

TOUCAN Tutorial

Muscle-specific genes

1. Sequence retrieval

File or ids

Select a file (1 column of identifiers) or type comma-separated ids

TIP: Ids can be Affy ids

Subsequence list

Sequence set

Feature list

New GeneList from identifiers

ID list or ID-file: D:\SAE\muscle\genes.txt Browse

ID file: Human HUGO

5' Upstream CDS bp before: 10000 bp within: 200

Add to current list Add (multiple) orthologs:

- ggallus
- mmusculus
- ptroglodytes
- rnorvegicus
- tnigroviridis

OK Cancel

Gene Info

Left mouse click on a gene to display info.

Feature Info

Left mouse click on a feature to display info.

Status


Supporting Ensembl release: 26

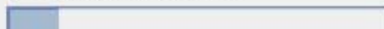
Subsequence list

Sequence set

Feature list

Progress... [Close]

 Retrieving sequences from Ensembl
Completed 2 out of 15.



Cancel

Gene Info
Left mouse click on a gene to display info.

Feature Info
Left mouse click on a feature to display info.

Status
Driver loaded (Ensembl - Mus_musculus)

Subsequence list

Feature list

Sequence set

Warning [X]

 Retrieval and/or parsing problems occurred for:
- homolog for des (ENSMUSG00000026208) >> Gene not found

OK

Gene Info
Left mouse click on a gene to display info.

Feature Info
Left mouse click on a feature to display info.

Status
Sequence retrieval done

Subsequence list

Sequence set

CpG Island - Human Ctrl-G

RevCompl Negatives

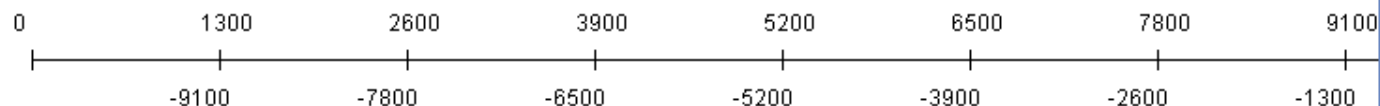
Cut in pieces

RepeatMasker

2. Reverse complement Genes on negative strand

Feature list

- CDS
- FivePrimeUTR
- ThreePrimeUTR
- exon
- gene



ENS00000104879

ENSMUS00000030399

ENS00000129991

ENS00000132438

ENSMUS00000045103

ENS00000197616

ENSMUS00000040752

Gene Info

Left mouse click on a gene to display info.

Feature Info

Left mouse click on a feature to display info.

Status

Sequence retrieval done

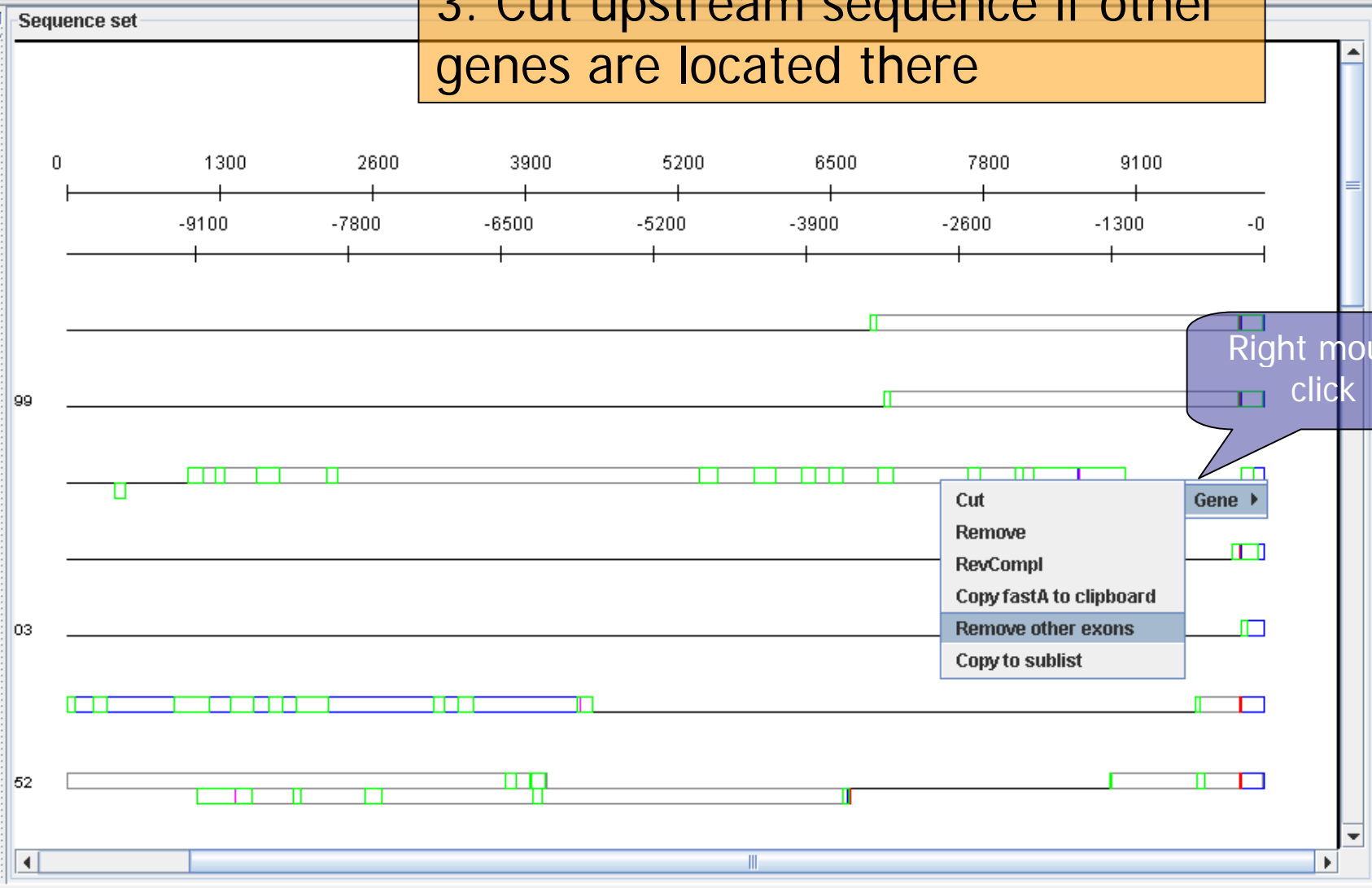
3. Cut upstream sequence if other genes are located there

Subsequence list

Sequence set

Feature list

- CDS
- FivePrimeUTR
- ThreePrimeUT
- exon
- gene



Right mouse click

- Cut
- Remove
- RevCompl
- Copy fastA to clipboard
- Remove other exons
- Copy to sublist

Gene Info

No ortho-family relations known for this gene.

Feature Info

Status

saved D:\SAE\musclegenes.embl

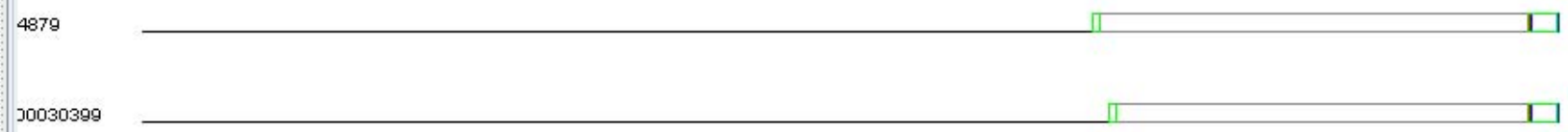
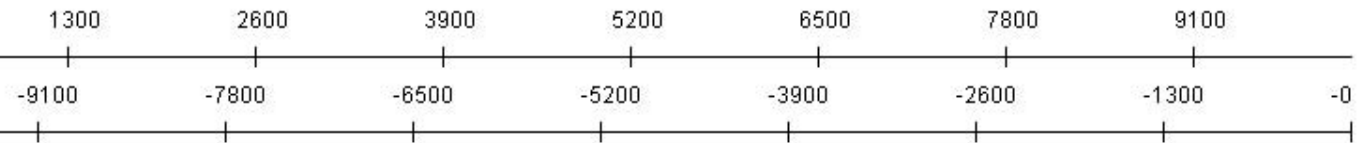
4. Align all human-mouse pairs with AVID

Choose All pairs from the Alignment menu

Subsequence list | Sequence set

Feature list

- CDS
- FivePrimeUTR
- ThreePrimeUT
- exon
- gene



Auto-align Web Service

Service URL * <http://www.esat.kuleuven.ac.be/axis/ToucanSOAPService.jws>

Choose one or more pairs

- ENSG00000104879-ENSMUSG00000030399
- ENSG00000132438-ENSMUSG00000045103
- ENSG00000197616-ENSMUSG00000040752
- ENSG00000181072-ENSMUSG00000045613
- ENSG00000175206-ENSMUSG00000041616

Algorithm: AVID

P3: Minimal % identity (Integer): 75

P4: Window length in bp (Integer): 100

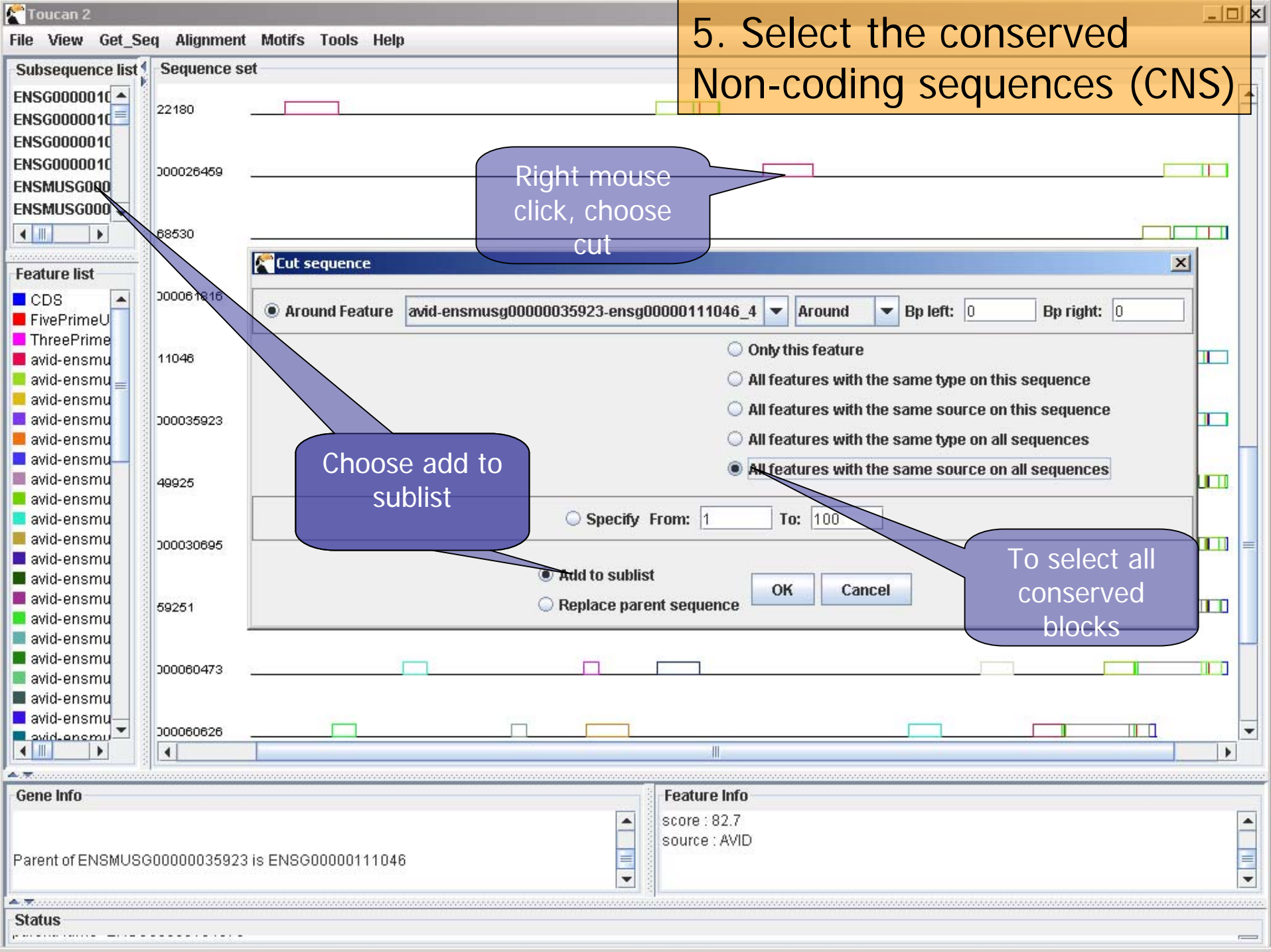
OK Cancel

Gene Info

Parent of ENSMUSG00000031972 is ENSG00000143632

Feature Info

Status



5. Select the conserved Non-coding sequences (CNS)

Right mouse click, choose cut

Choose add to sublist

To select all conserved blocks

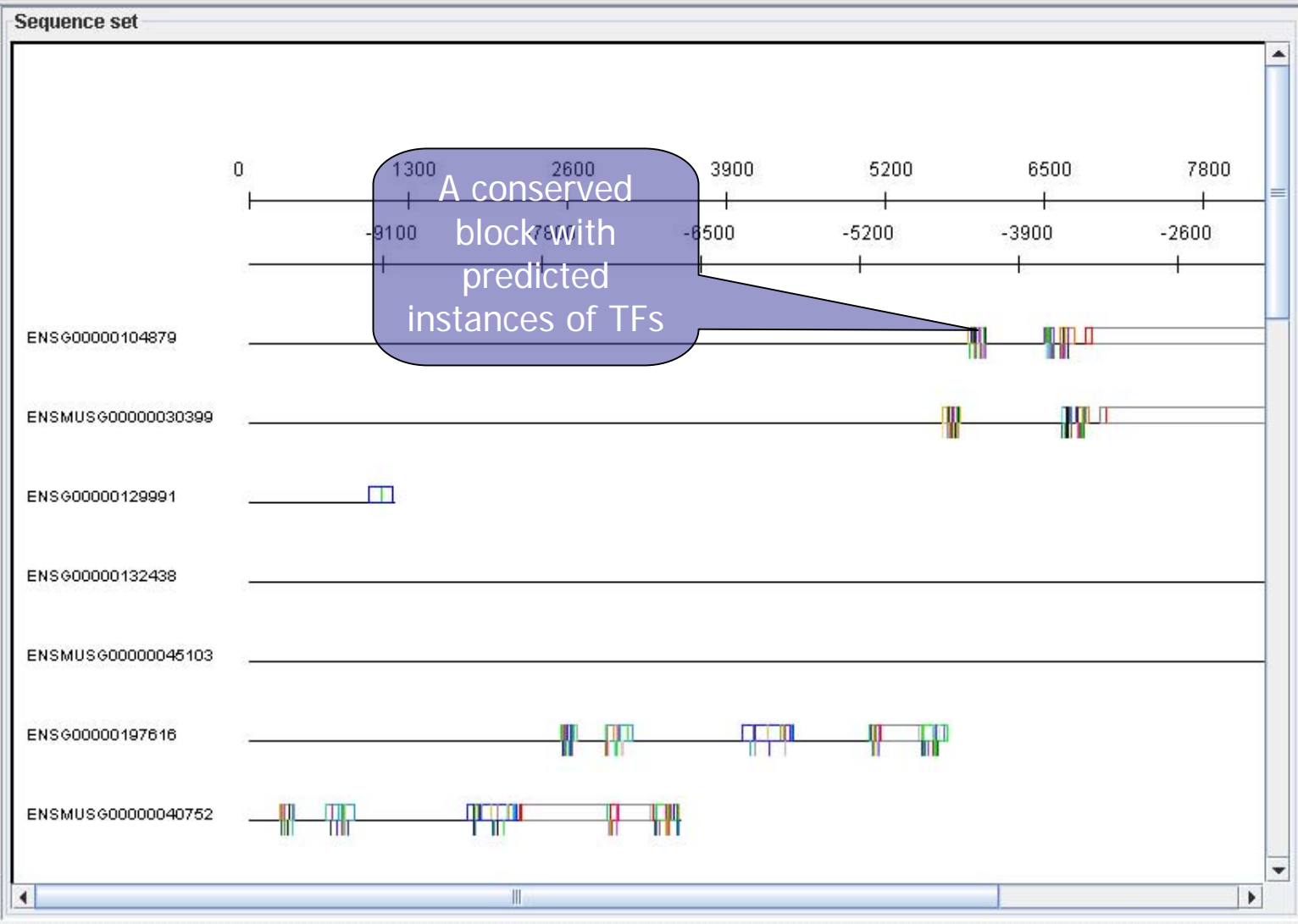
Gene Info
Parent of ENSMUSG00000035923 is ENSG00000111046

Feature Info
score : 82.7
source : AVID

Status

- Subsequence list**
- ENSG00000104879_5884_6
 - ENSG00000104879_9994_1
 - ENSG00000104879_6639_6
 - ENSG00000104879_6514_6
 - ENSMUSG00000030399_66
 - ENSMUSG00000030399_56

- Feature list**
- CDS
 - FivePrimeUTR
 - M00001-V\$MYOD_01
 - M00002-V\$E47_01
 - M00003-V\$VMYB_01
 - M00004-V\$CMYB_01
 - M00005-V\$AP4_01
 - M00006-V\$MEF2_01
 - M00007-V\$ELK1_01
 - M00008-V\$SP1_01
 - M00011-V\$EV1_06
 - M00017-V\$ATF_01
 - M00024-V\$E2F_01
 - M00025-V\$ELK1_02
 - M00026-V\$RSRFC4_01
 - M00032-V\$CETS1P54_01
 - M00033-V\$P300_01
 - M00035-V\$VMAF_01
 - M00037-V\$NFE2_01
 - M00039-V\$CREB_01
 - M00042-V\$SOX5_01
 - M00045-V\$E4BP4_01
 - M00050-V\$E2F_02



Gene Info

ID ENSG00000197616

No ortho-family relations known for this gene.

Feature Info

type : M00960-V\$PR_Q2

score : 90.7132

site : CAAGGAACAA

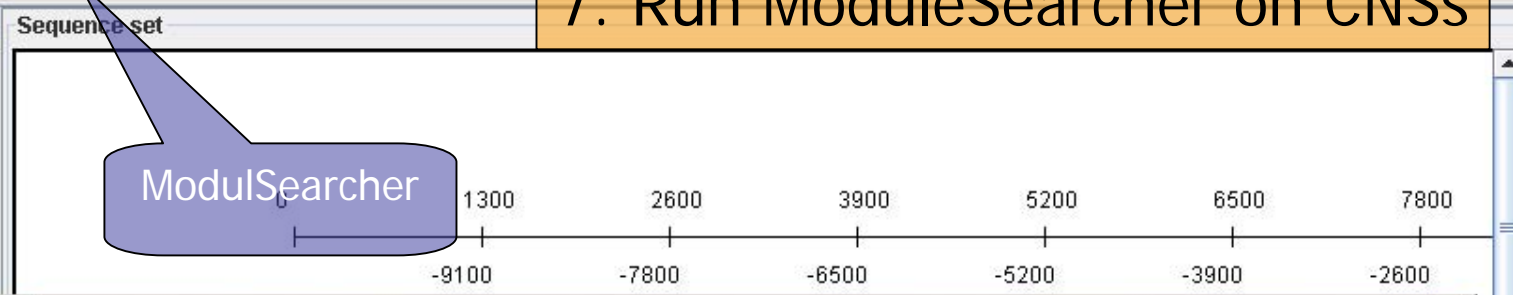
source : MotifScanner

Status

7. Run ModuleSearcher on CNSs

Subsequence list

- ENSG00000104879_5884_6
- ENSG00000104879_9994_1
- ENSG00000104879_6639_6
- ENSG00000104879_6514_6
- ENSMUSG00000030399_66
- ENSMUSG00000030399_56



Feature list

- CDS
- FivePrimeUTR
- M00001-V\$MYOD_01
- M00002-V\$E47_01
- M00003-V\$VMYB_01
- M00004-V\$CMYB_01
- M00005-V\$AP4_01
- M00006-V\$MEF2_01
- M00007-V\$ELK1_01
- M00008-V\$SP1_01
- M00011-V\$EV1_06
- M00017-V\$ATF_01
- M00024-V\$E2F_01
- M00025-V\$ELK1_02
- M00026-V\$RSRFC4_01
- M00032-V\$CETS1P54_01
- M00033-V\$P300_01
- M00035-V\$VMAF_01
- M00037-V\$NFE2_01
- M00039-V\$CREB_01
- M00042-V\$SOX5_01
- M00045-V\$E4BP4_01
- M00050-V\$E2F_02

ModuleSearcher Web Service

Service URL * <http://www.esat.kuleuven.ac.be/axis/ToucanSOAPService.jws>

Feature source (String) * MotifScanner

Algorithm Genetic Algorithm

Nr elements (Integer) * 5

Size (Integer) * 200

Allow overlap

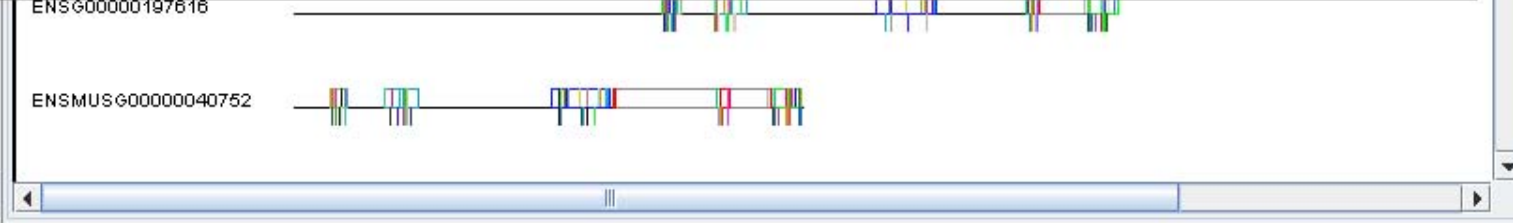
Use penalisation of incomplete instances

Exclude matrices (comma separated) (String)

Nr of top scoring modules to return (Integer) * 20

GET

OK Cancel



Gene Info

ID ENSG00000197616

No ortho-family relations known for this gene.

Feature Info

type : M00960-V\$PR_Q2

score : 90.7132

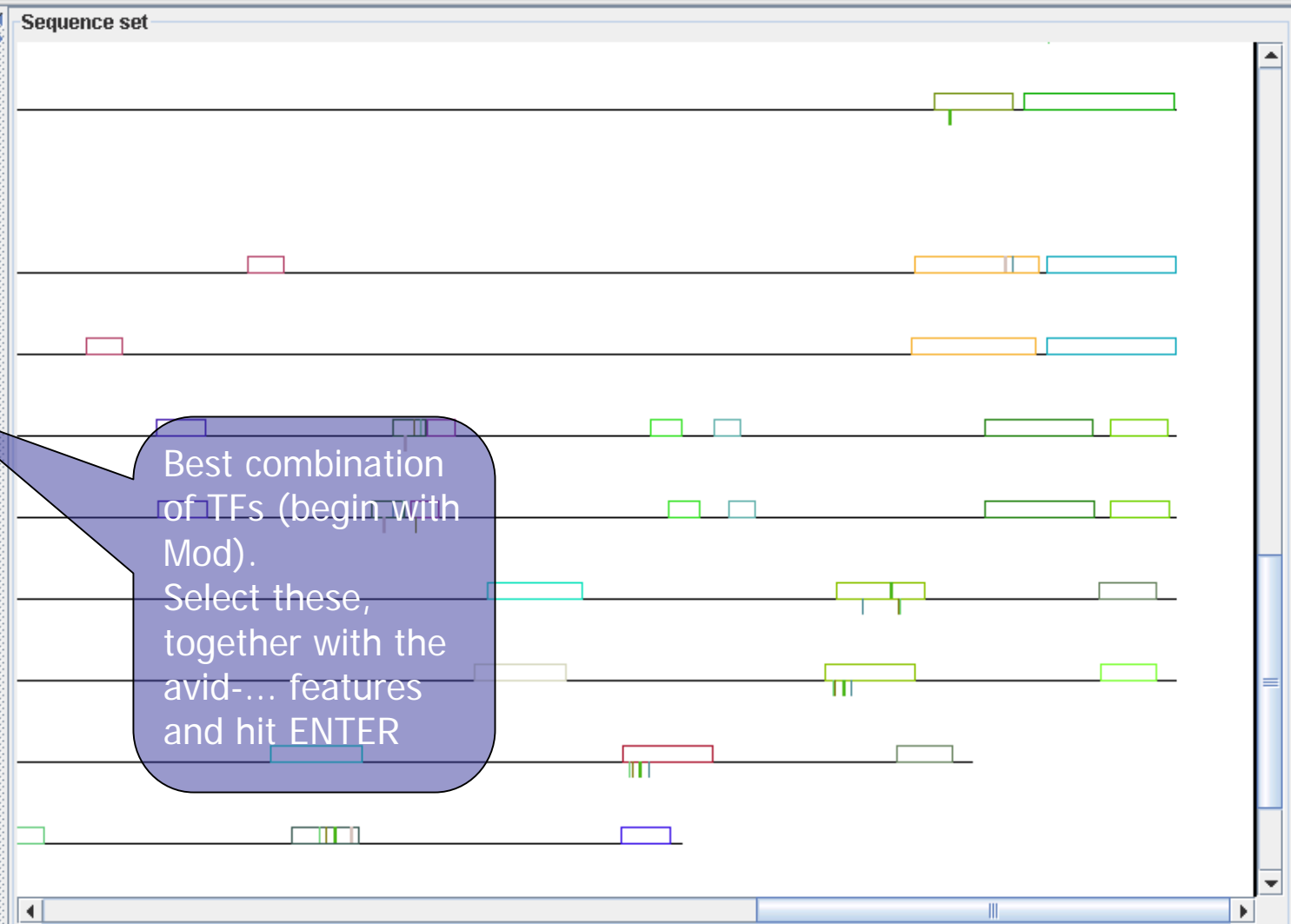
site : CAAGGAACAA

source : MotifScanner

Status

- Subsequence list**
- ENSG00000104879_5884_6034
 - ENSG00000104879_9994_10202
 - ENSG00000104879_6639_6757
 - ENSG00000104879_6514_6603
 - ENSMUSG00000030399_6651_67
 - ENSMUSG00000030399_5674_58

- Feature list**
- Mod5M00986-V\$CHCH_01
 - Mod6M00649-V\$MAZ_Q6
 - Mod6M00720-V\$CACBINDINGPR
 - Mod6M00922-V\$SRF_Q5_01
 - Mod6M00941-V\$MEF2_Q6_01
 - Mod6M00986-V\$CHCH_01
 - Mod7M00215-V\$SRF_C
 - Mod7M00646-V\$LFA1_Q6
 - Mod7M00649-V\$MAZ_Q6
 - Mod7M00720-V\$CACBINDINGPR
 - Mod7M00980-V\$TBP_Q6
 - Mod8M00405-V\$MMEF2_Q6
 - Mod8M00646-V\$LFA1_Q6
 - Mod8M00649-V\$MAZ_Q6
 - Mod8M00810-V\$SRF_Q4
 - Mod8M00933-V\$SP1_Q2_01
 - Mod9M00215-V\$SRF_C
 - Mod9M00649-V\$MAZ_Q6
 - Mod9M00915-V\$AP2_Q6_01
 - Mod9M00922-V\$SRF_Q5_01
 - Mod9M00931-V\$SP1_Q6_01
 - avid-ensmusg00000026459-er



Best combination of TFs (begin with Mod).
 Select these, together with the avid-... features and hit ENTER

Gene Info

ID ENSG00000197616

No ortho-family relations known for this gene.

Feature Info

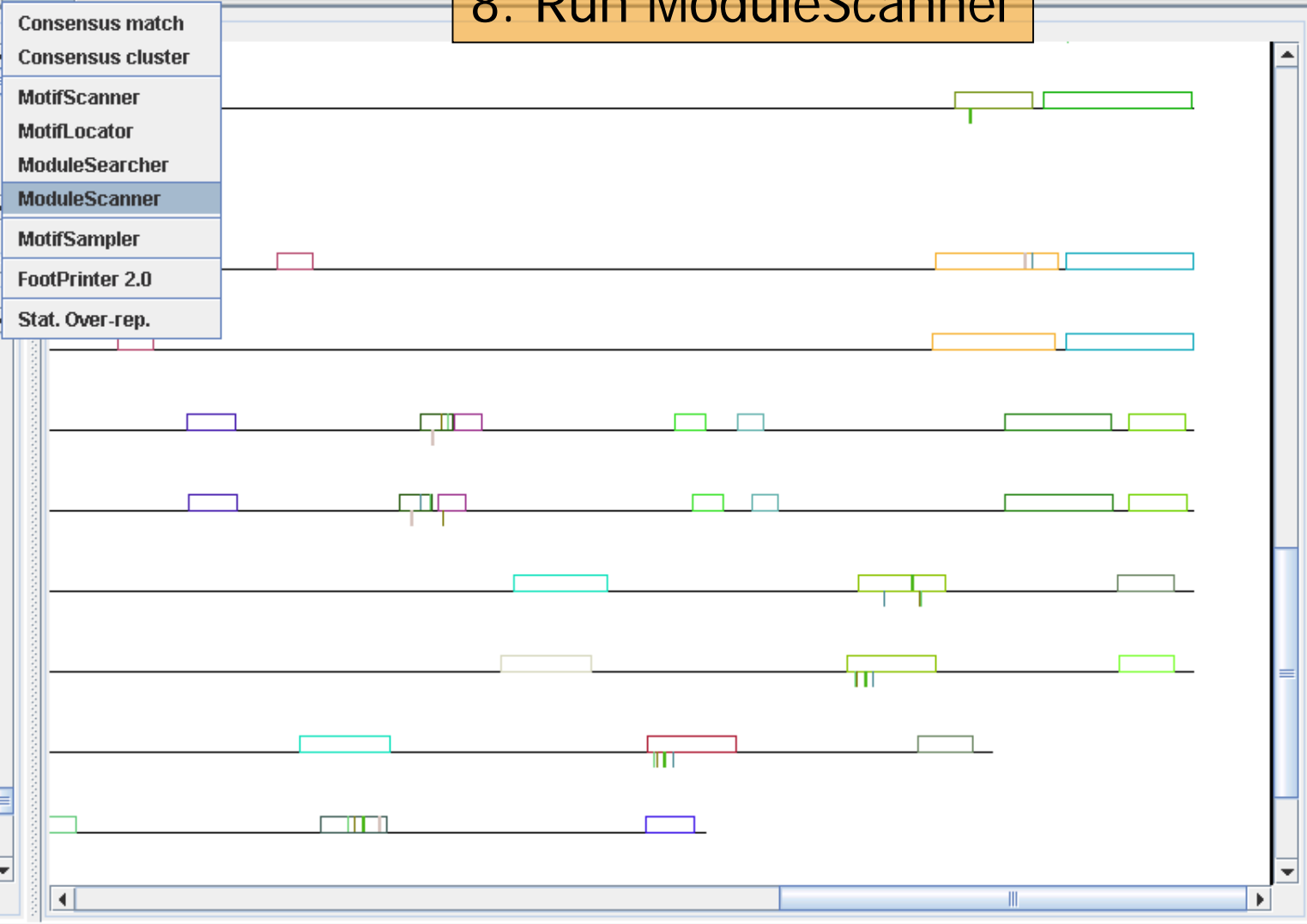
type : M00960-V\$PR_Q2
 score : 90.7132
 site : CAAGGAACAA
 source : MotifScanner

Status

8. Run ModuleScanner

- Subsequence list**
- ENSG00000104879_5884_6034
 - ENSG00000104879_9994_10202
 - ENSG00000104879_6639_6757
 - ENSG00000104879_6514_6603
 - ENSMUSG00000030399_6651_6757
 - ENSMUSG00000030399_5674_5782

- Feature list**
- Mod5M00986-V\$CHCH_Q1
 - Mod6M00649-V\$MAZ_Q6
 - Mod6M00720-V\$CACBINDINGPR_Q6
 - Mod6M00922-V\$SRF_Q5_Q1
 - Mod6M00941-V\$MEF2_Q6_Q1
 - Mod6M00986-V\$CHCH_Q1
 - Mod7M00215-V\$SRF_C
 - Mod7M00646-V\$LFA1_Q6
 - Mod7M00649-V\$MAZ_Q6
 - Mod7M00720-V\$CACBINDINGPR_Q6
 - Mod7M00980-V\$TBP_Q6
 - Mod8M00405-V\$MMEF2_Q6
 - Mod8M00646-V\$LFA1_Q6
 - Mod8M00649-V\$MAZ_Q6
 - Mod8M00810-V\$SRF_Q4
 - Mod8M00933-V\$SP1_Q2_Q1
 - Mod9M00215-V\$SRF_C
 - Mod9M00649-V\$MAZ_Q6
 - Mod9M00915-V\$AP2_Q6_Q1
 - Mod9M00922-V\$SRF_Q5_Q1
 - Mod9M00931-V\$SP1_Q6_Q1
 - avid-ensmusg00000026459-er



Gene Info

ID ENSG00000197616

No ortho-family relations known for this gene.

Feature Info

type : M00960-V\$PR_Q2

score : 90.7132

site : CAAGGAACAA

source : MotifScanner

Status

- Subsequence list**
- ENSG00000104879_5884_6034
 - ENSG00000104879_9994_10202
 - ENSG00000104879_6639_6757
 - ENSG00000104879_6514_6603
 - ENSMUSG00000030399_6651_6757
 - ENSMUSG00000030399_5674_5774



ModuleScanner Web Service

Service URL * <http://www.esat.kuleuven.ac.be/axis/ToucanSOAPService.jws>

Subject * (String) Press GET and choose a DB

Module * **Select**

Size in base pairs

Top N to return

Overlap *

Penalize *

Choose one of the database values:

- Human-Mouse Conserved non-coding regions, Ensembl r24, Human regions, prior 0.2, Transfac
- Human-Mouse Conserved non-coding regions, Ensembl r24, Human regions, prior 0.2, Transfac
- Human-Zebrafish Conserved non-coding regions, Ensembl r24, Human regions, prior 0.2, Transfac
- Human-Zebrafish Conserved non-coding regions, Ensembl r24, Zebrafish regions, prior 0.2, Transfac
- Human-Mouse Conserved non-coding regions, Ensembl r24, Human regions, prior 0.2, Jaspas
- Human-Fugu Conserved non-coding regions, Ensembl r24, Fugu regions, prior 0.2, Transfac
- Human-Fugu Conserved non-coding regions, Ensembl r24, Human regions, prior 0.2, Jaspas
- Human-Fugu Conserved non-coding regions, Ensembl r24, Fugu regions, prior 0.2, Jaspas
- Human-Mouse Conserved non-coding regions, Ensembl r18, Human regions, prior 0.2, Transfac

Gene Info

ID ENSG00000197616

No ortho-family relations known for this gene.

Feature Info

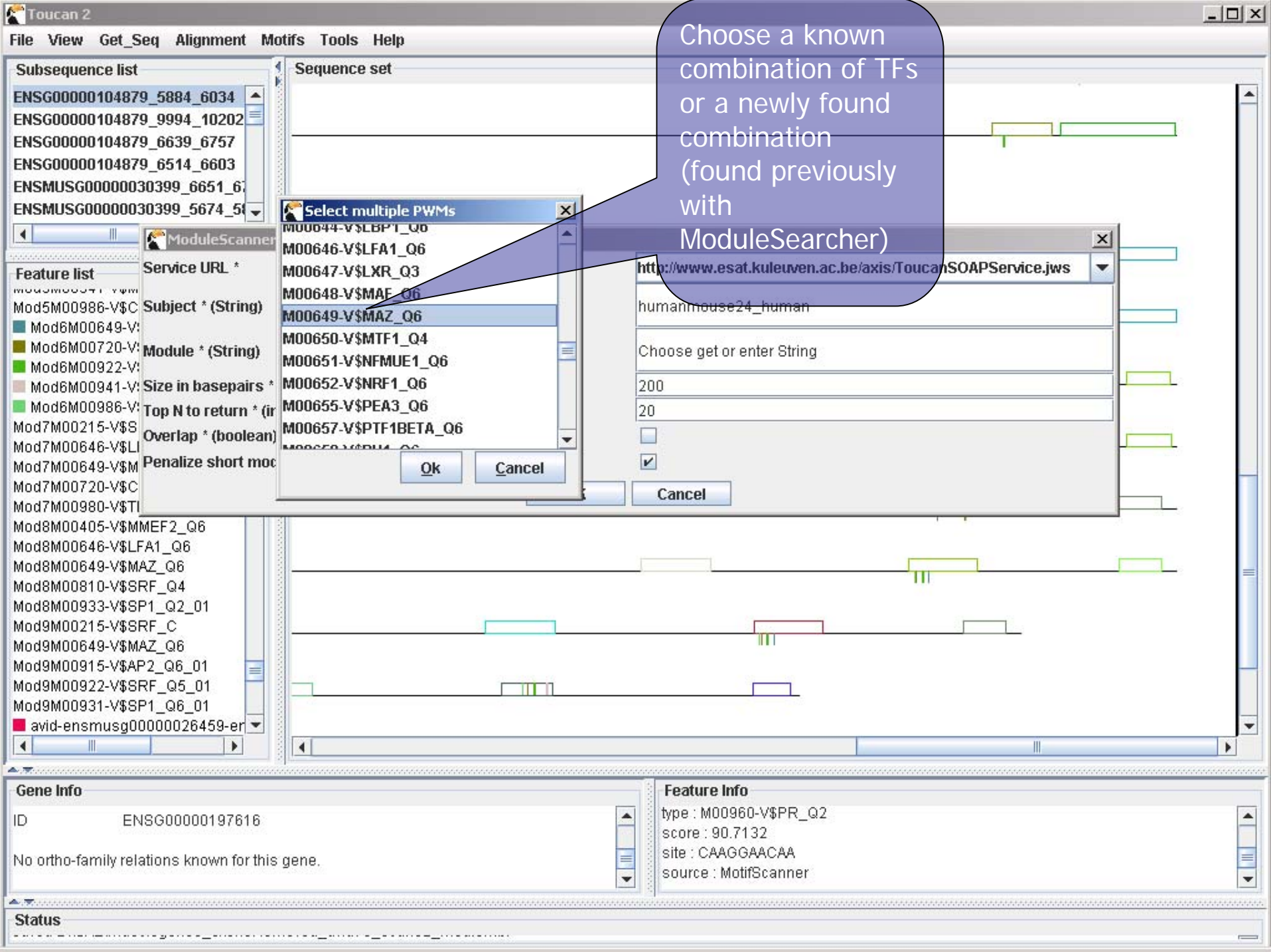
type : M00960-V\$PR_Q2

score : 90.7132

site : CAAGGAACAA

source : MotifScanner

Status



Choose a known combination of TFs or a newly found combination (found previously with ModuleSearcher)

Select multiple PWMs

- M00644-V\$LBPT_Q0
- M00646-V\$LFA1_Q6
- M00647-V\$LXR_Q3
- M00648-V\$MAE_Q6
- M00649-V\$MAZ_Q6
- M00650-V\$MTF1_Q4
- M00651-V\$NFMUE1_Q6
- M00652-V\$NRF1_Q6
- M00655-V\$PEA3_Q6
- M00657-V\$PTF1BETA_Q6
- M00659-V\$PR_Q2

Ok Cancel

ModuleScanner

Service URL *
http://www.esat.kuleuven.ac.be/axis/ToucanSOAPService.jws

Subject * (String)
humanmouse24_human

Module * (String)
Choose get or enter String

Size in basepairs *
200

Top N to return * (ir)
20

Overlap * (boolean)

Penalize short motifs

Cancel

Subsequence list

- ENSG00000104879_5884_6034
- ENSG00000104879_9994_10202
- ENSG00000104879_6639_6757
- ENSG00000104879_6514_6603
- ENSMUSG00000030399_6651_6757
- ENSMUSG00000030399_5674_5774

Feature list

- Mod5M00986-V\$C
- Mod6M00649-V\$
- Mod6M00720-V\$
- Mod6M00922-V\$
- Mod6M00941-V\$
- Mod6M00986-V\$
- Mod7M00215-V\$S
- Mod7M00646-V\$L
- Mod7M00649-V\$M
- Mod7M00720-V\$C
- Mod7M00980-V\$T
- Mod8M00405-V\$MMEF2_Q6
- Mod8M00646-V\$LFA1_Q6
- Mod8M00649-V\$MAZ_Q6
- Mod8M00810-V\$SRF_Q4
- Mod8M00933-V\$SP1_Q2_01
- Mod9M00215-V\$SRF_C
- Mod9M00649-V\$MAZ_Q6
- Mod9M00915-V\$AP2_Q6_01
- Mod9M00922-V\$SRF_Q5_01
- Mod9M00931-V\$SP1_Q6_01
- avid-ensmusg00000026459-er

Gene Info

ID ENSG00000197616

No ortho-family relations known for this gene.

Feature Info

type : M00960-V\$PR_Q2
score : 90.7132
site : CAAGGAACAA
source : MotifScanner

Status

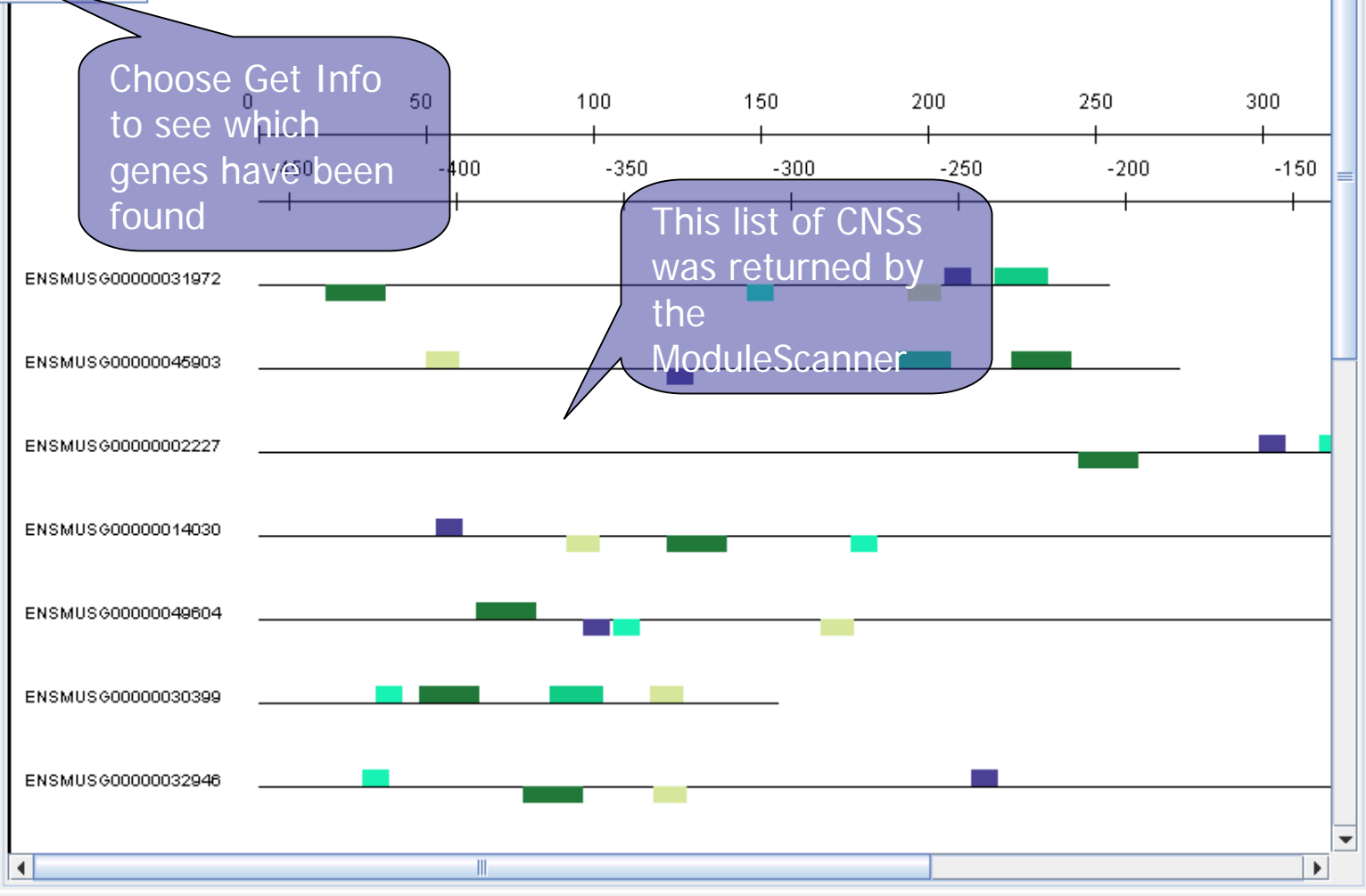
Subsequer

- From Ensembl ▶ New/Add Ctrl-N
- From EMBL ▶ Get Info Ctrl-I

Choose Get Info to see which genes have been found

This list of CNSs was returned by the ModuleScanner

- Feature list**
- ModM00405-V\$MMEF2_Q6
 - ModM00646-V\$LFA1_Q6
 - ModM00649-V\$MAZ_Q6
 - ModM00810-V\$SRF_Q4
 - ModM00933-V\$SP1_Q2_01



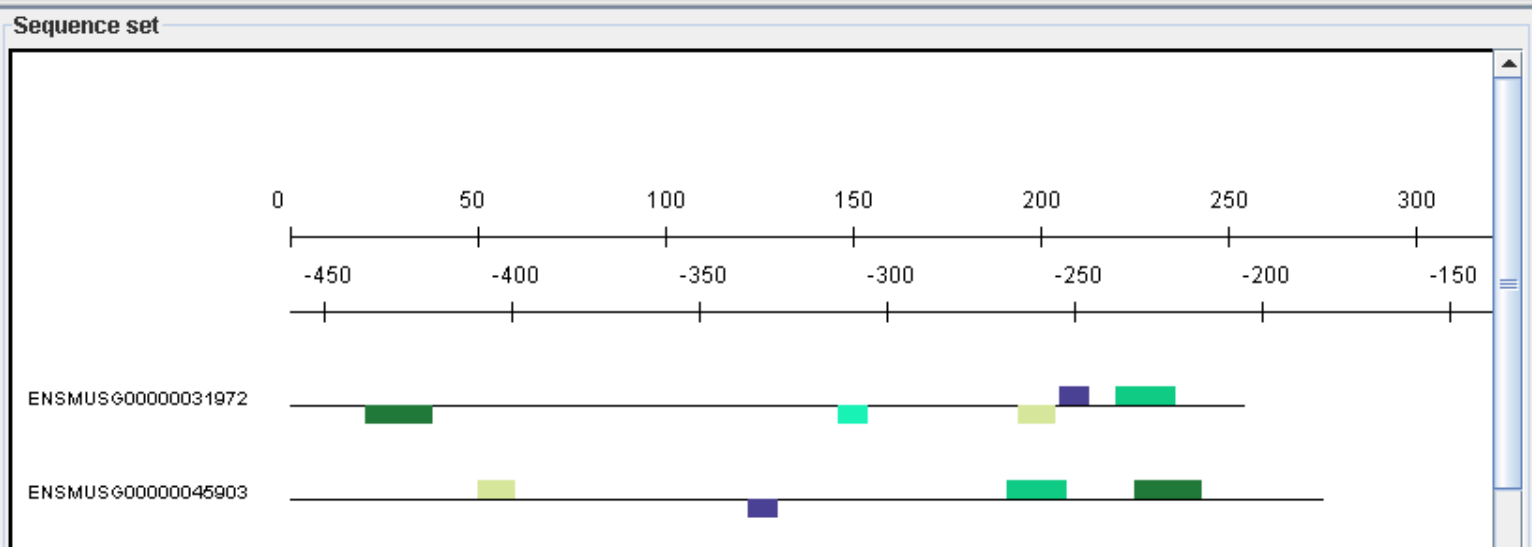
Gene Info
Left mouse click on a gene to display info.

Feature Info
Left mouse click on a feature to display info.

Status
Retrieved info

Subsequence list

- Feature list**
- ModM00405-V\$MMEF2_Q6
 - ModM00646-V\$LFA1_Q6
 - ModM00649-V\$MAZ_Q6
 - ModM00810-V\$SRF_Q4
 - ModM00933-V\$SP1_Q2_01



Gene Information from Ensembl

ocusLink	AFFY_Mu11...	Uniprot/SPT...	EMBL	protein_id	MarkerSym...	Description
464 114...	m12347_s...		J00073 J0...	AAB59619....	Actg2 Acta1...	Actin, alpha skeletal muscle (Alp...
ENSMUSG00...	5872	Q8BGD7	AB049835 ...	BAC19831....	Nxf	bHLH-PAS type transcription fact...
	454		X52574 AK...	CAA36803....	Mov10	Potentail helicase MOV-10. [Sou...
	507		M97013	AAA37325.1	Pax5	Paired box protein Pax-5 (B-cell s...
ENSMUSG00...	408		U57051 B...	AAC52769....	Hoxb13	Homeobox protein Hox-B13. [Sou...
	715		X03233	CAA26979.1	Ckm	Creatine kinase, M chain (EC 2.7...
	395	Q80WC0 Q...	BC051474 ...	AAH51474....	Rasgrp2	RAS, guanyl releasing protein 2; I...
ENSMUSG00...	778	Q8CHS2	BC039633	AAH39633.1	2810451A0...	
ENSMUSG00...	348	AJ001598_...	AB080232 ...	BAC44888....	Viaat	Vesicular inhibitory amino acid tra...
	071	m82874_s...	M97506 M8...	AAA37801....	Nhlh1	Helix-loop-helix protein 1 (HEN1)...
	0935	Q9ESJ5	AF099933	AAG31020.1	Grid2ip	glutamate receptor, ionotropic, de...
	260	Q8BGU8 Q...	AK048800 ...	BAC33461....	Ing1l	inhibitor of growth family, membe...
	aa241085_...					
	878	AB006103_...	AB006103 ...	BAA28284....	Arx	Homeobox protein ARX (Aristales...
		Q9D6E4	AK013779	BAB28991.1	2900074C...	
	816	d00466_f_at	D00466 M1...	BAA00361....	Apoe	Apolipoprotein E precursor (Apo-1...
	816	J05154_s_at	J05154 X5...	AAA39419....	Lcat	Phosphatidylcholine-sterol acyltra...
	484		S79554 S7...	AAC60711....	Hsd11b2	Corticosteroid 11-beta-dehydroge...

Gene Info

Left mouse click on a gene to display info.

Status

Retrieved info