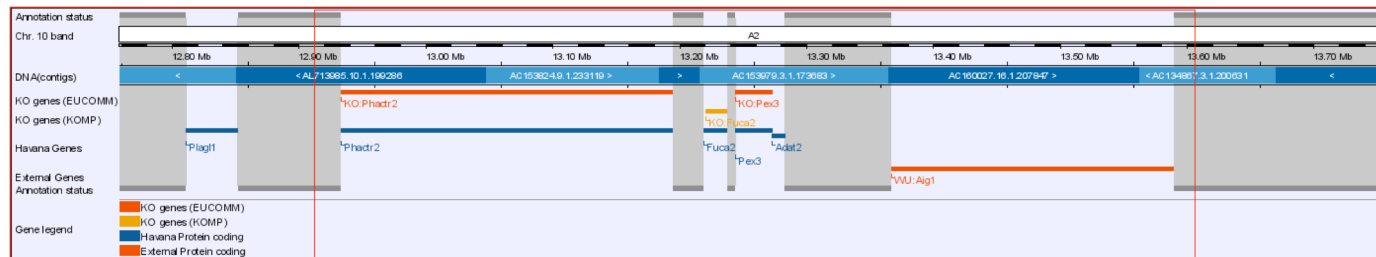
- [Home](#)
- [Human](#)
- [Mouse](#)
- [Zebrafish](#)
- [Gorilla](#)
- [Wallaby](#)
- [Pig](#)
- [Dog](#)
- [Ensembl Mouse](#)



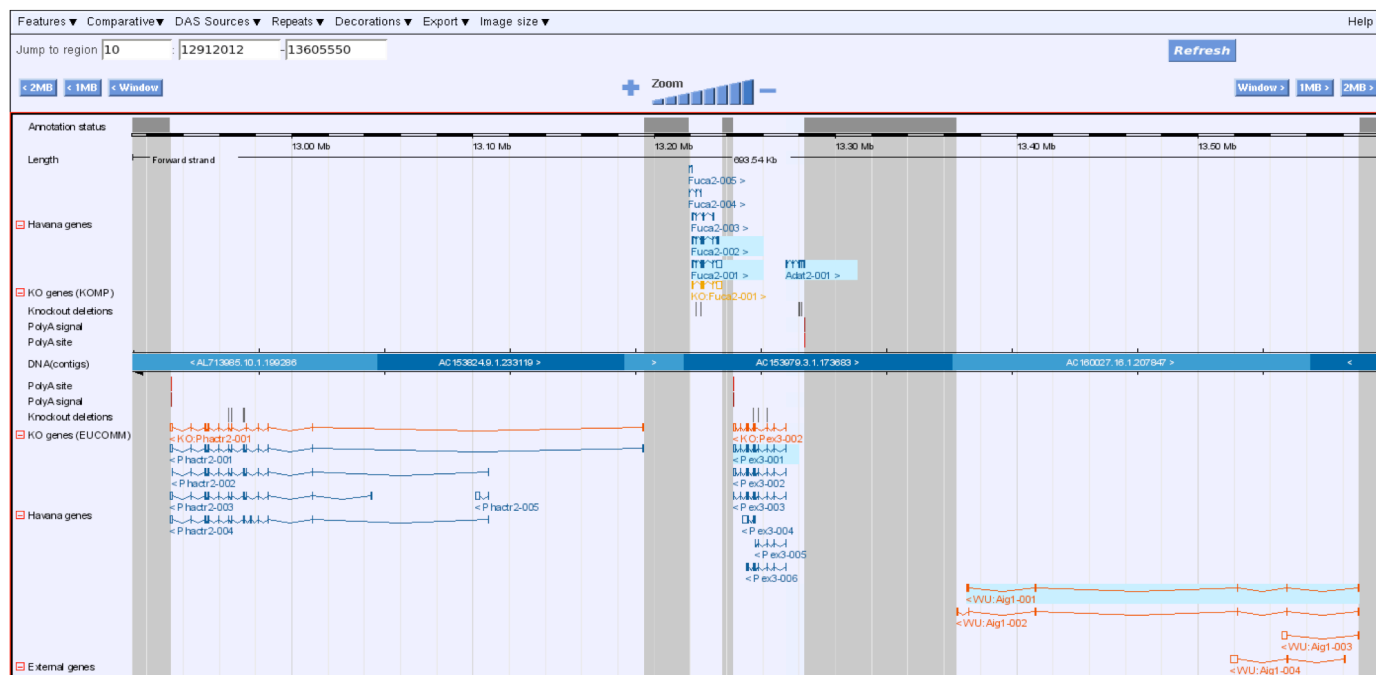
**Chromosome 10**



**Overview**



**Detailed view**



Nature Precedings : doi:10.1038/npre.2009.3181.1 : Posted 27 Apr 2009

# VEGA

## Vega Mouse ContigView

Vega v35 Mar 2009

Your Vega

- ★ Login or Register
- ★ About User Accounts

Chromosome 10  
12,912,012 - 13,605,550

- ★ View of Chromosome 10
- ★ Graphical view
- ★ Graphical overview

Export data

- ★ Export from region...

Use Vega to...

- ★ BLAST / SSAHA
- ★ Text search
- ★ Export data
- ★ Download data

Information

- Site Map (Mouse)
- What's New
- Information
- Acknowledgements

Other links

- ★ Home
- ★ Human
- ★ Mouse
- ★ Zebrafish
- ★ Gorilla
- ★ Wallaby
- ★ Pig
- ★ Dog
- ★ Ensembl Mouse



Chromosome 10

Overview

Detailed view

Chr. 10

- Annotation status
- Chr. 10 band
- DNA(contigs)
- KO genes (EUCOMM)
- KO genes (KOMP)
- Havana Genes
- External Genes
- Annotation status
- Gene legend

Features

Jump to region

< 2MB < 1MB

Annotation status

Length

Havana genes

KO genes (KOMP)

Knockout deletion

PolyA signal

PolyA site

DNA(contigs)

PolyA site

PolyA signal

Knockout deletion

KO genes (EUCOMM)

Havana genes

External genes

Features Comparative DAS Sources Repeats Decorations Export Image size

Jump to region 10 : 12912012 - 13605550

< 2MB < 1MB < Window



Annotation status

Length

Havana genes

KO genes (KOMP)

Knockout deletions

PolyA signal

PolyA site

DNA(contigs)

PolyA site

PolyA signal

Knockout deletions

KO genes (EUCOMM)

Havana genes

External genes

Havana genes

Vega\_transcript: Pex3-001

Transcript class: Known protein coding

Gene type: Known protein coding

Author: Havana

CCDS: CCDS23702.1

Gene: OTTMUSG00000020905

Transcr: OTTMUST00000049414

Peptide: OTTMUSP00000022888

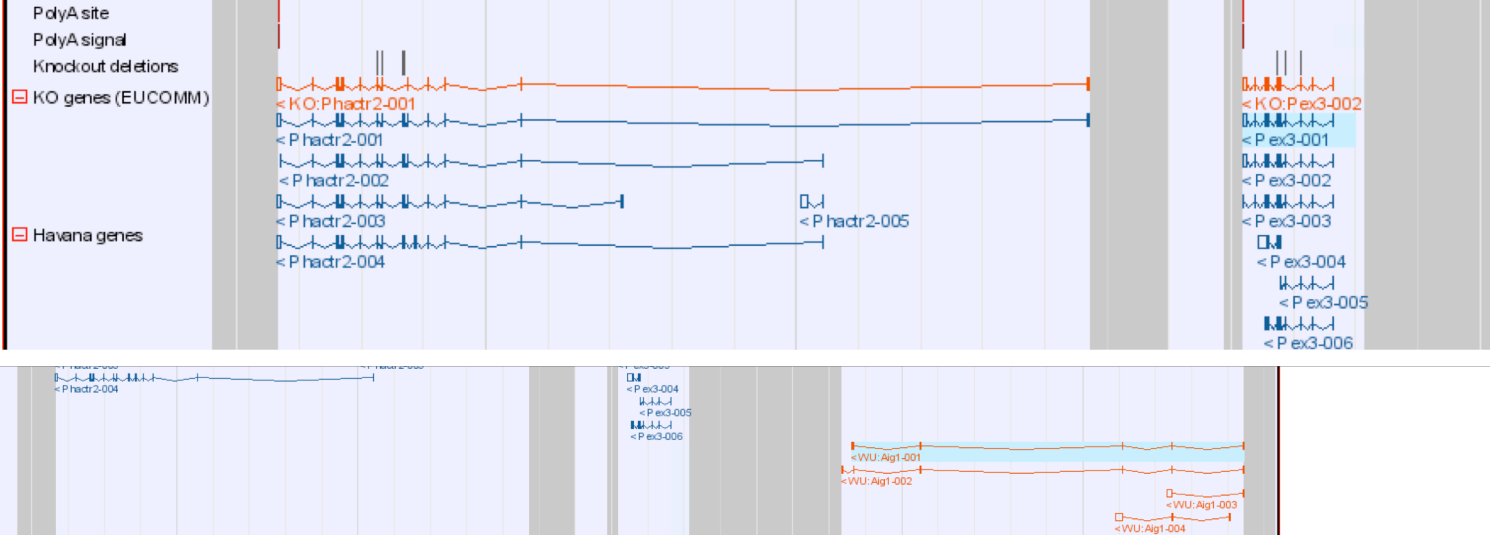
Exon: OTTMUST00000049414

Supporting evidence

Export cDNA

Export Peptide

< AL713685.10.1.189286 AC15362.4.9.1.233119 > > AC153679.3.1.173683 >



Nature Precedings : doi:10.1038/npre.2009.3181.1 : Posted 27 Apr 2009

# Results

# Results - current status

## Annotation

Annotated protein coding genes	12,585
Genes requiring annotation (up to)	4,183
<b>Total</b>	<b>16,768</b>

## Vectors

Awaiting design	811
Design completed	7,010
Vectors in construction	3,927
Complete Targeting Vectors	3,004
ES Cells	981

# Results - High Throughput Gene Targeting

Nature Precedings : doi:10.1038/npre.2009.3181.1 : Posted 27 Apr 2009

Database of targeted mouse mutant resources generated by the WTSI for the EUCOMM, KOMP and NORCOMM projects

<http://www.sanger.ac.uk/htgt/>

<http://www.eucomm.org>

## High Throughput Gene Targeting

Search for a gene by Marker Symbol, MGI ID, Ensembl ID, Vega ID, Keyword or EPD clone  
e.g. Cbx1, MGI:105369, ENSMUSG00000018666, OTTMUSG00000001636, zinc finger or EPD0027\_1\_F05

[BioMart \(Advanced Query Interface\)](#)

[Browse for Designs](#)

### Pipeline Summary

Genes selected for targeted conditional knockout by pipeline stage / project:

Stage	EUCOMM	KOMP	MGP	NorCOMM
Mice - Available for distribution	0	0	0	0
Mice - Genotype confirmed	148	89	124	0
Mice - Germline transmission	67	23	21	0
Mice - Microinjection unsuccessful	0	0	0	0
Mice - Microinjection in progress	217	55	43	0
ES Cells - Targeting Confirmed	981 (866 shipped)	1042 (1008 shipped)	83 (223 shipped)	0 (0 shipped)
ES Cells - Targeting Unsuccessful - Project Terminated	2	2	1	0
ES Cells - No QC Positives	312	315	48	0
ES Cells - Electroporation Unsuccessful	233	182	29	0
ES Cells - Electroporation in Progress	210	405	20	0
Vector - DNA Not Suitable for Electroporation	288	375	15	0
Vector Complete	546	715	62	0
Vector - Initial Attempt Unsuccessful	2212	1751	141	0
Vector Unsuccessful - Alternate Design in Progress	105	7	19	0
Vector Unsuccessful - Project Terminated	4	2	3	0
Vector Construction in Progress	1610	1495	27	627
Design Completed	75	160	3	88
Design Not Possible	7	893	27	0
VEGA Annotation Requested	765	149	0	47
Alternate Design Requested	221	5	9	1
Design Requested	46	1429	0	435
Withdrawn From Pipeline	324	428	21	29
Transferred to NorCOMM	216	0	0	0
Transferred to KOMP	0	58	0	0
On Hold	3	8	0	108

# Conclusions

- Detailed manual annotation is vital for ensuring a high quality conditional knockout
  - Benefits include;
    - Accurate identification of exon splice variation, including NMD
    - Annotation of UTRs
    - Analysis of possible reinitiation events in KO transcripts
    - Ensures targeting of all coding variants
- Annotation contributed by mouse KO projects available to wider research community



# Acknowledgements

## HAVANA

Tim Hubbard

Jen Harrow

Catherine Snow

Clara Amid

Denise Carvalho-Silva

Marie-Marthe Suner

- Everyone else

## Otterlace and Zmap

James Gilbert

Ed Griffiths

## VEGA

Stephen Trevanion

## ES Cell Mutagenesis

Bill Skarnes

Manousos Koutsourakis (Vector Construction)

Alejandro Mujica

Vivek Iyer (Informatics)

David Jackson

Daniel Klose

## ENSEMBL

Funded by

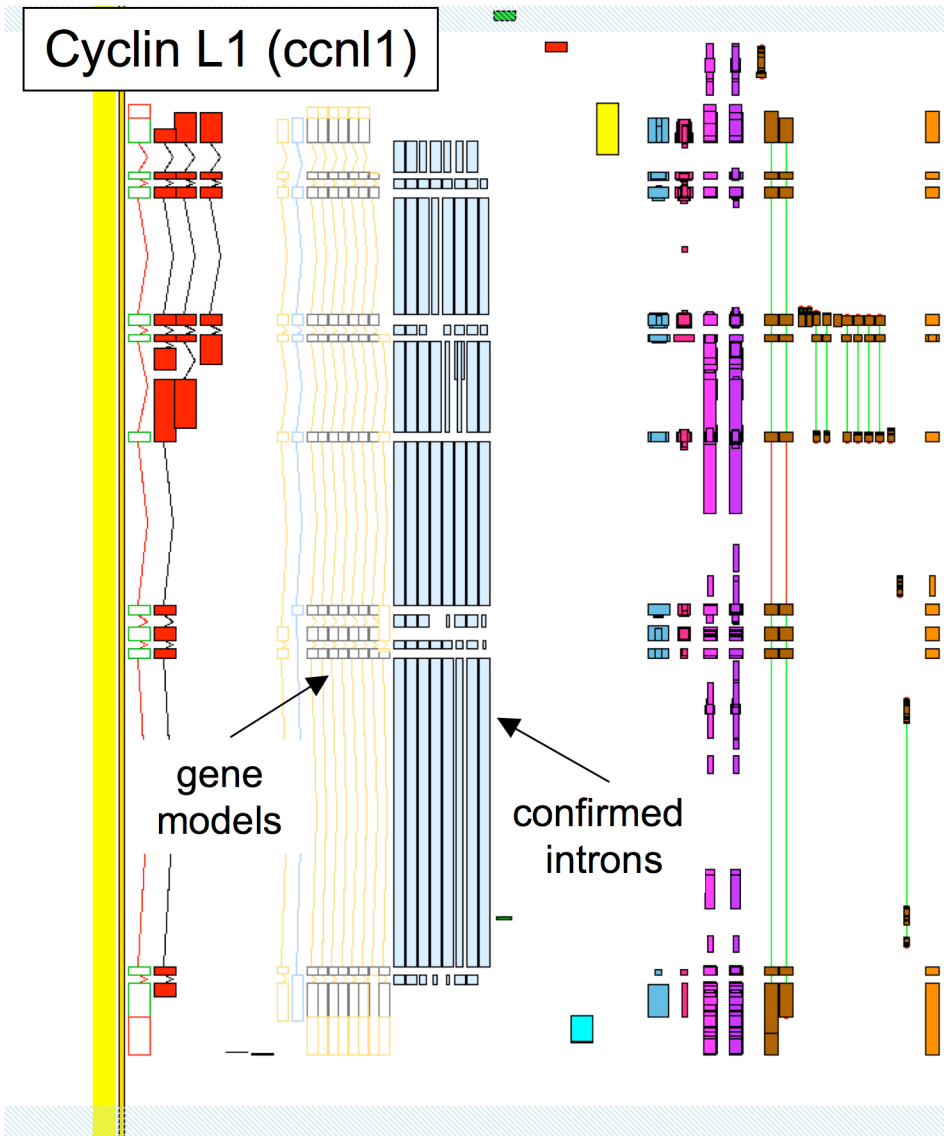
**welc**ome trust



SIXTH FRAMEWORK PROGRAMME

# RNA-Seq data

## Cyclin L1 (ccn1)



- Solexa data imported from Ensembl as;
  - gene models
  - confirmed introns
- Data currently only available for human and zebrafish
- Example from zebrafish - Cyclin L1 (ccn1)