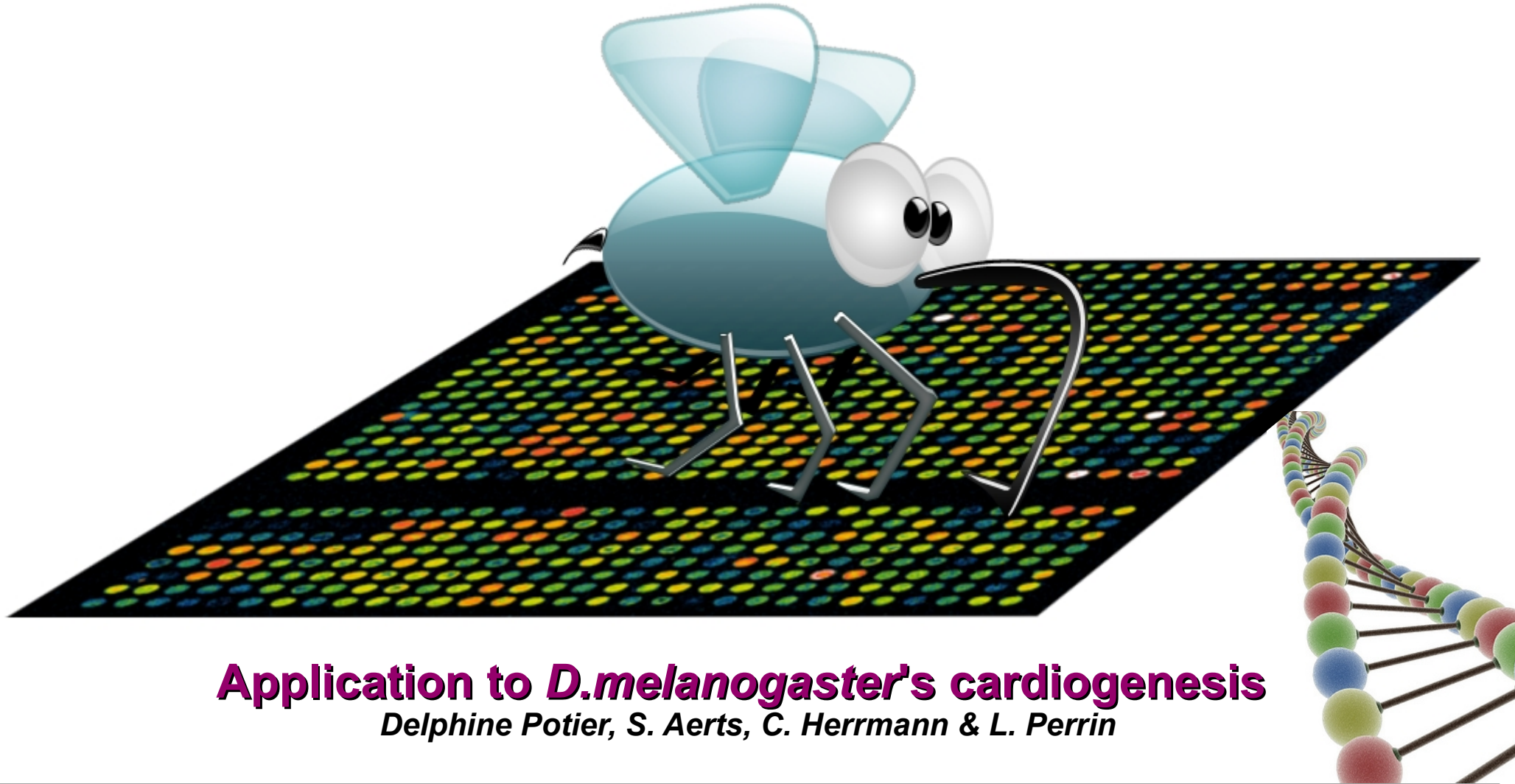


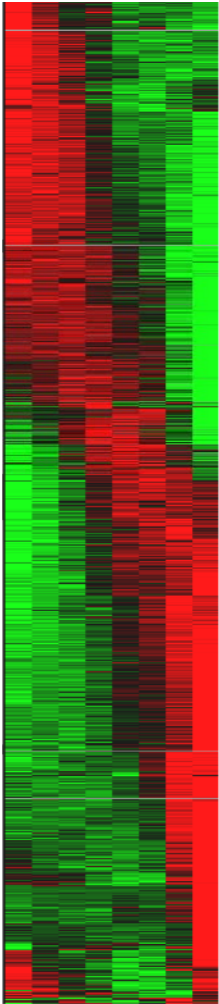
Bioinformatics predictions, experimental validation and analysis of cis-regulatory modules



Application to *D.melanogaster*'s cardiogenesis

Delphine Potier, S. Aerts, C. Herrmann & L. Perrin

Can we identify a cis-regulatory code?



Similarity of several genes expression



Potential similarity in the
regulating sequences

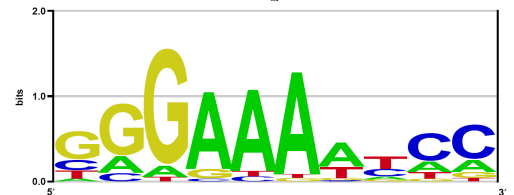
→ ***A bioinformatic approach can help to identify the mechanisms of genes transcriptional regulation: identify cis-regulatory modules (CRM)***

How can bioinformatics help us?

Currently:

- Few known Cis-Regulatory Modules (CRMs)
($< 2\%$ of *Drosophila* genes are associated to identified CRMs)
- Few (26% of known CRM) have been characterized in detail
- Not all binding sites for transcription factors (TFs) are characterized

Halfon & al. 2008



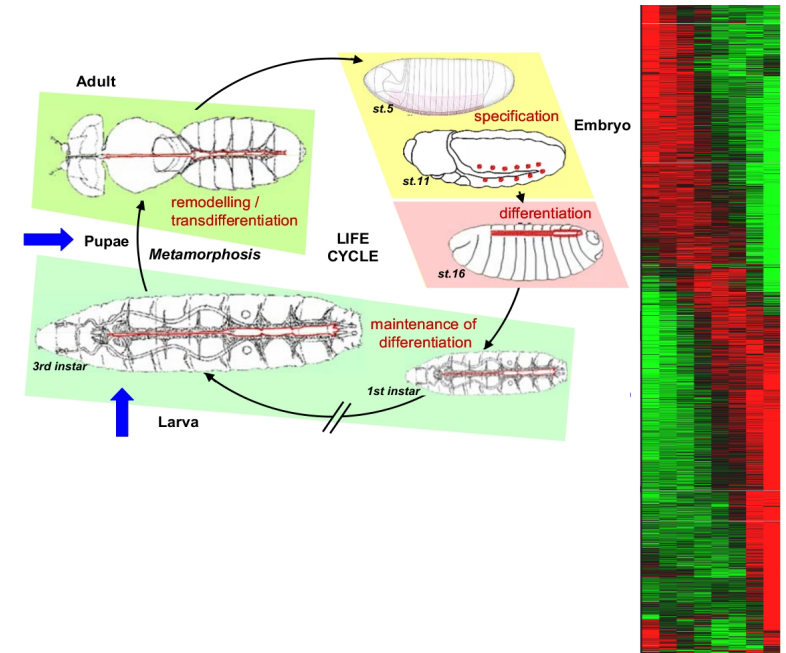
Bioinformatics can:

- Predict CRMs before experimental validation
- Help to understand CRMs syntax / Discover a cis-regulatory code

Generic strategy applied to *D.melanogaster*

We have a biological question:

- co-expressed gene data
- no *a priori* on the TFs involved



Our in silico approach: **cisTargetX**

Goal: predict motifs involved in the regulation of co-expressed genes

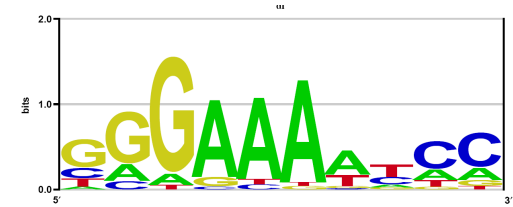
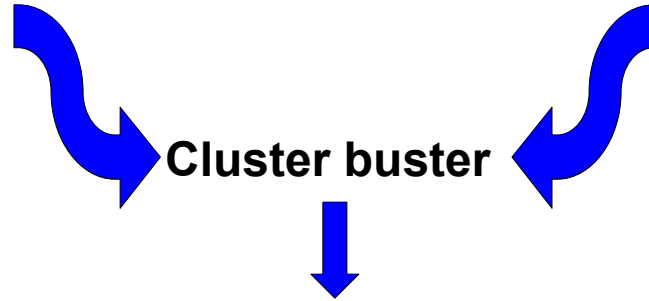
Method: rank all *Drosophila* genes using

- a comprehensive set of position-weight matrices (~2000)
- « conservation » between 12 *drosophila* species

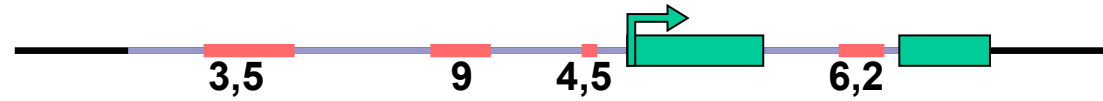
CisTargetX principle



— All 5kb up + introns for all *D. mel* genes



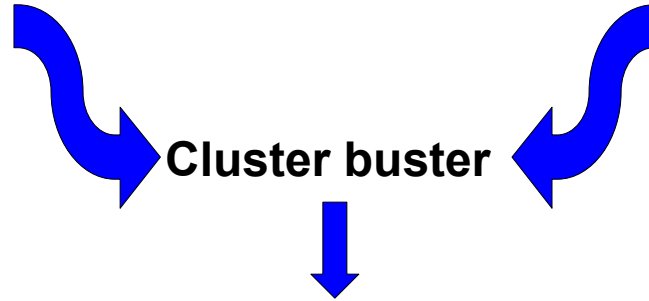
A matrix



CisTargetX principle

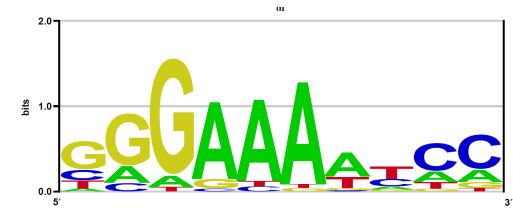
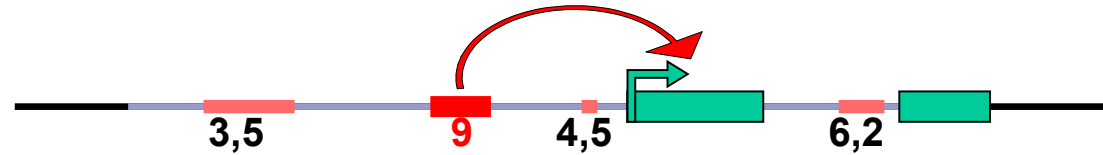


— All 5kb up + introns for all *D. mel* genes



Cluster buster

Take the best score among identified clusters / region

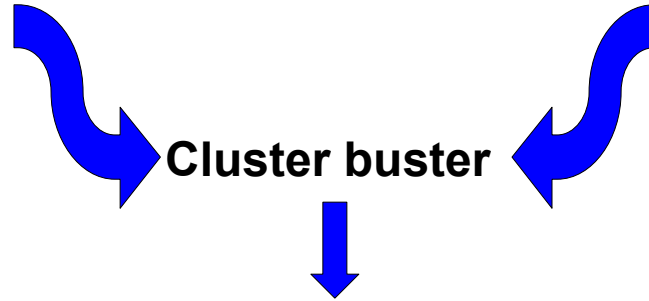


A matrix

CisTargetX principle



— All 5kb up + introns for all *D. mel* genes

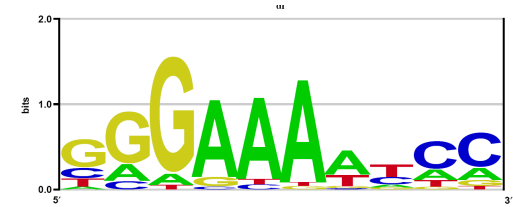


Cluster buster

Take the best score among identified clusters / region

Gene Score	
CGxxx	9
CGyyy	8
*	
*	

Ranking of all
Drosophila genes



A matrix

CisTargetX principle



All 5kb up + introns for all *D. mel* genes

Cluster buster



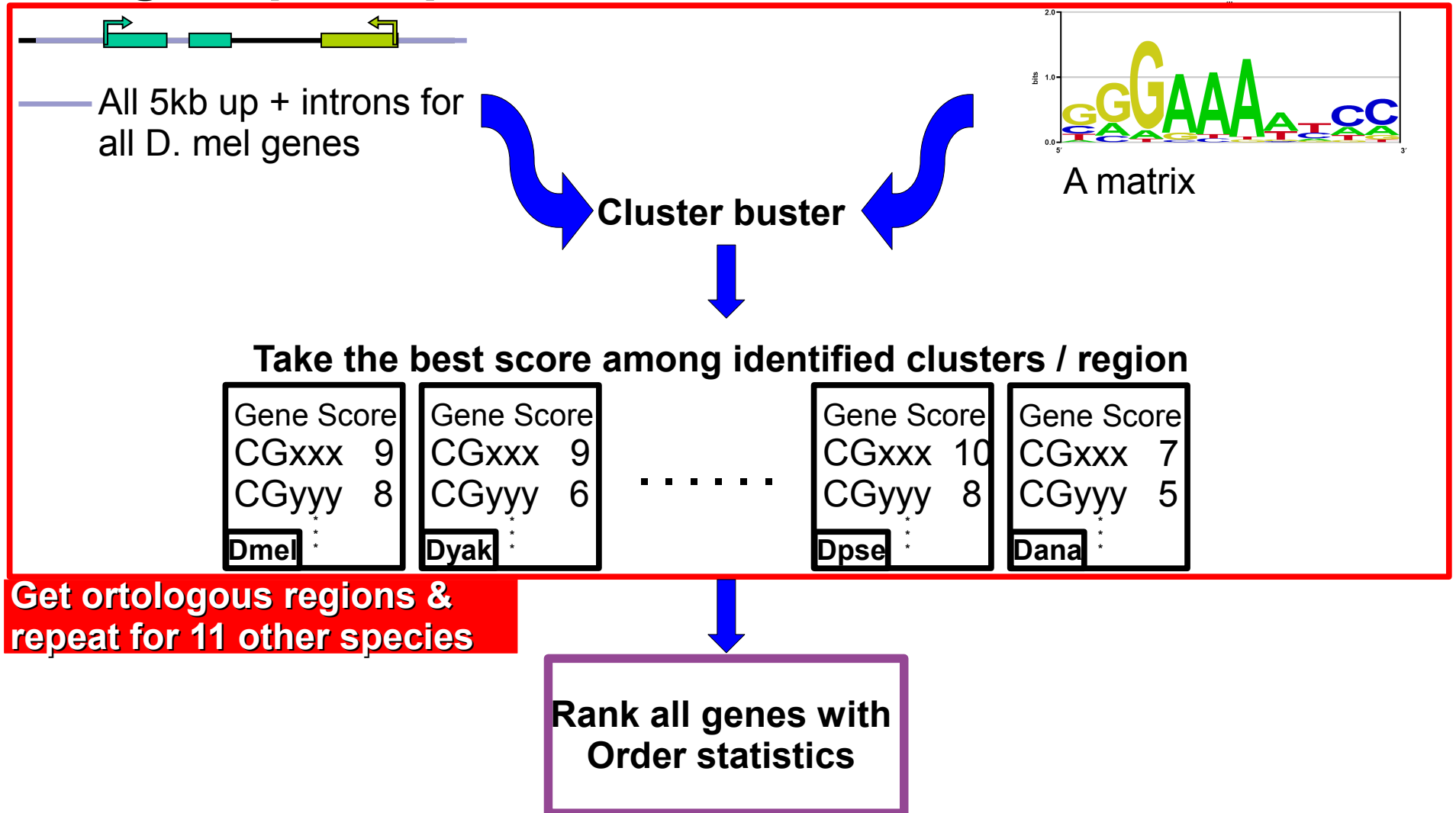
A matrix

Take the best score among identified clusters / region

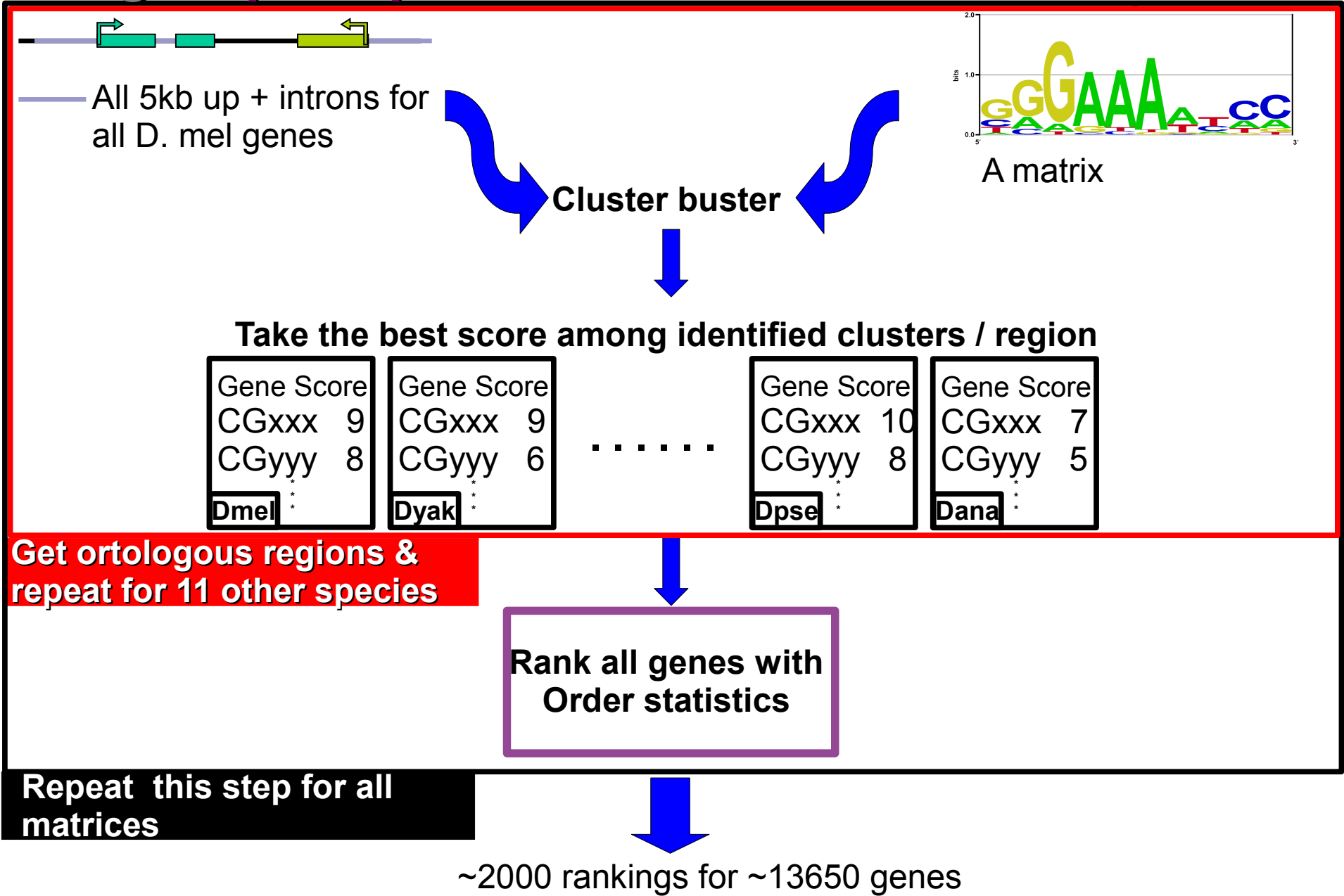
Gene Score	Gene Score		Gene Score	Gene Score
CGxxx 9	CGxxx 9	...	CGxxx 10	CGxxx 7
CGyyy 8	CGyyy 6		CGyyy 8	CGyyy 5
Dmel	Dyak		Dpse	Dana

Get ortologous regions & repeat for 11 other species

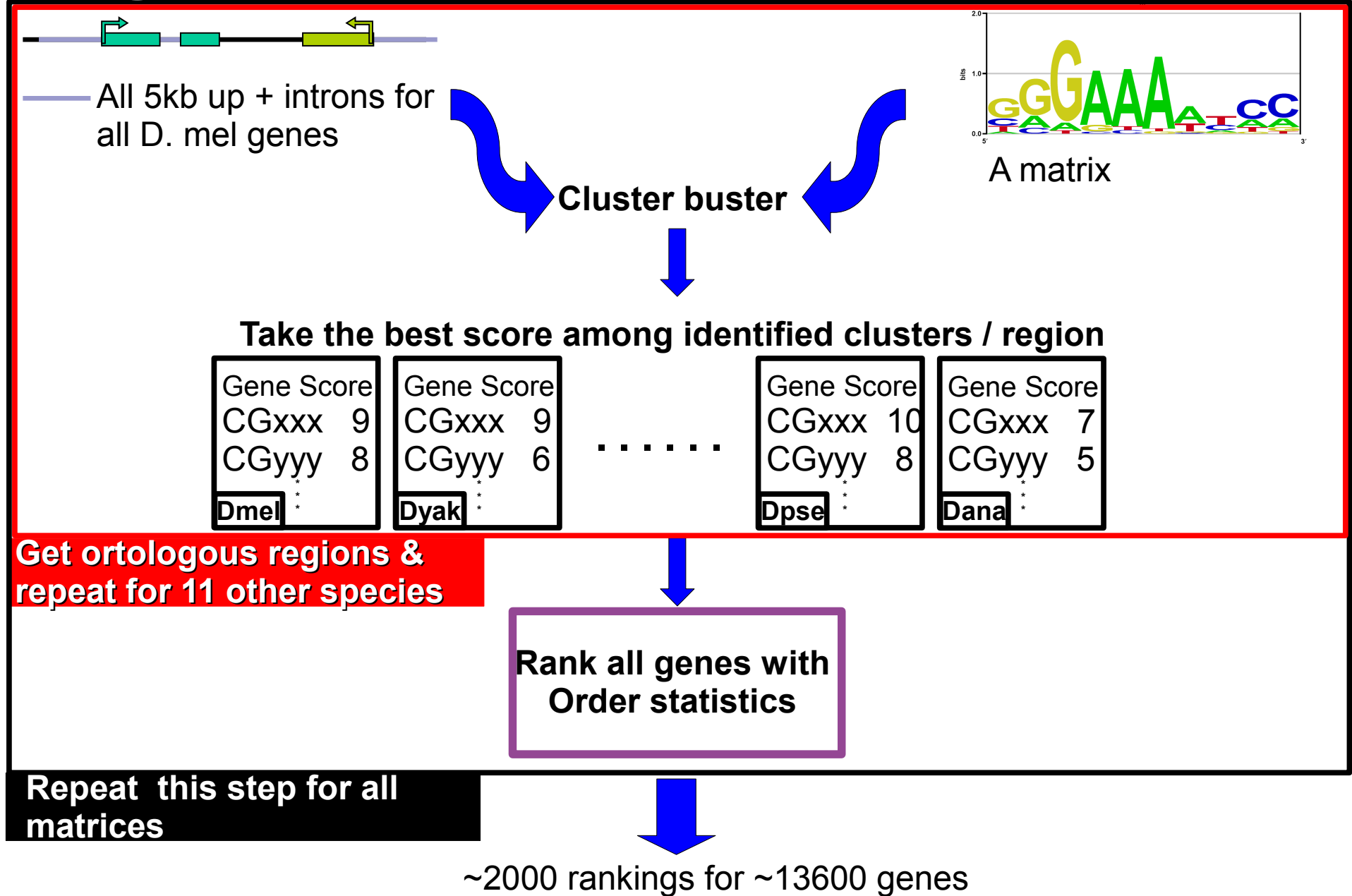
CisTargetX principle



CisTargetX principle



CisTargetX principle

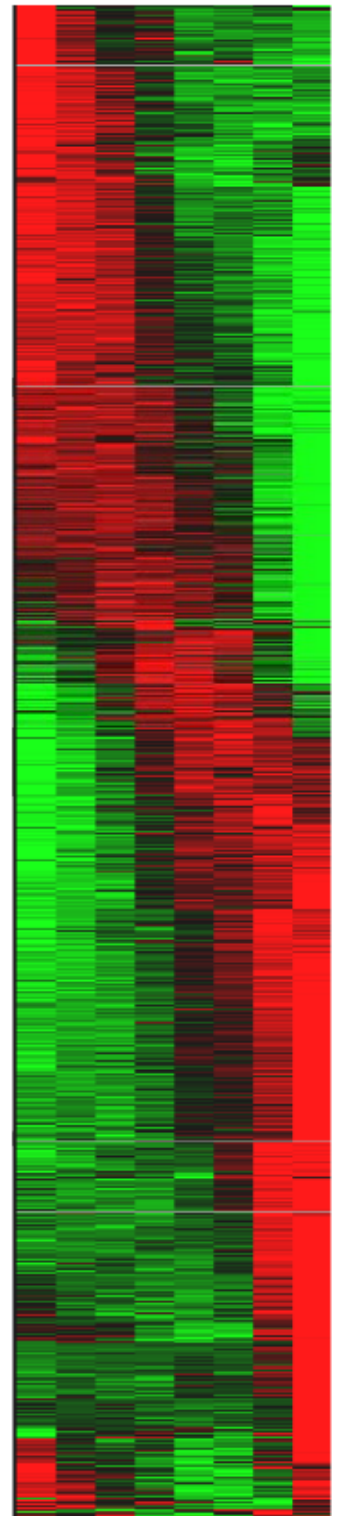
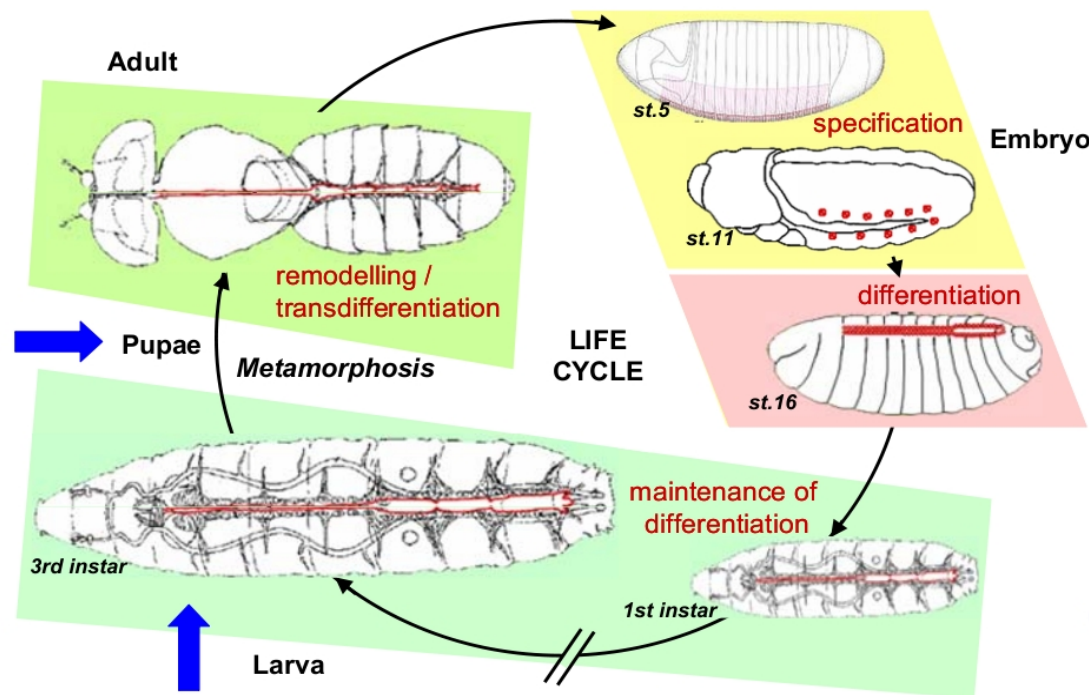


Which matrices rank my genes best?

Studied system :

Drosophila cardiac tube remodelling

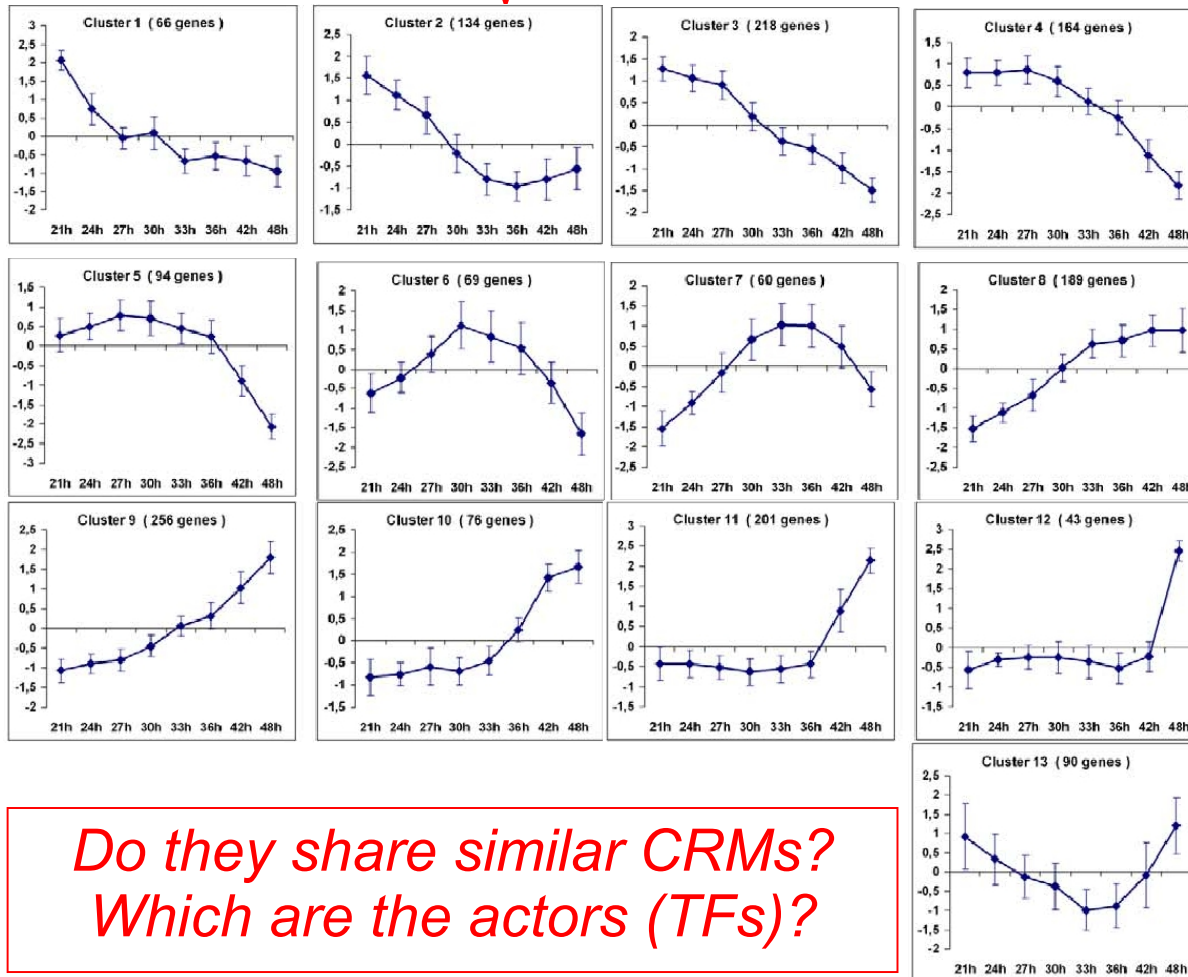
- Drosophila: genomic and genetic tool
- Drosophila heart :
 - Simple linear tube (~100 cells)
 - Metamorphosis is well characterised



Data:

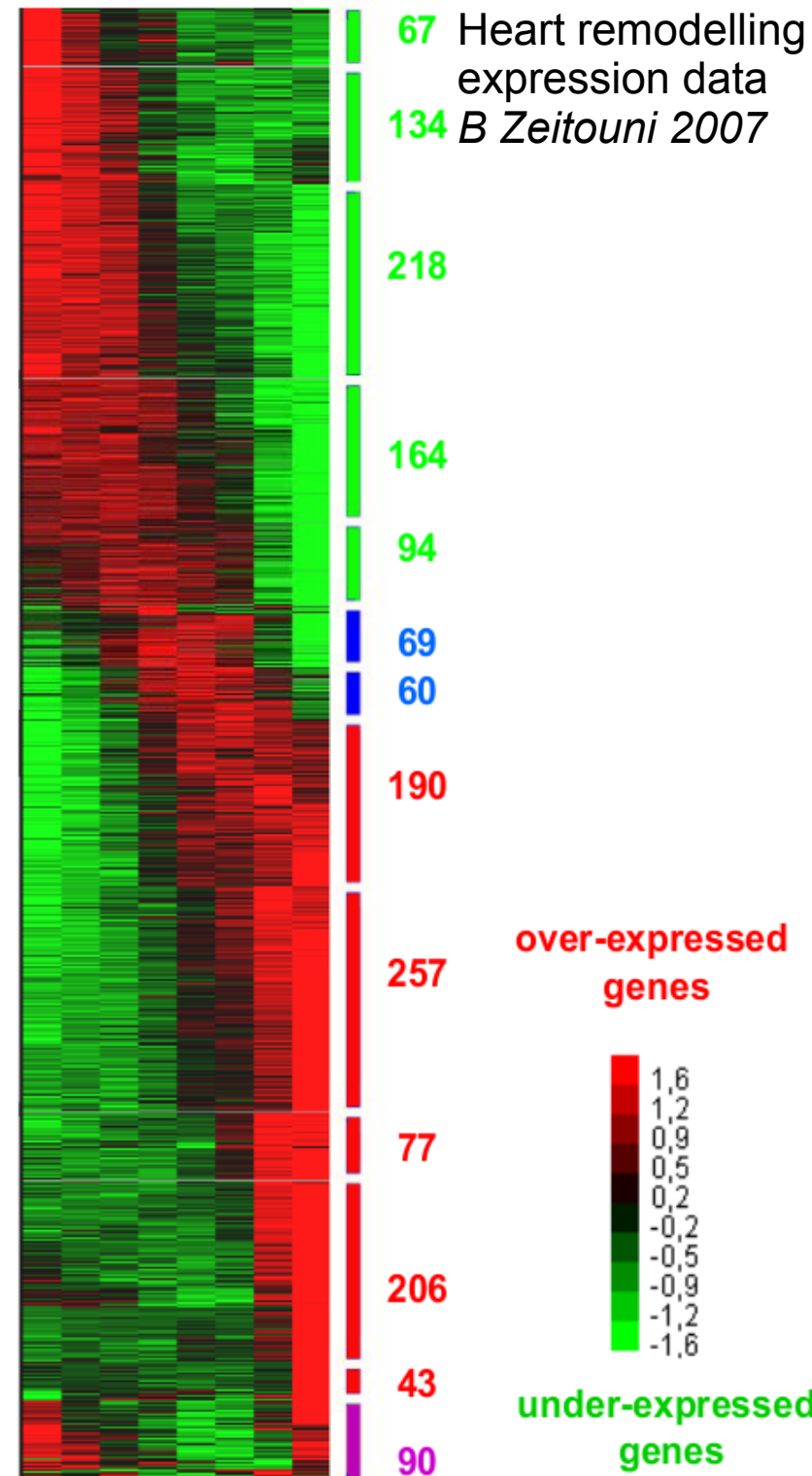
Large scale approach:
MicroArray data of
heart metamorphosis

13 clusters



Do they share similar CRMs?
Which are the actors (TFs)?

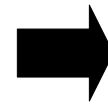
21 24 27 30 33 36 42 48



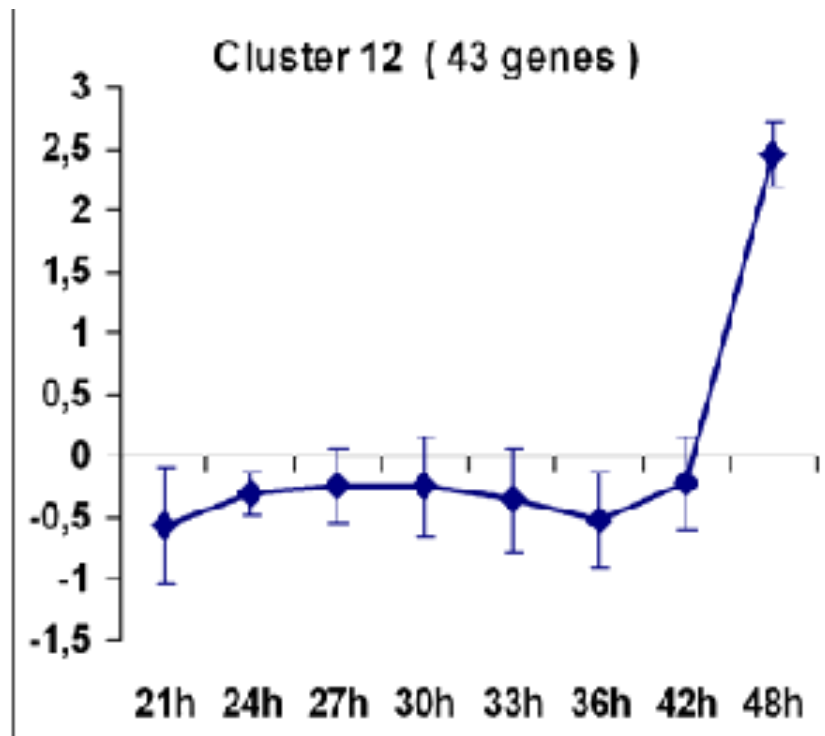
Data:

Large scale approach:
MicroArray data of
heart metamorphosis

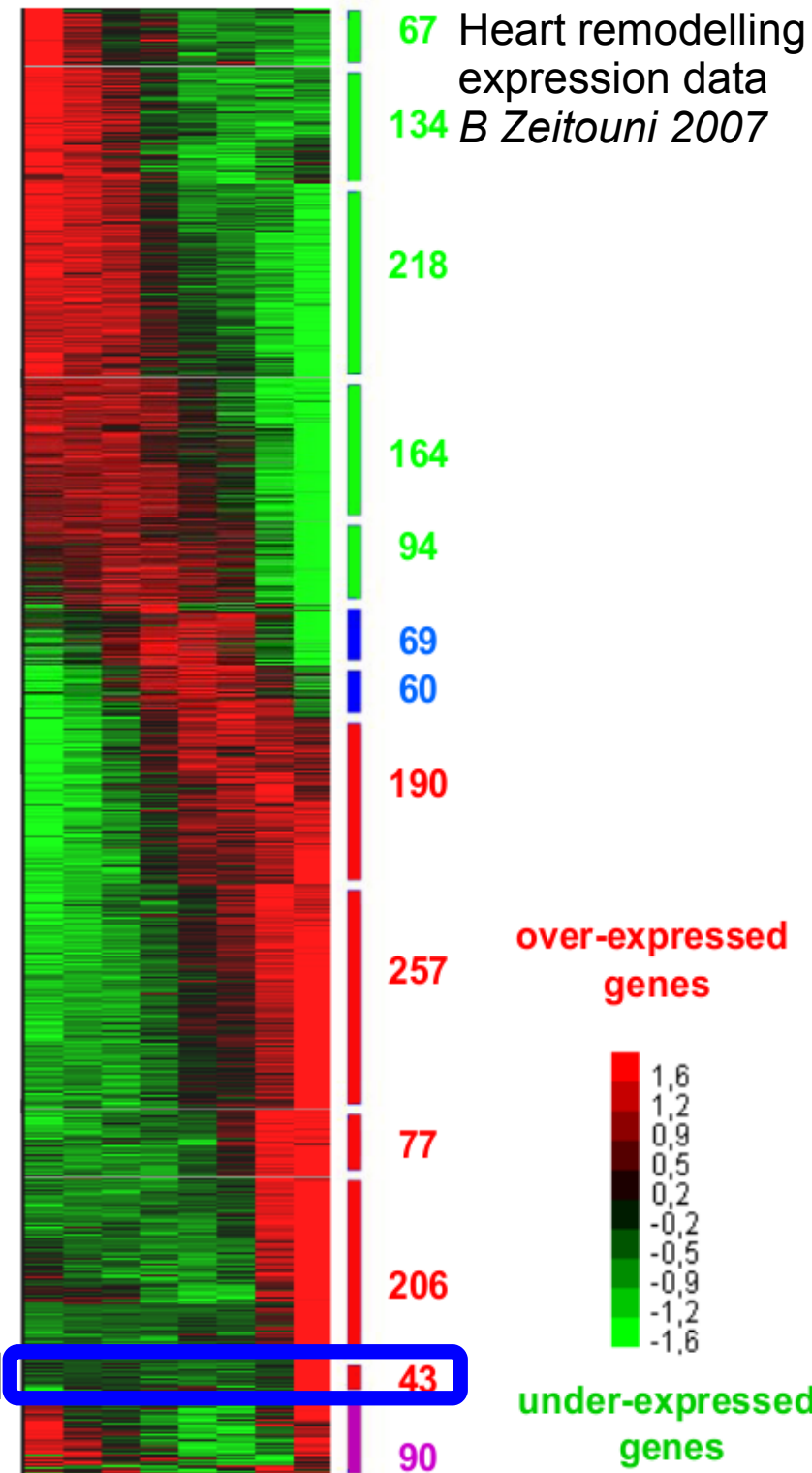
13 clusters



Best results with CisTarget-X:

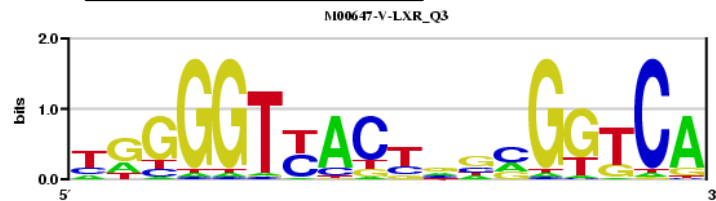


21 24 27 30 33 36 42 48

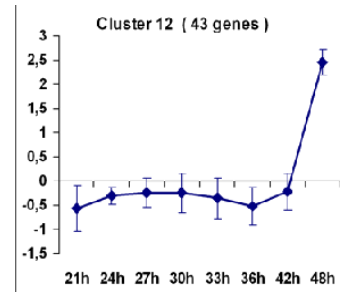
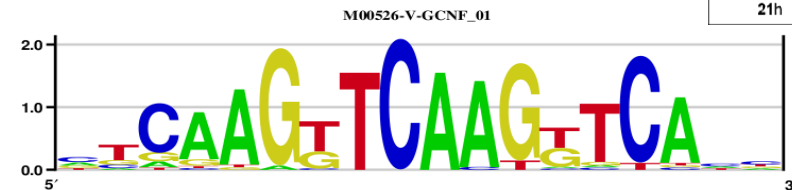


Motif identification by gene ranking : Cluster 12 example

Matrix 1: LXR
gene ranking



Matrix 83: GCNF
gene ranking



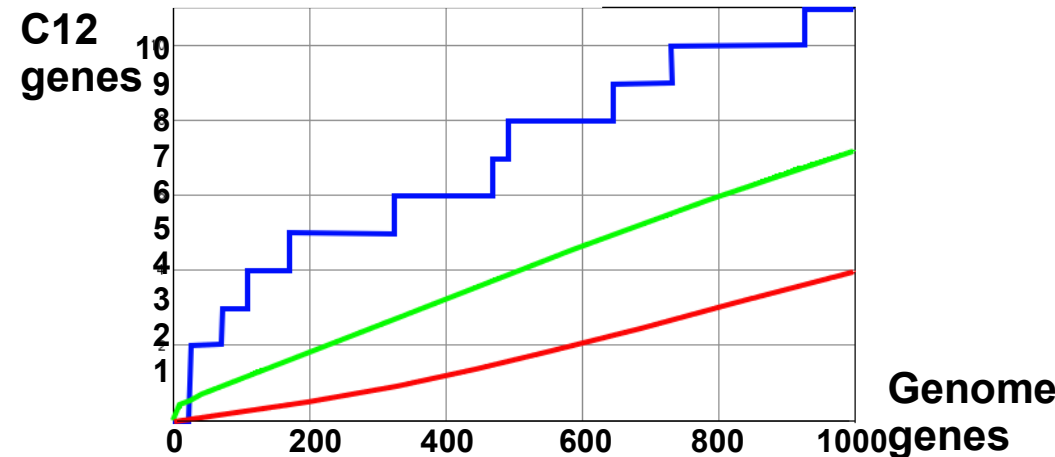
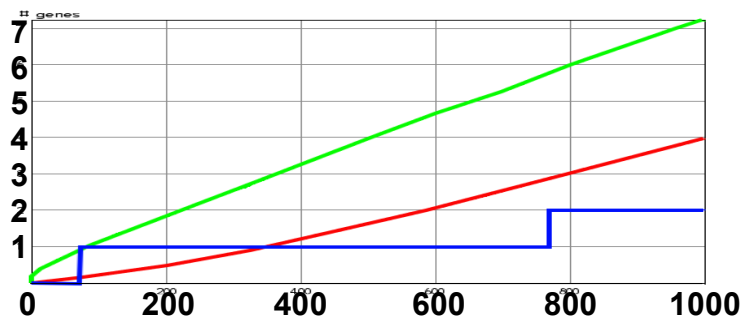
Matrices
(~2000)

Gene ranking

1-xxxx
2-xxxx
...
72-[CG17907](#)
73-xxxx
...
772-[CG10076](#)
773-xxxx
...
13667-xxxx

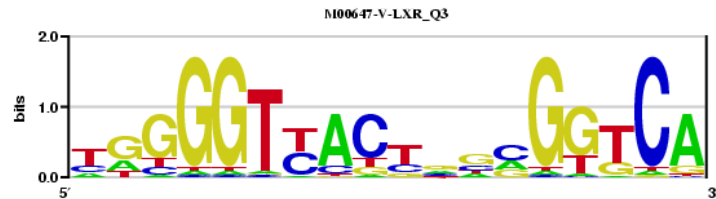
1-xxxx
2-xxxx
...
24-[CG15545](#)
25-xxxx
26-[CG17298](#)
27-xxxx
...
72-[CG4998](#)
73-xxxx
...
109-[CG11380](#)
110-xxxx
...
171- [CG10175](#)
172-xxxx
...
325- [CG3902](#)
326-xxxx
...
13667-xxxx

ROC curves

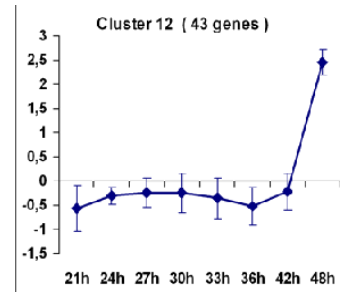
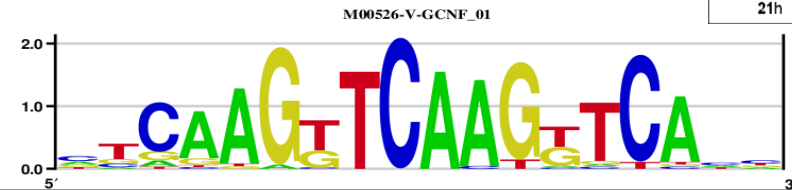


Motif identification by gene ranking : Cluster 12 example

Matrix 1: LXR
gene ranking



Matrix 83: GCNF
gene ranking



Matrices
(~2000)

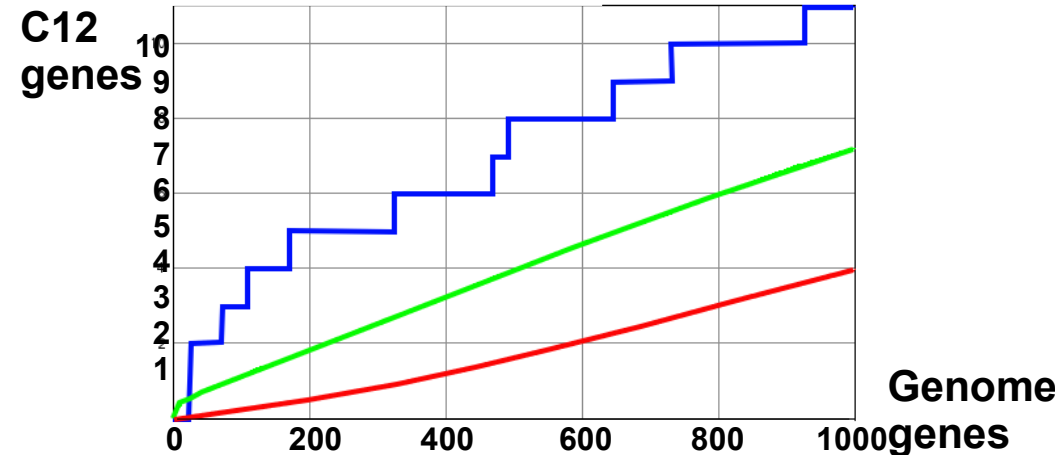
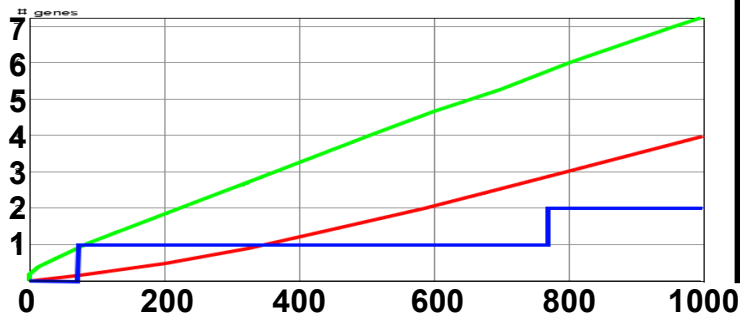
Gene ranking

1-xxxx
2-xxxx
...
72-[CG17907](#)
73-xxxx
...
772-[CG10076](#)
773-xxxx
...
13667-xxxx

non relevant

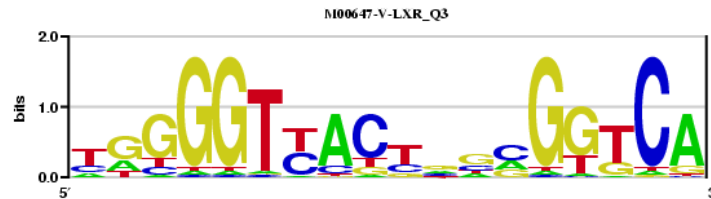
1-xxxx
2-xxxx
...
24-[CG15545](#)
25-xxxx
26-[CG17298](#)
27-xxxx
...
72-[CG4998](#)
73-xxxx
...
109-[CG11380](#)
110-xxxx
...
171- [CG10175](#)
172-xxxx
...
325- [CG3902](#)
326-xxxx
...
13667-xxxx

ROC curves

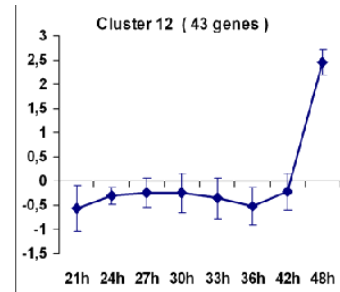
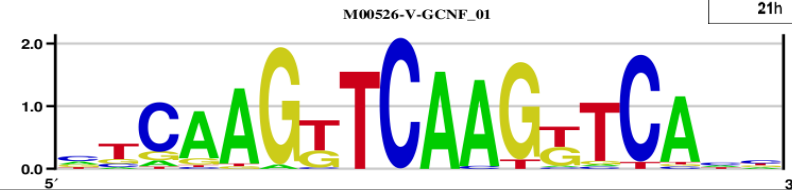


Motif identification by gene ranking : Cluster 12 example

Matrix 1: LXR
gene ranking



Matrix 83: GCNF
gene ranking



Matrices
(~2000)

Gene ranking

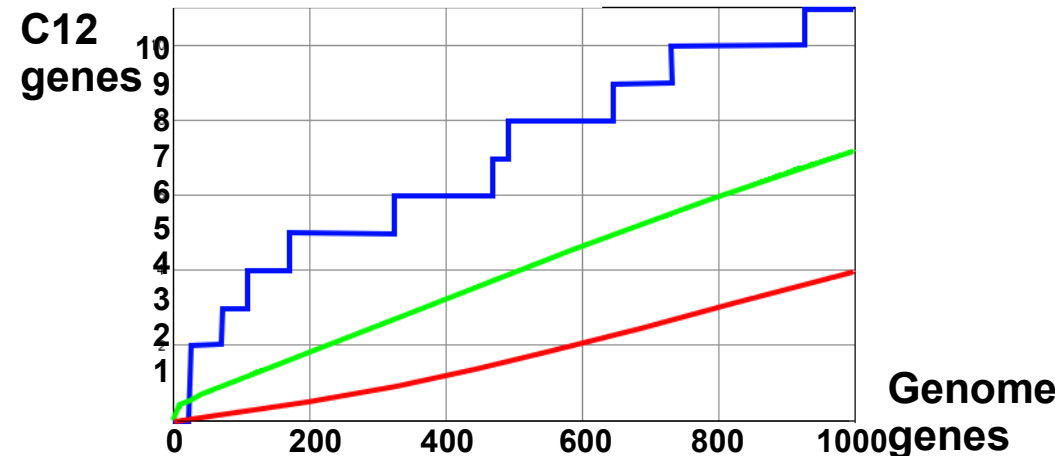
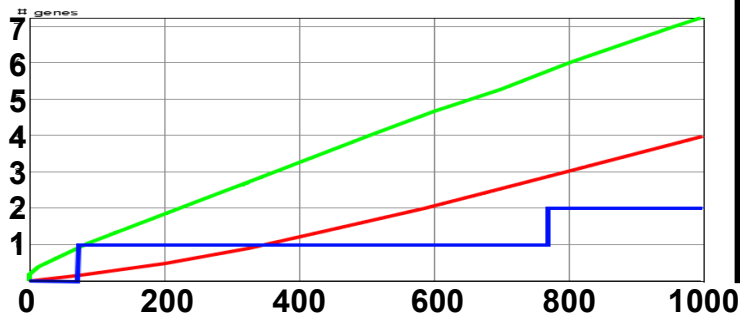
1-xxxx
2-xxxx
...
72-[CG17907](#)
73-xxxx
...
772-[CG10076](#)
773-xxxx
...
13667-xxxx

non relevant

1-xxxx
2-xxxx
...
24-[CG15545](#)
25-xxxx
26-[CG17298](#)
27-xxxx
...
72-[CG4998](#)
73-xxxx
...
109-[CG11380](#)
110-xxxx
...
171- [CG10175](#)
172-xxxx
...
325- [CG3902](#)
326-xxxx
...
13667-xxxx

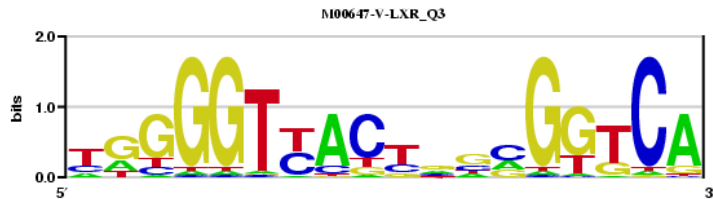
relevant

ROC curves

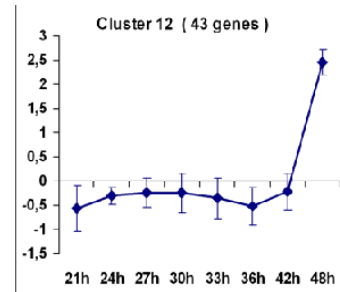
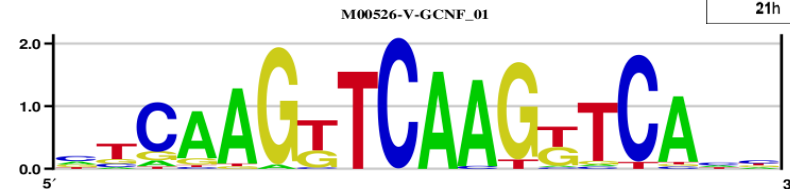


Motif identification by gene ranking : Cluster 12 example

Matrix 1: LXR
gene ranking



Matrix 83: GCNF
gene ranking



Matrices
(~2000)

Gene ranking

1-XXXX
2-XXXX
...
72-[CG17907](#)
73-XXXX
...
772-[CG10076](#)
773-XXXX
...
13667-XXXX

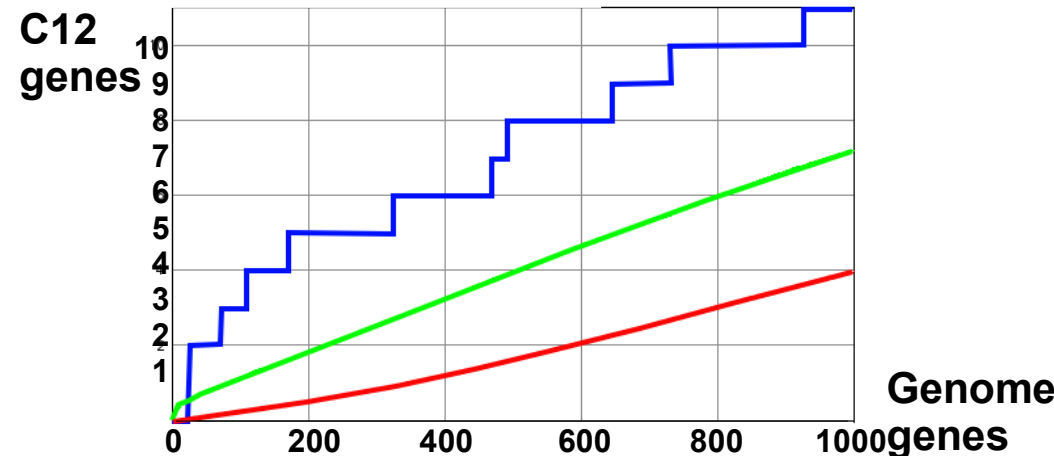
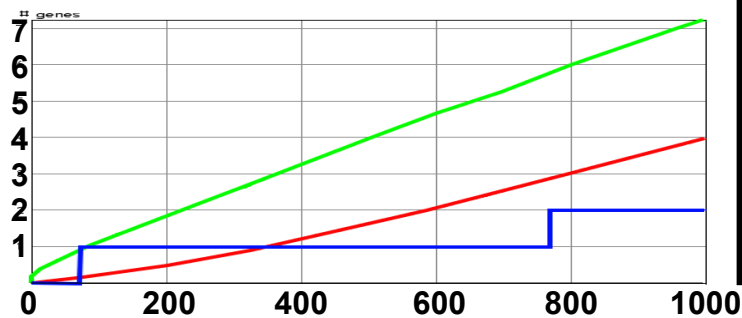
non relevant

1-XXXX
2-XXXX
...
24-[CG15545](#)
25-XXXX
26-[CG17298](#)
27-XXXX
...
72-[CG4998](#)
73-XXXX
...
109-[CG11380](#)
110-XXXX
...
171- [CG10175](#)
172-XXXX
...
325- [CG3902](#)
326-XXXX
...
13667-XXXX

→ GCNF motif might
represent a common
motif for cluster 12

relevant

ROC curves



Selection of putative CRMs

CisTarget-X: prediction of CRM around 6 co-expressed genes containing GCNF-like motifs



6 significantly highly ranked genes

24 CG15545
26 CG17298
72 CG4998
109 CG11380
171 CG10175
325 CG3902



Take 200bp up & downstream
of GCNF binding sites
(in 5kb up +introns)



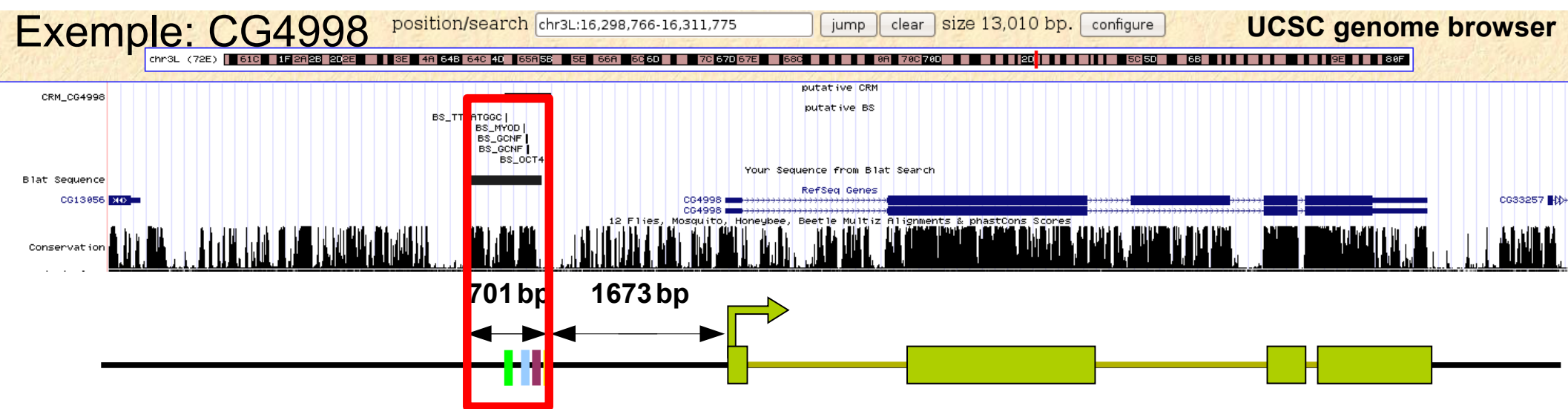
Clover

Selection of putative CRMs

CisTarget-X: prediction of CRM around 6 co-expressed genes containing GCNF-like motifs

+

Other shared motifs



→ Selection of putative CRM for 5 other genes of cluster 12

In vivo validation of predicted CRM

→ Predicted CRMs were tested by reporter gene expression (lacZ)

1. PCR on →
genomic DNA



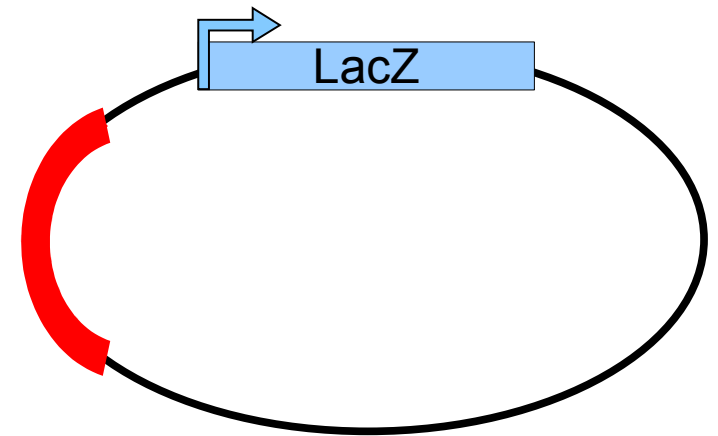
In vivo validation of predicted CRM

→ Predicted CRMs were tested by reporter gene expression (lacZ)

1. PCR on
genomic DNA



2. Cloning into LacZ vector



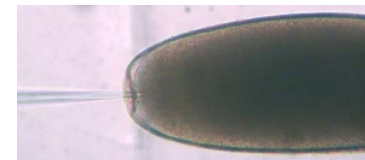
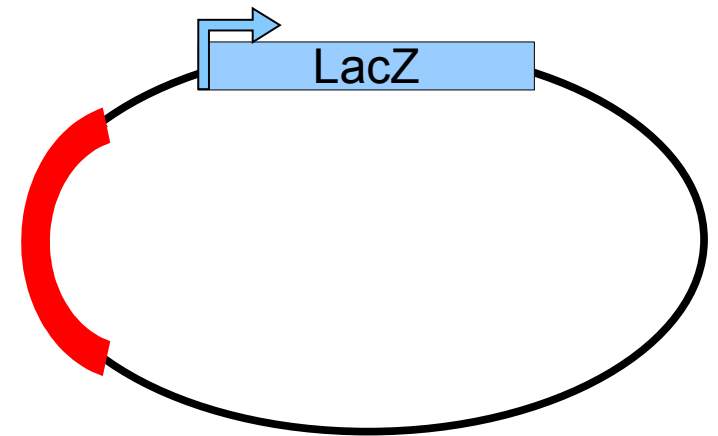
In vivo validation of predicted CRM

→ Predicted CRMs were tested by reporter gene expression (lacZ)

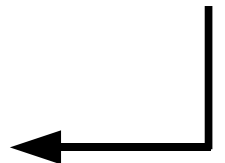
1. PCR on
genomic DNA



2. Cloning into LacZ vector



3. flies injection



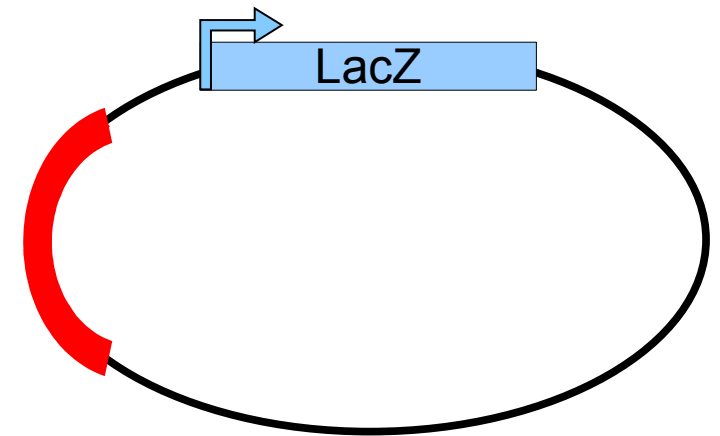
In vivo validation of predicted CRM

→ Predicted CRMs were tested by reporter gene expression (lacZ)

1. PCR on genomic DNA

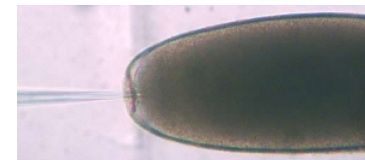


2. Cloning into LacZ vector



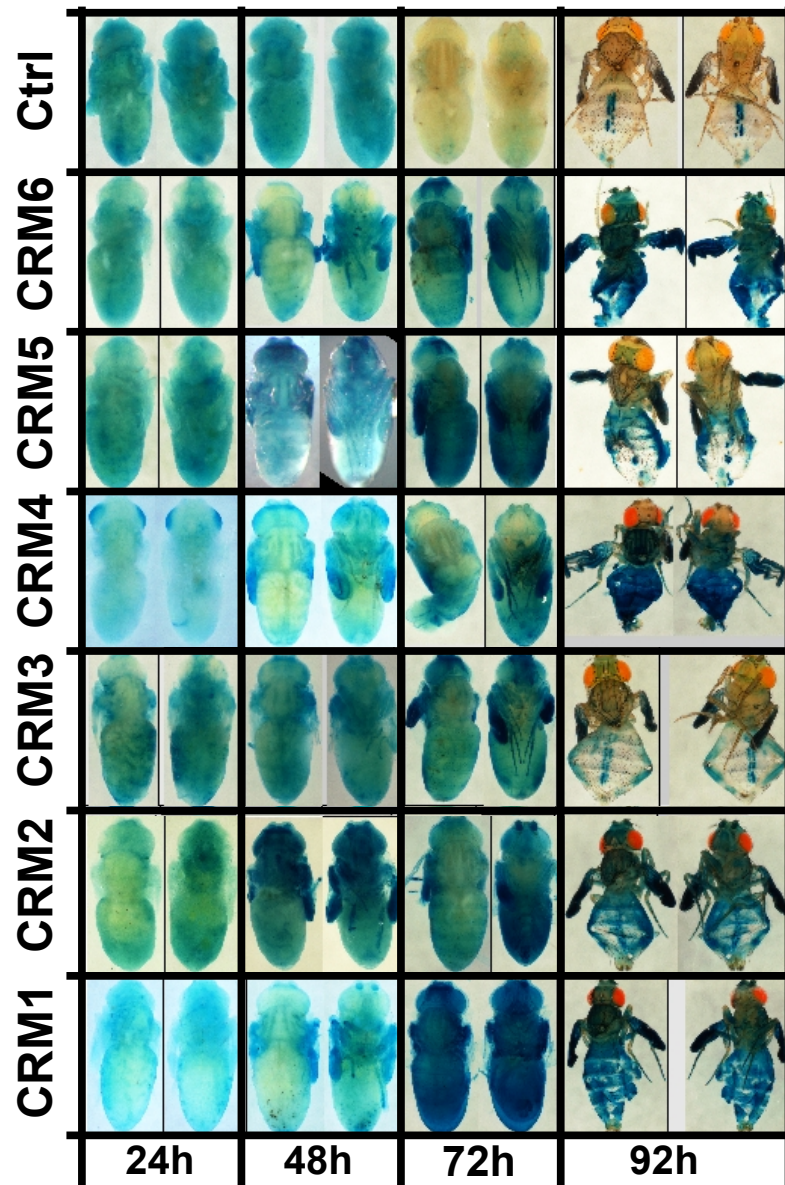
4. Look at LacZ expression

3. flies injection



In vivo validation of predicted CRM

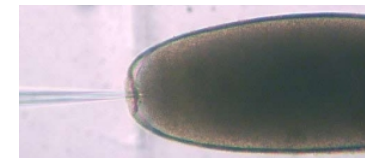
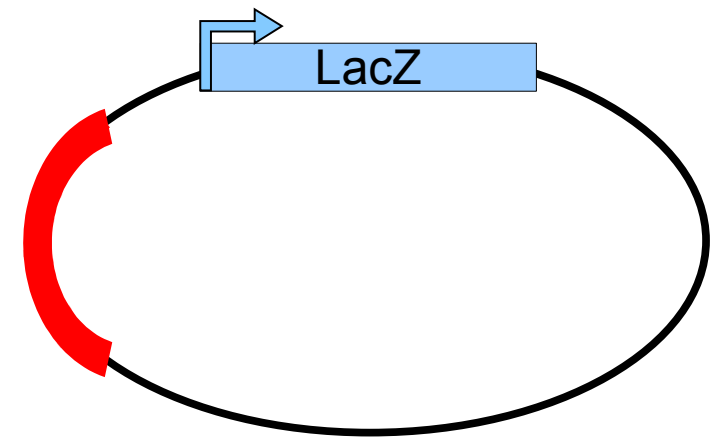
→ Predicted CRMs were tested by reporter gene expression (lacZ)



1. PCR on genomic DNA



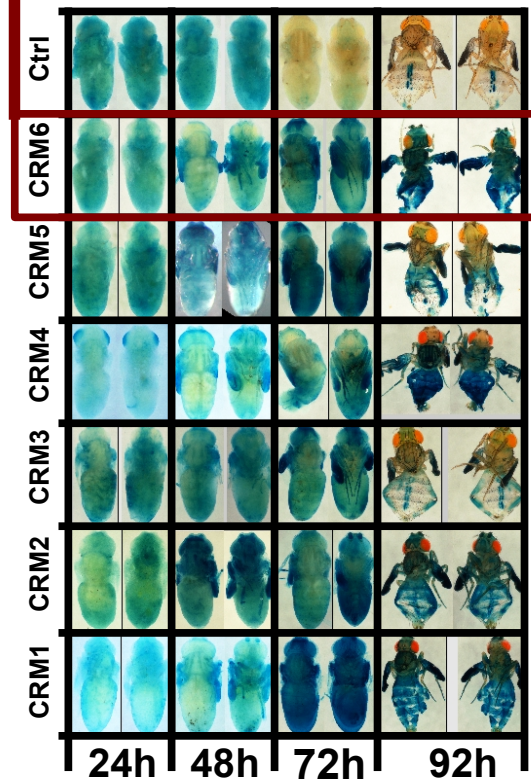
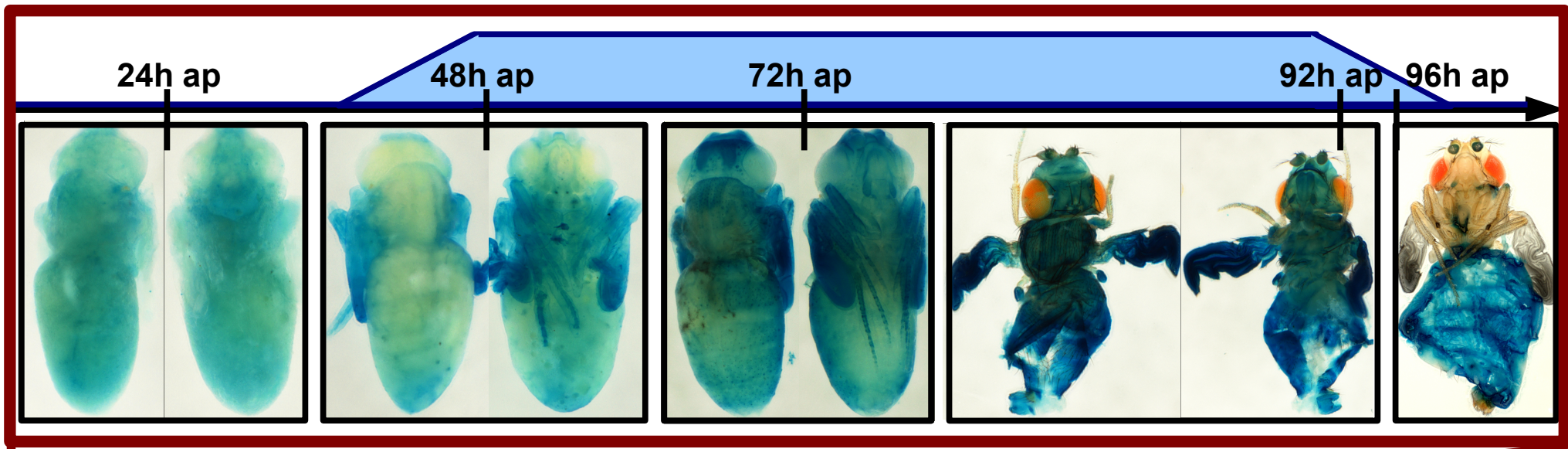
2. Cloning into LacZ vector



4. Look at LacZ expression

3. flies injection

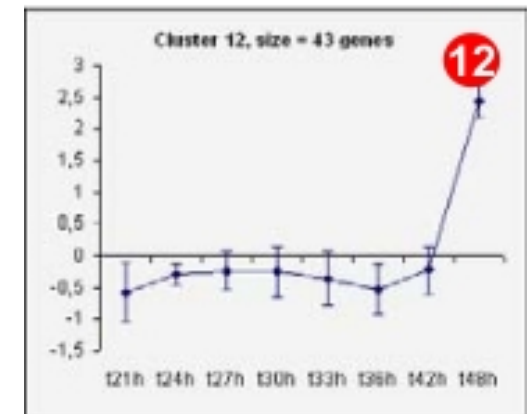
In vivo validation of predicted CRM



Results:

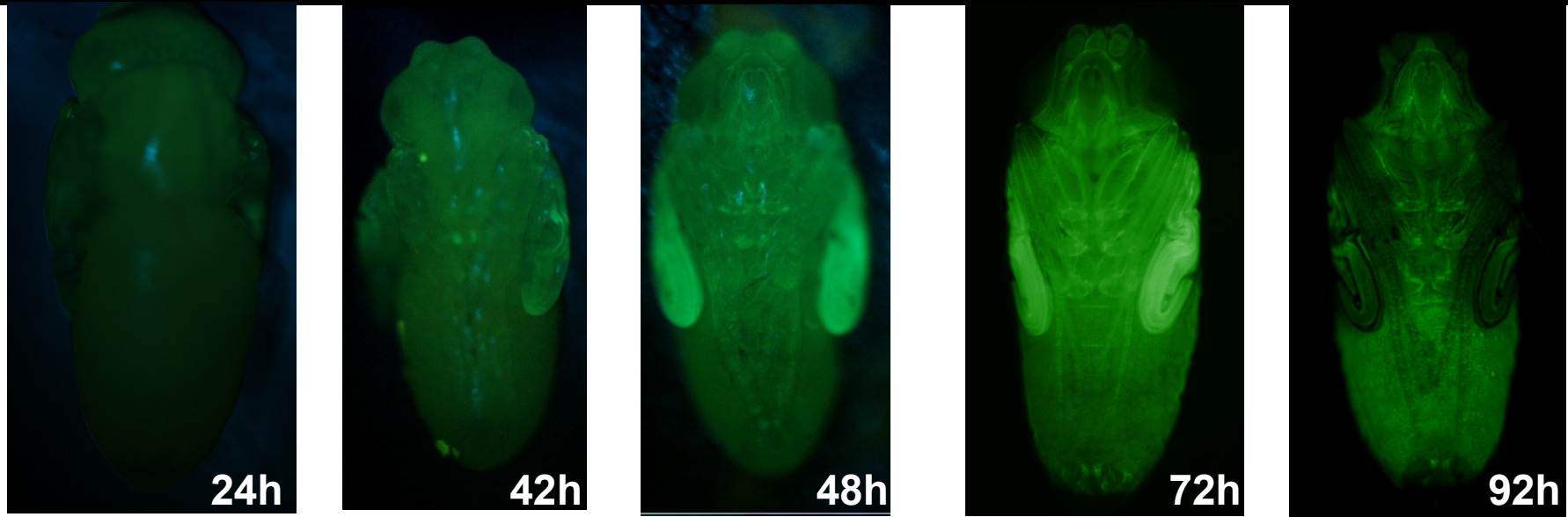
All tested CRMs have similar temporal expression pattern with a major induction at 48h

Fit to cluster 12 temporal expression pattern

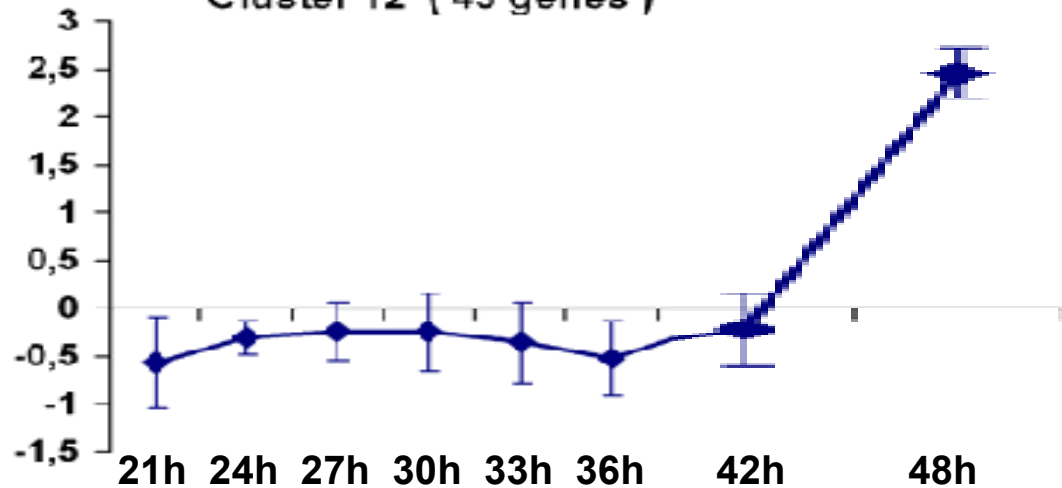


Tested CRMs reproduce the temporal expression pattern

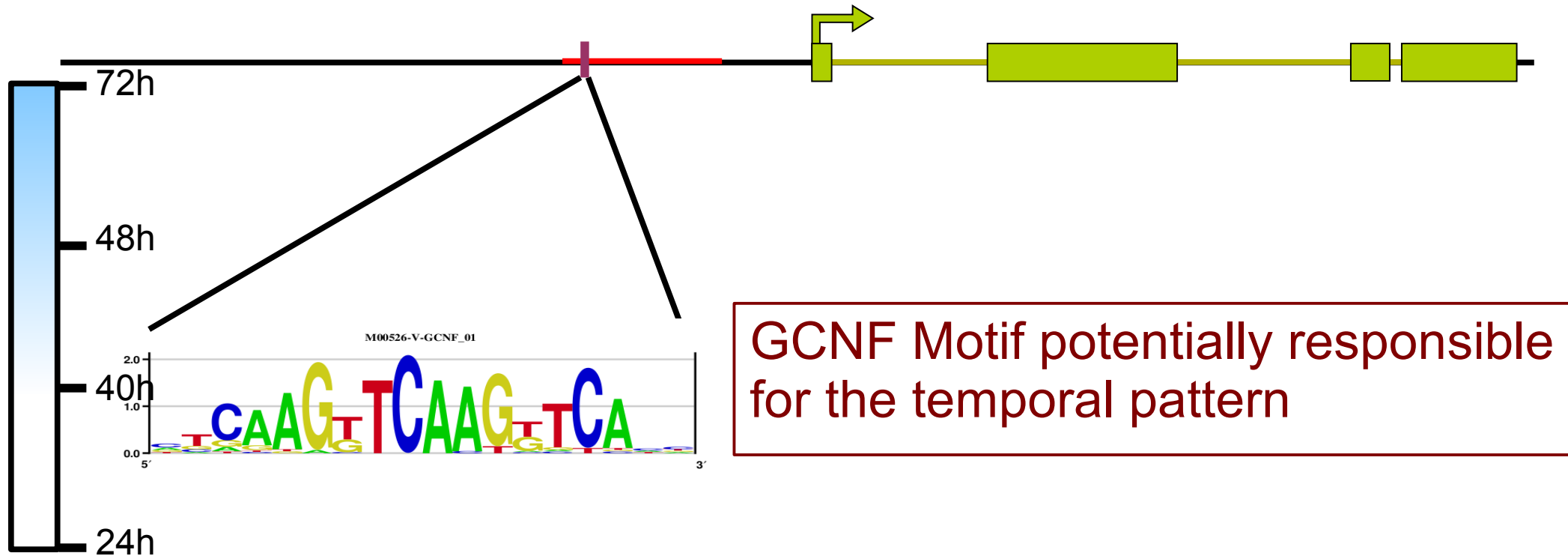
CG15545-GFP



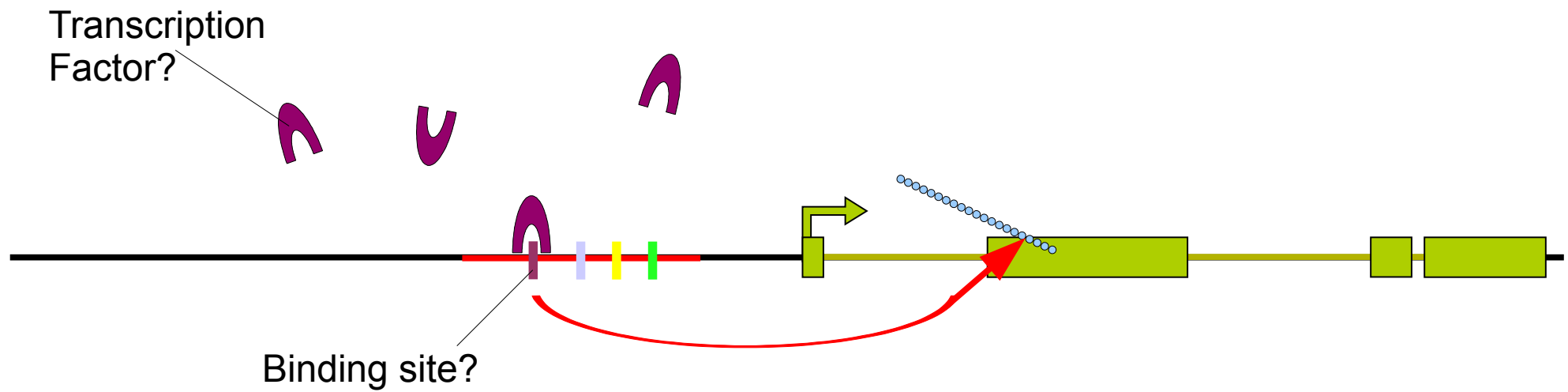
Cluster 12 (43 genes)



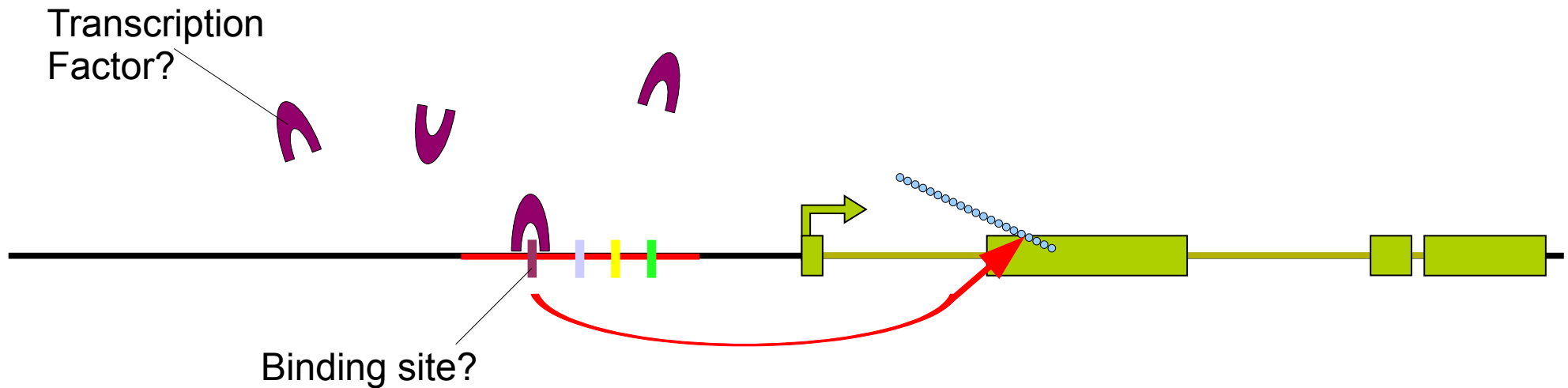
Tested CRMs reproduce the temporal expression pattern



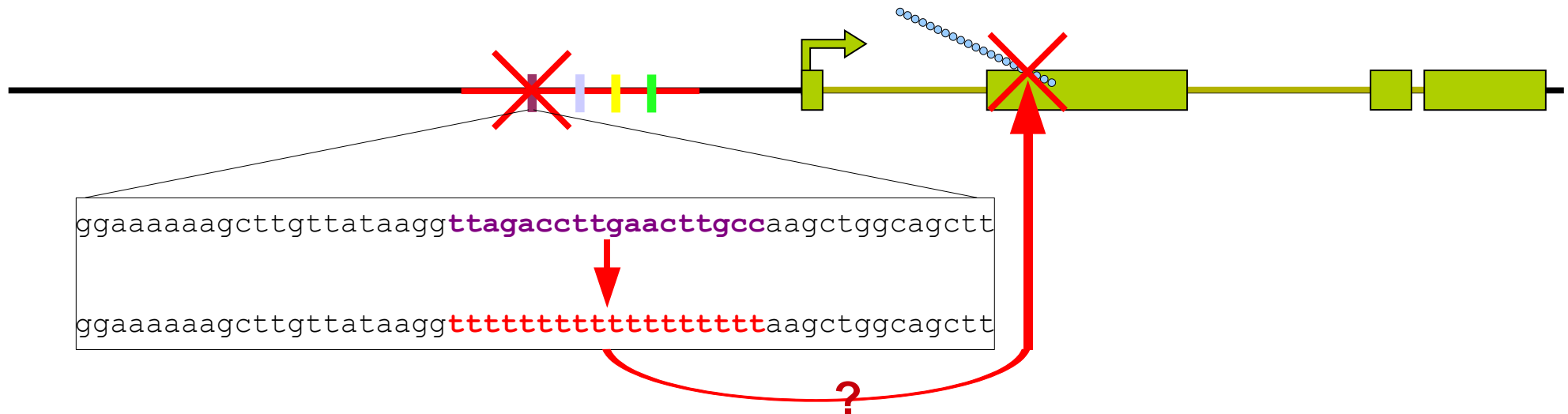
Validation of cis- and trans-actors



Validation of cis- and trans-actors

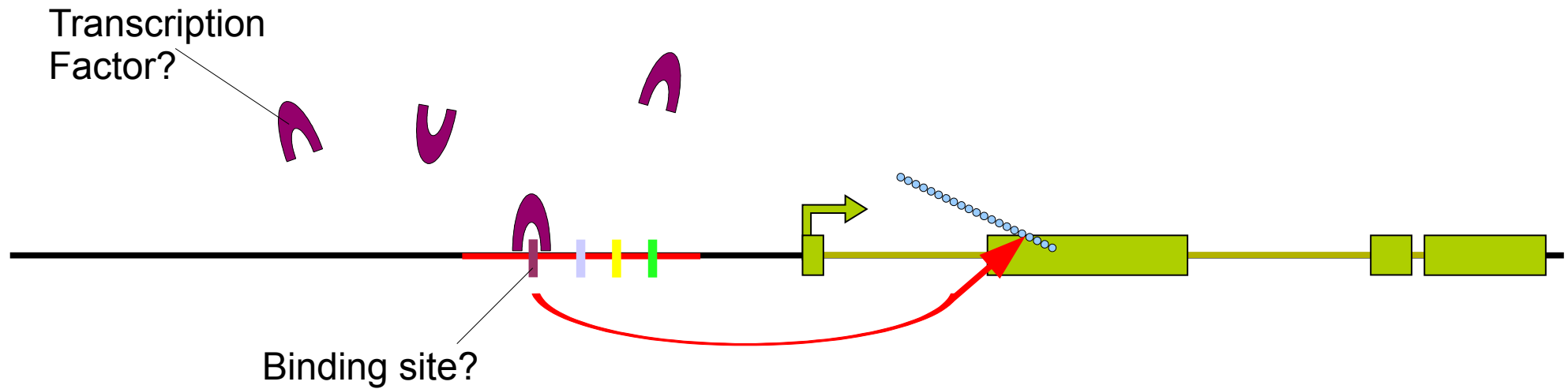


(1) Validation of cis-actors: Mutate putative GCNF sites

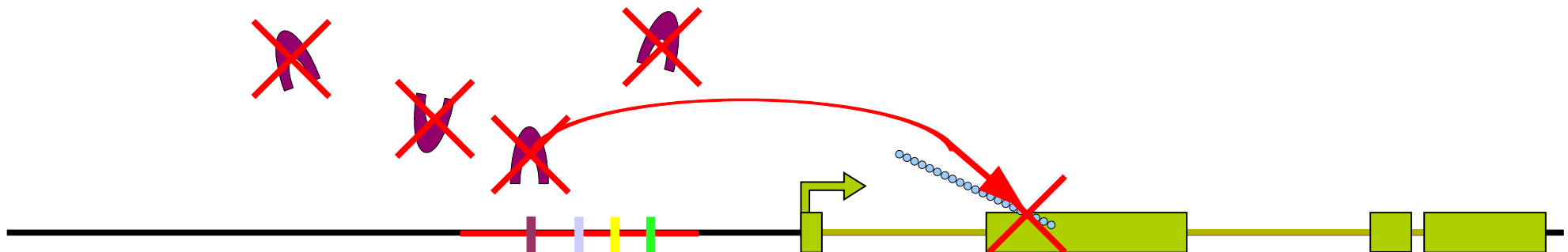


Results in ~ one week...

Validation of cis- and trans-actors

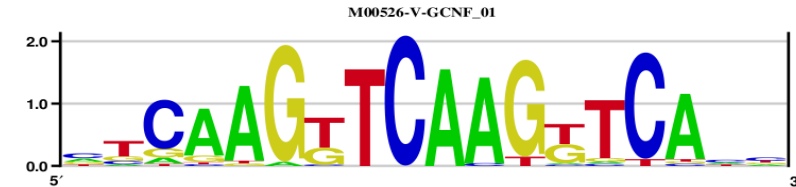


(2) Validation of trans-actors: RNAi for the putative involved TF

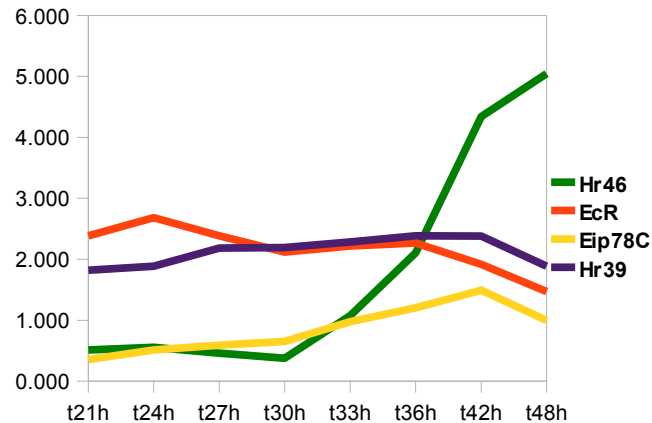


Which TF could bind GCNF motif?

→ mouse nuclear receptor(NR)
temporal gene expression regulator during fly pupation

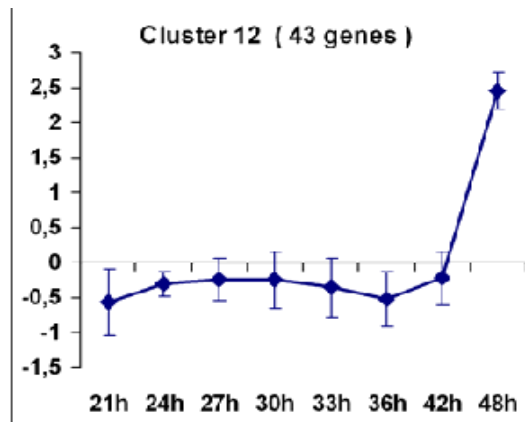


4 Drosophila NRs are expressed differentially from 21h to 48h ap in cardiac tube:



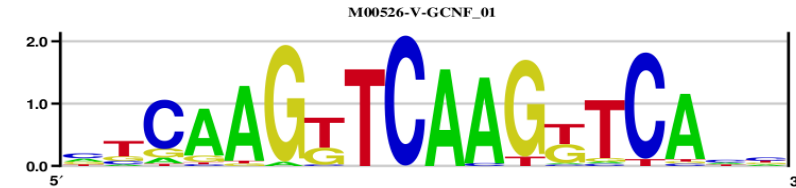
B. Zeitouni 2007

C12 genes expression (21h to 48h ap):

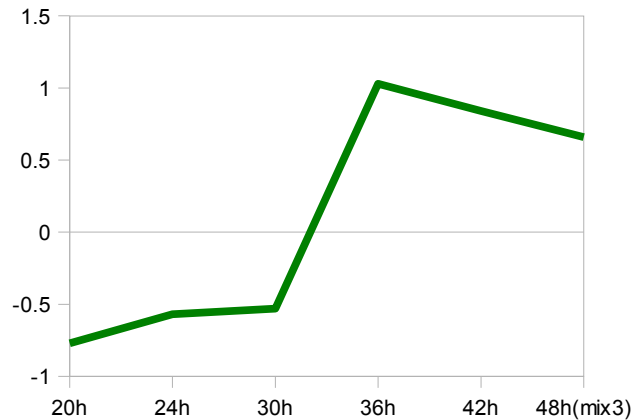


Which TF could bind GCNF motif?

→ mouse nuclear receptor(NR)
temporal gene expression regulator during fly pupation



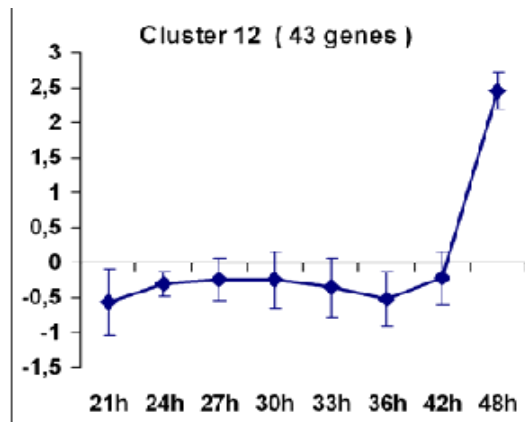
Drosophila Hr46 expression from 20h to 48h ap in total pupae:



Arbeitman 2002

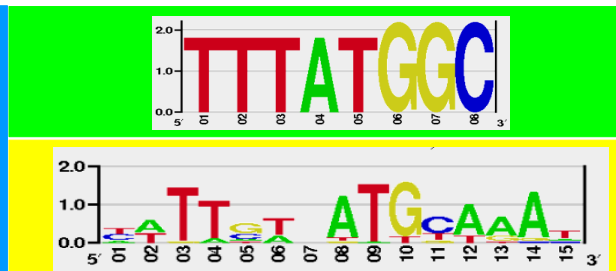
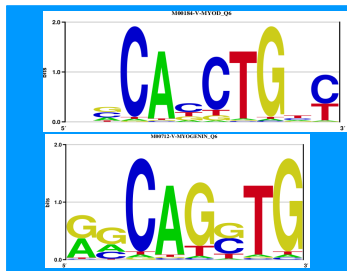
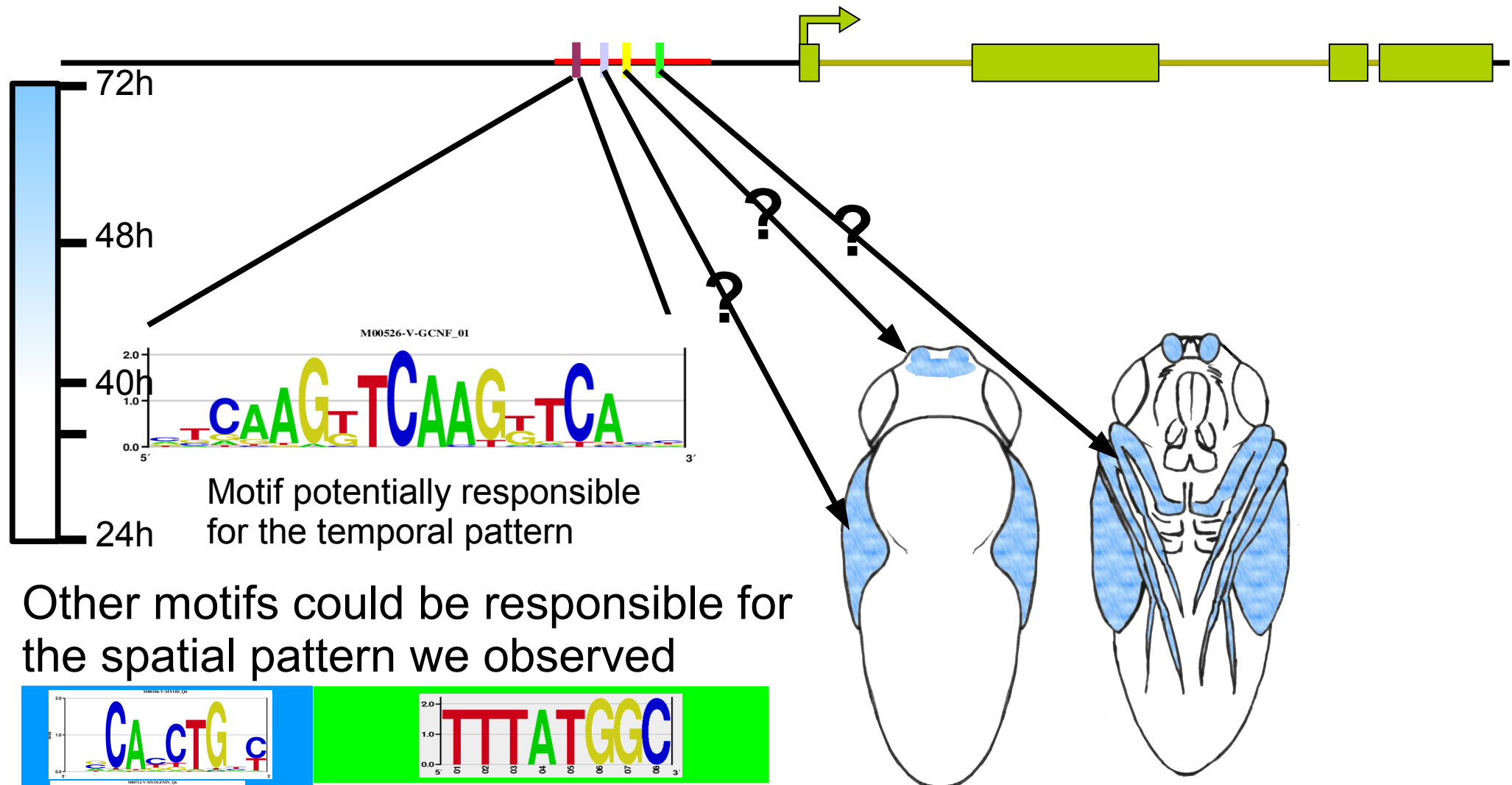
Hr46 is a potential candidate

C12 genes expression (21h to 48h ap):

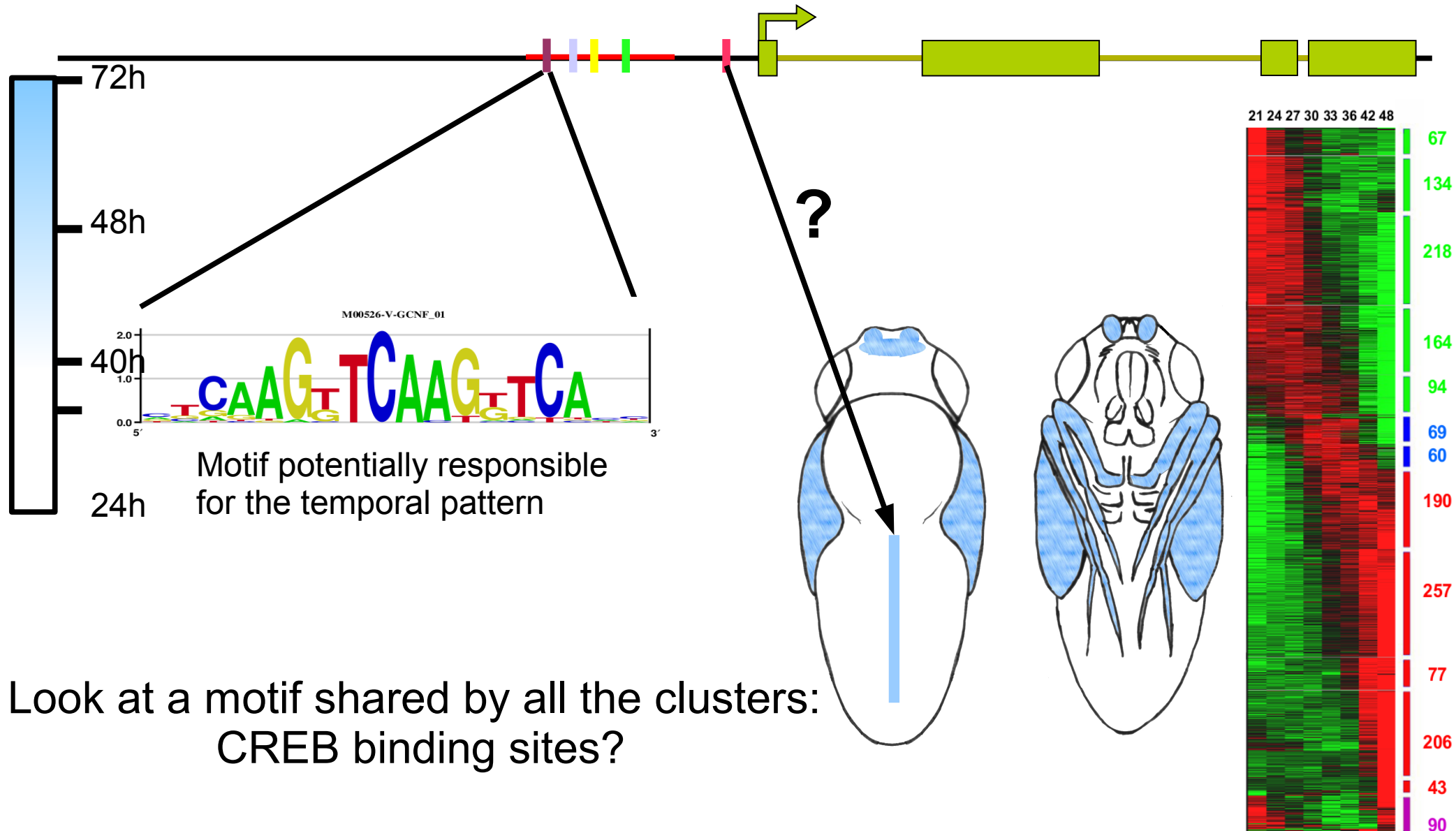


Results in ~ few weeks...

How is the spatial expression pattern encoded?



Motif responsible for the spatial expression in heart?

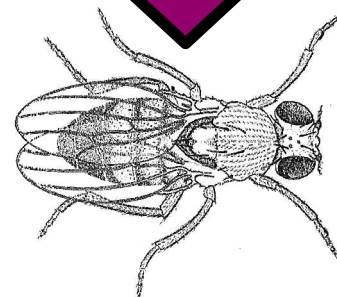




M. Semeriva's team



Thanks



Stein Aerts
Laboratory of Neurogenetics
Department of Human Genetics
VIB-KULeuven





M. Semeriva's team



Stein Aerts
Laboratory of Neurogenetics
Department of Human Genetics
VIB-KULeuven