Butterfly effect in molecular biology: the case of microRNAs

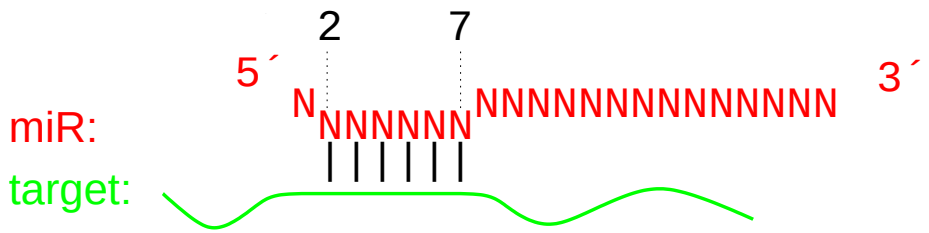
Hervé Seitz

IGH (CNRS)

April 24, 2012

microRNA target identification

microRNA target identification



$\longrightarrow$ the "seed"

Identification of miRNA targets

Computational programs for target prediction: look for seed matches in 3´ UTRs, select the ones that were conserved in evolution.

Identification of miRNA targets

Computational programs for target prediction: look for seed matches in 3´ UTRs, select the ones that were conserved in evolution.

Such short matches are very frequent (60 % of human coding genes seem to be targeted: Friedman *et al.*, 2009).

Identification of miRNA targets

Computational programs for target prediction: look for seed matches in 3´ UTRs, select the ones that were conserved in evolution.

Such short matches are very frequent (60 % of human coding genes seem to be targeted: Friedman *et al.*, 2009).

$\Longrightarrow$ miRNAs are implicated in every physiological process in animals.

Identification of miRNA targets

Computational programs for target prediction: look for seed matches in 3´ UTRs, select the ones that were conserved in evolution.

Such short matches are very frequent (60 % of human coding genes seem to be targeted: Friedman *et al.*, 2009).

$\Longrightarrow$ miRNAs are implicated in every physiological process in animals.

miRNA-mediated repression is very modest (usually $<$ 2-fold): have these binding sites been conserved because they trigger a butterfly effect (large consequences for minor causes)?

An alternative hypothesis

An alternative hypothesis

- Most computationally predicted targets are not *functionally* targeted (not repressed enough).

An alternative hypothesis

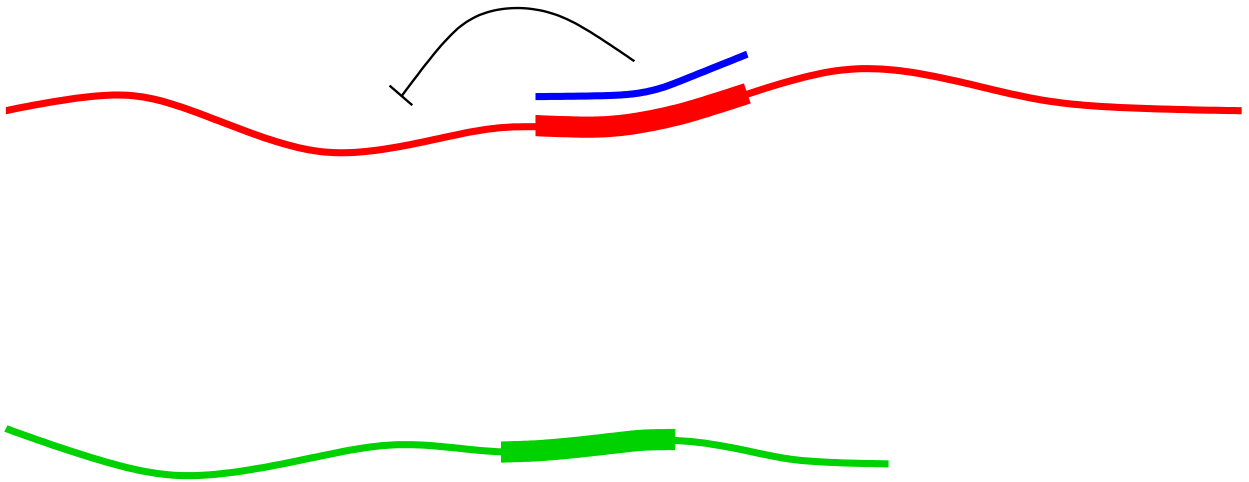
- Most computationally predicted targets are not *functionally* targeted (not repressed enough).
- Their phylogenetic conservation means that these binding sites have a function.

An alternative hypothesis

- Most computationally predicted targets are not *functionally* targeted (not repressed enough).
- Their phylogenetic conservation means that these binding sites have a function.
- That function could be to repress the miRNA by titrating it.

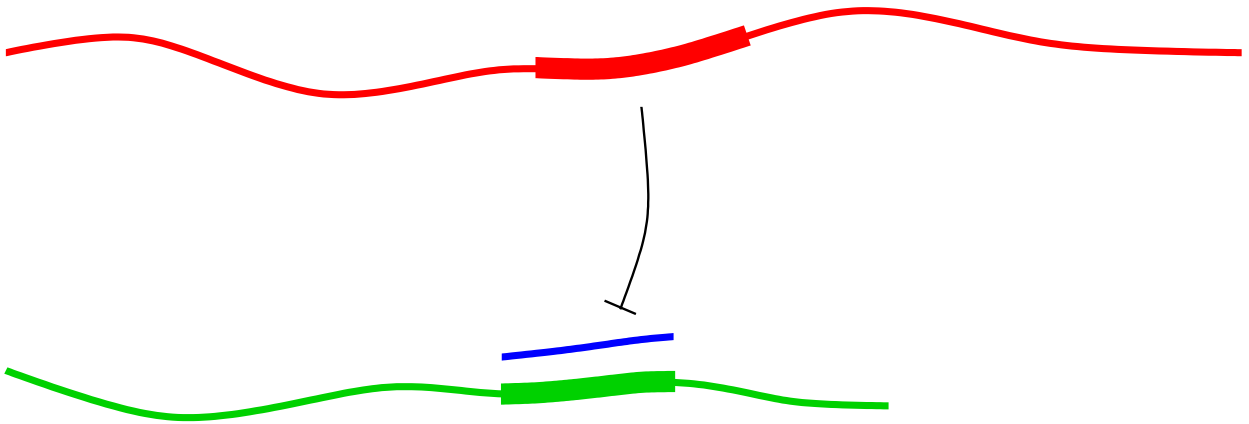
An alternative hypothesis

- Most computationally predicted targets are not *functionally* targeted (not repressed enough).
- Their phylogenetic conservation means that these binding sites have a function.
- That function could be to repress the miRNA by titrating it.



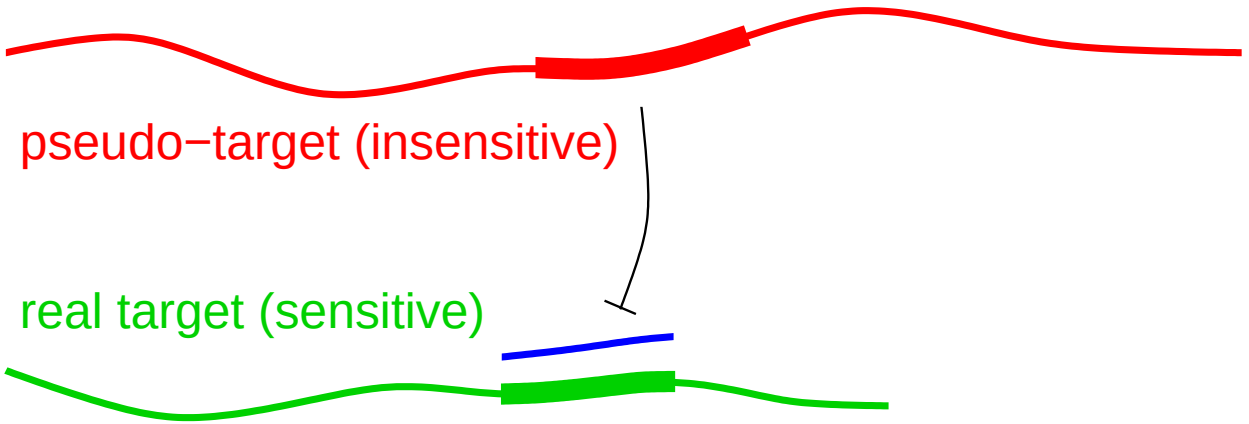
An alternative hypothesis

- Most computationally predicted targets are not *functionally* targeted (not repressed enough).
- Their phylogenetic conservation means that these binding sites have a function.
- That function could be to repress the miRNA by titrating it.



An alternative hypothesis

- Most computationally predicted targets are not *functionally* targeted (not repressed enough).
- Their phylogenetic conservation means that these binding sites have a function.
- That function could be to repress the miRNA by titrating it.



A new interpretation

The molecular event is the same (interaction between an mRNA and a miRNA), but the interpretation of that interaction is different:

- Current theory: the miRNA represses the mRNA.
- New hypothesis: the mRNA represses the miRNA (except for a few real targets).

Confronting predictions

Predicted by the current theory:

Confronting predictions

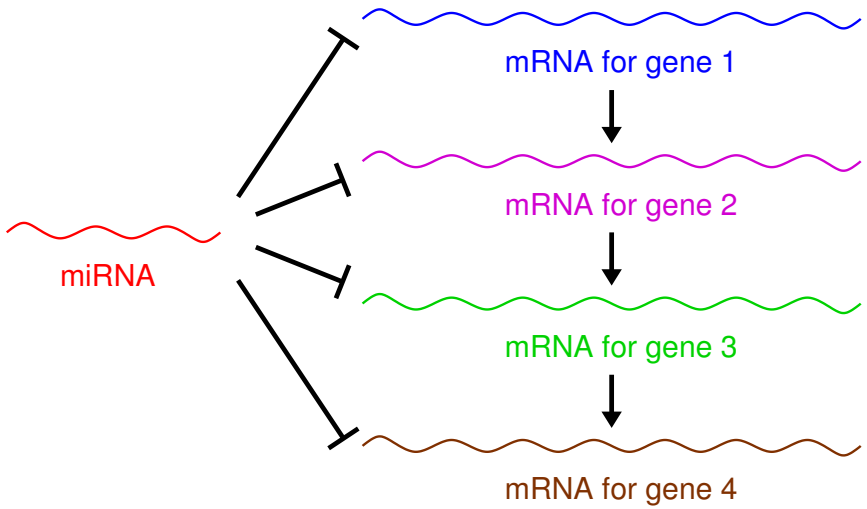
Predicted by the current theory:

1. Targets for a given miRNA often belong to the same biological pathways.

Confronting predictions

Predicted by the current theory:

1. Targets for a given miRNA often belong to the same biological pathways.



Confronting predictions

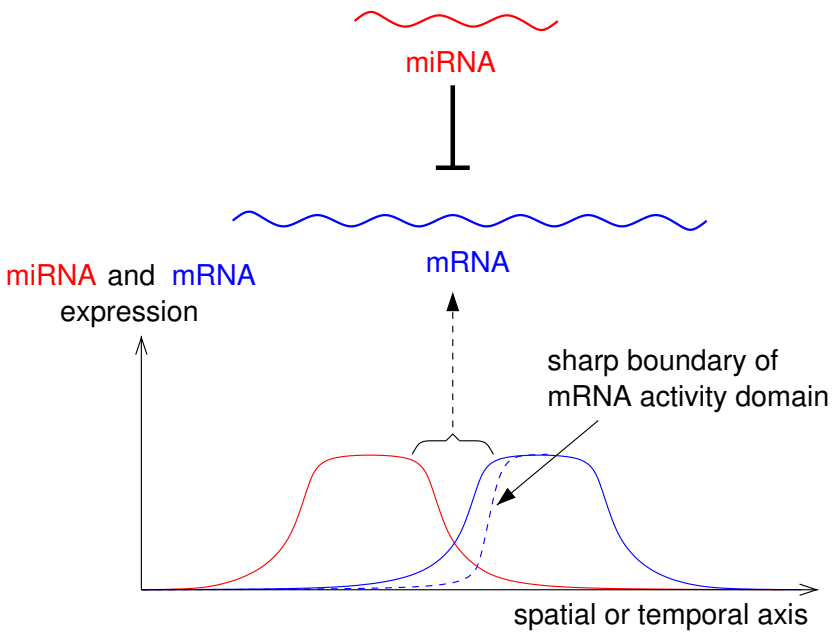
Predicted by the current theory:

1. Targets for a given miRNA often belong to the same biological pathways.
2. miRNA and predicted target expression patterns overlap at their boundaries.

Confronting predictions

Predicted by the current theory:

1. Targets for a given miRNA often belong to the same biological pathways.
2. miRNA and predicted target expression patterns overlap at their boundaries.



Confronting predictions

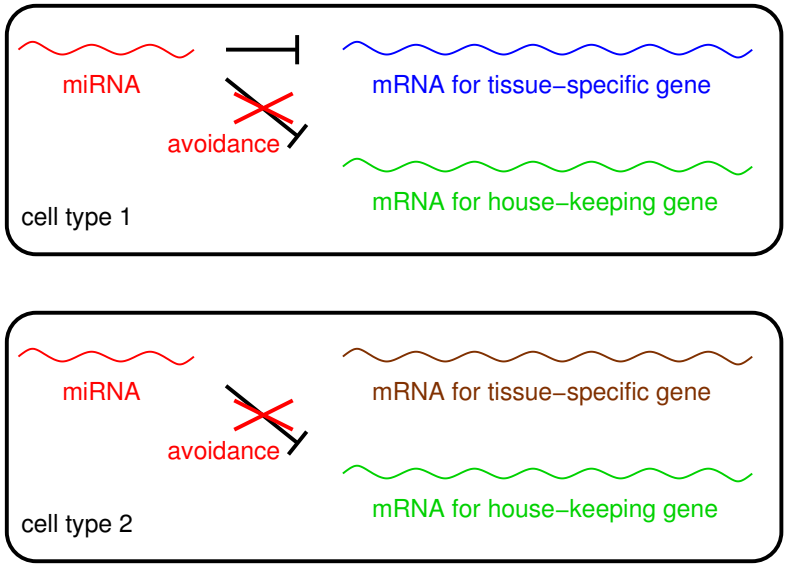
Predicted by the current theory:

1. Targets for a given miRNA often belong to the same biological pathways.
2. miRNA and predicted target expression patterns overlap at their boundaries.
3. Predicted targets are depleted in ubiquitously-expressed genes.

Confronting predictions

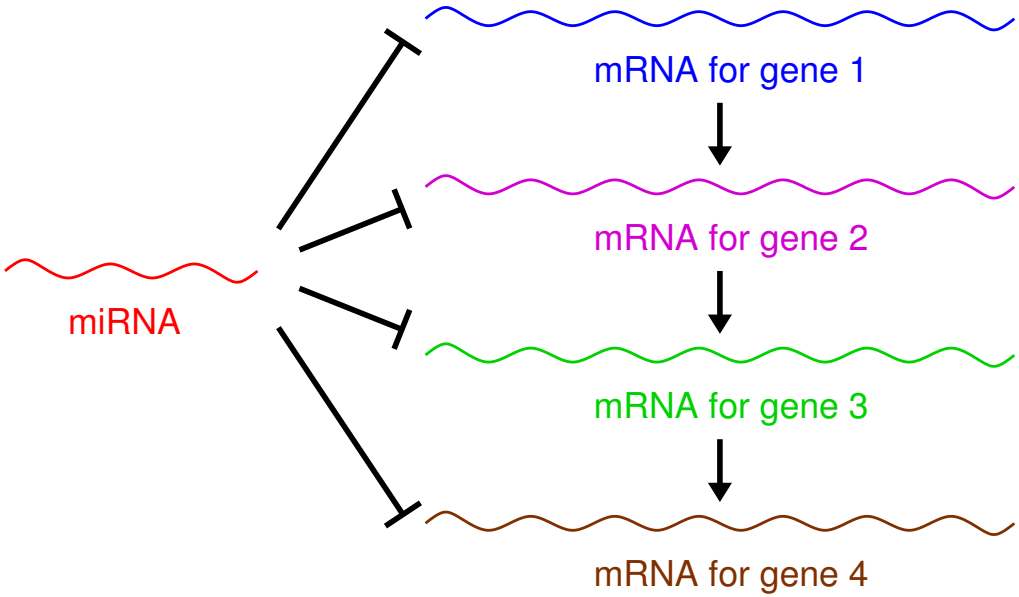
Predicted by the current theory:

1. Targets for a given miRNA often belong to the same biological pathways.
2. miRNA and predicted target expression patterns overlap at their boundaries.
3. Predicted targets are depleted in ubiquitously-expressed genes.

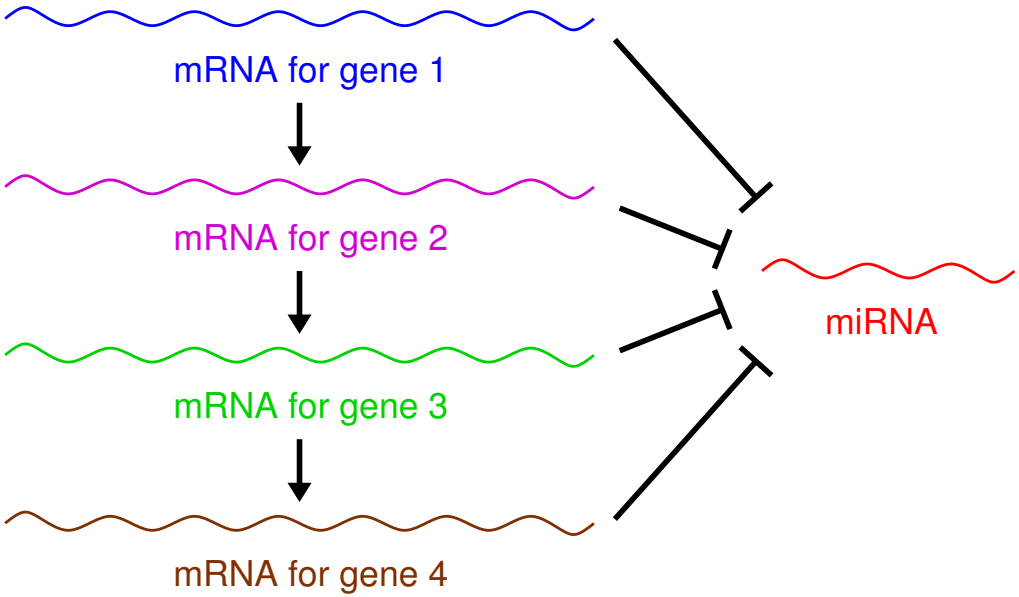


Confronting predictions

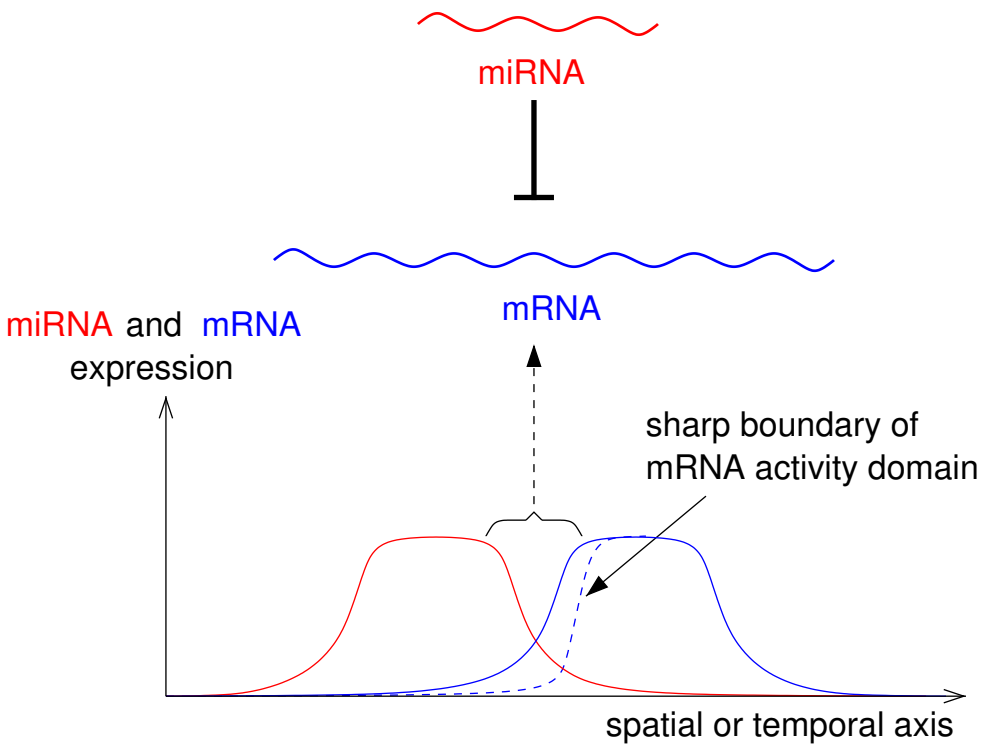
Confronting predictions



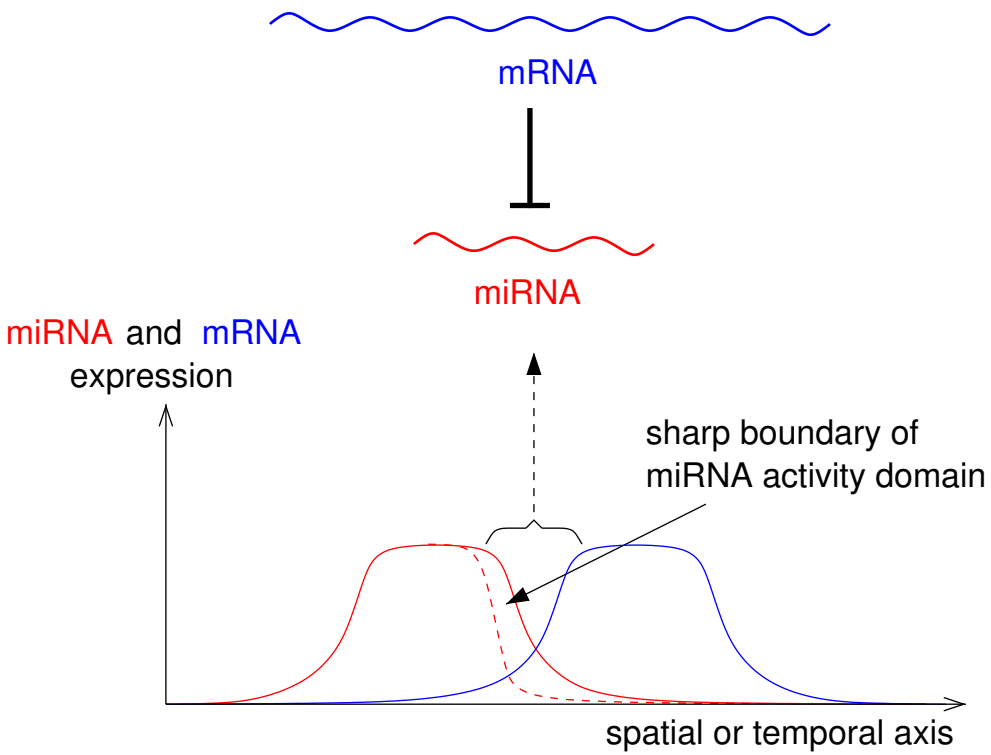
Confronting predictions



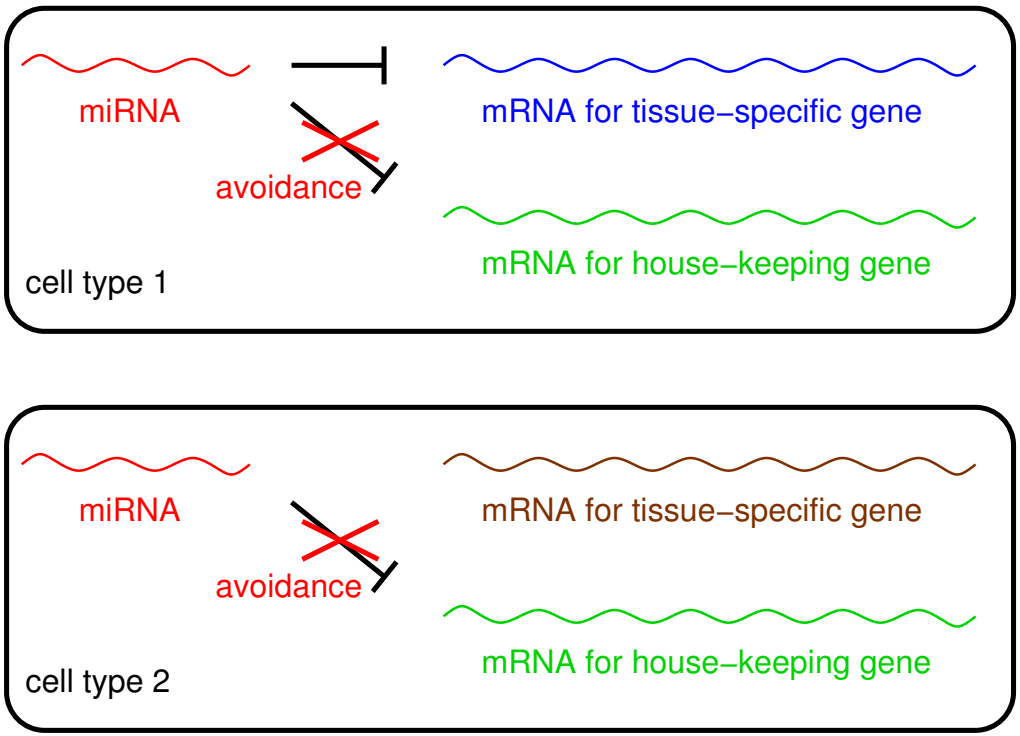
Confronting predictions



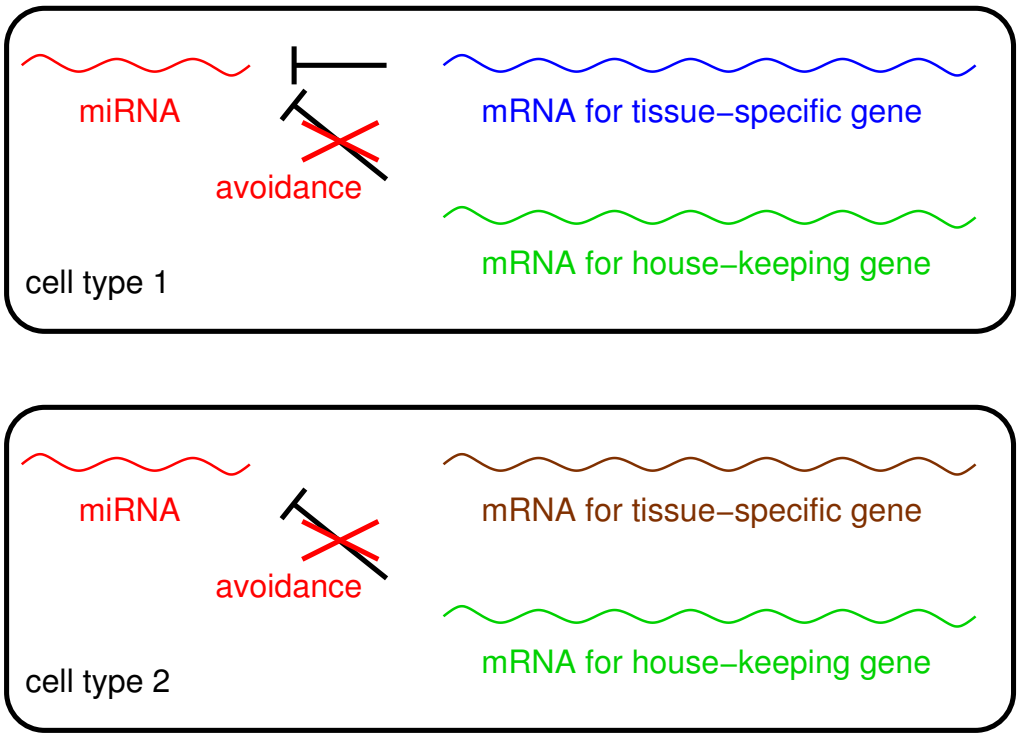
Confronting predictions



Confronting predictions



Confronting predictions



Discriminative predictions

According to the new hypothesis:

According to the current theory:

Discriminative predictions

According to the new hypothesis:

- Abundantly expressed pseudo-targets have a stronger miRNA-modulating effect, so their interaction with the miRNA should be better conserved

According to the current theory:

Discriminative predictions

According to the new hypothesis:

- Abundantly expressed pseudo-targets have a stronger miRNA-modulating effect, so their interaction with the miRNA should be better conserved $\Longrightarrow$ mRNA abundance should be positively correlated with miRNA binding site conservation.

According to the current theory:

Discriminative predictions

According to the new hypothesis:

- Abundantly expressed pseudo-targets have a stronger miRNA-modulating effect, so their interaction with the miRNA should be better conserved $\Longrightarrow$ mRNA abundance should be positively correlated with miRNA binding site conservation.

According to the current theory:

- Each target is a real target, its interaction is more or less conserved depending on the functional importance of the regulation (it is not expected to correlate with gene expression).

mRNA abundance and seed match conservation

Does mRNA abundance correlate positively with miRNA binding site conservation?

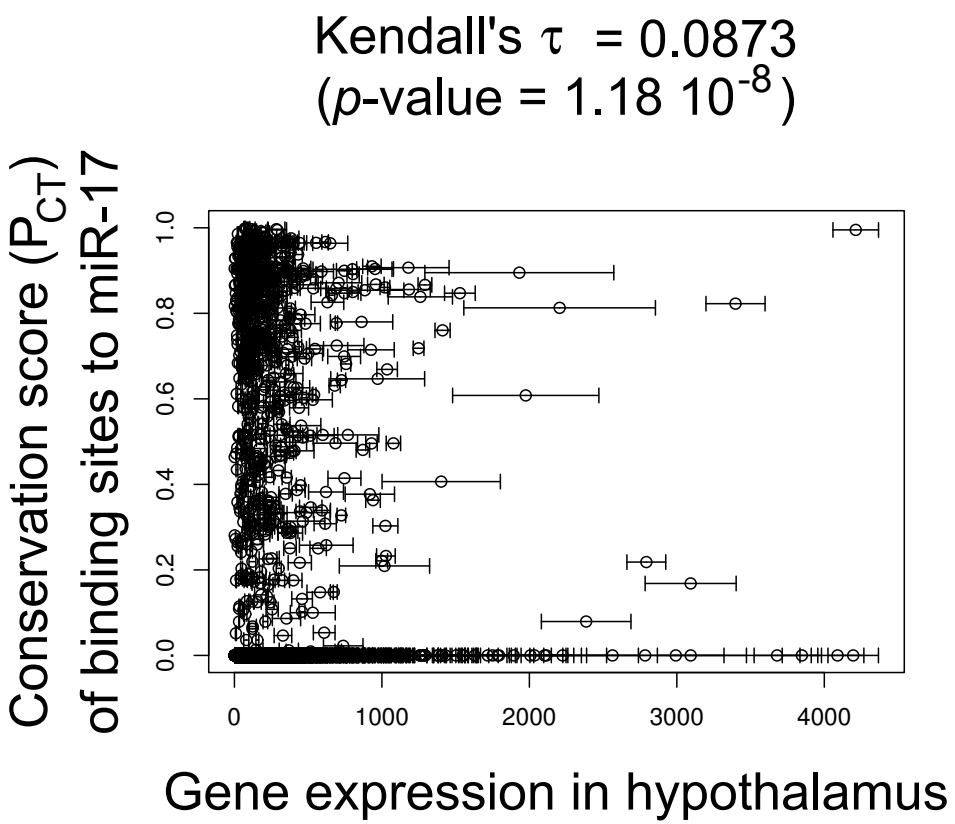
mRNA abundance and seed match conservation

Does mRNA abundance correlate positively with miRNA binding site conservation?

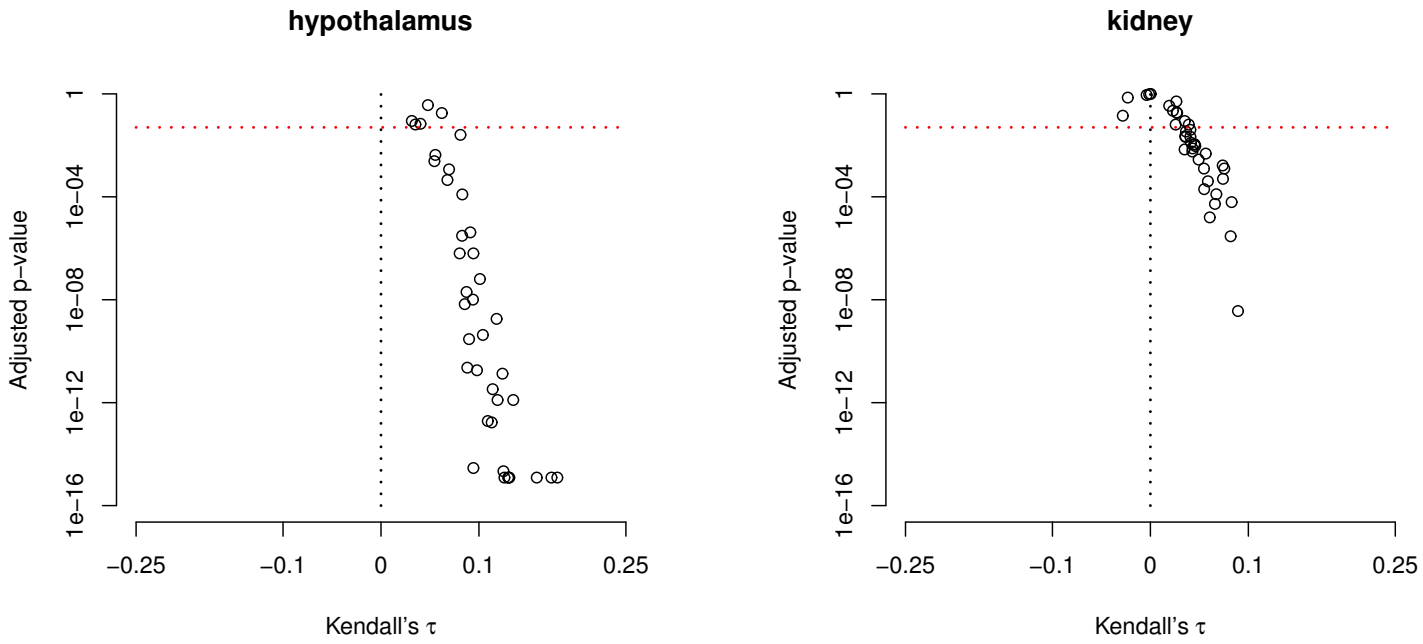
miRNA binding site conservation: measured by TargetScan's "P_{CT}" score. mRNA abundance in published microarray experiments.

▸ Supplementary data

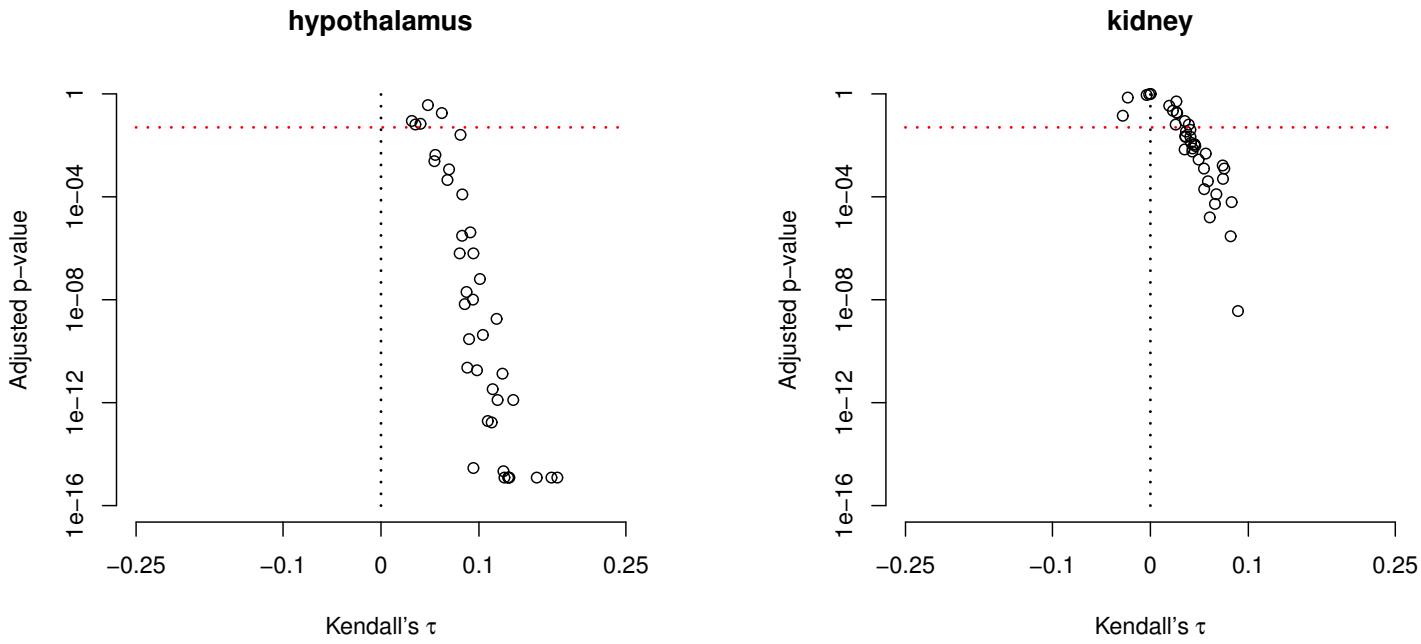
mRNA abundance and seed match conservation



mRNA abundance and seed match conservation



mRNA abundance and seed match conservation



+34 other mouse tissues, same results: ▸ Supplementary data

+33 fly tissues, same results: ▸ Supplementary data

Discriminative predictions

According to the new hypothesis:

According to the current theory:

Discriminative predictions

According to the new hypothesis:

- Pseudo-targets can be replaced (in their miRNA-modulating role) by other mRNAs with the same expression pattern; real targets cannot be replaced

According to the current theory:

Discriminative predictions

According to the new hypothesis:

- Pseudo-targets can be replaced (in their miRNA-modulating role) by other mRNAs with the same expression pattern; real targets cannot be replaced $\Longrightarrow$ genes with the most conserved miRNA binding sites should be enriched in dose-sensitive genes.

According to the current theory:

Discriminative predictions

According to the new hypothesis:

- Pseudo-targets can be replaced (in their miRNA-modulating role) by other mRNAs with the same expression pattern; real targets cannot be replaced $\Longrightarrow$ genes with the most conserved miRNA binding sites should be enriched in dose-sensitive genes.

According to the current theory:

- miRNA binding site conservation is not expected to correlate with dose-sensitivity.

Dose-sensitivity and seed match conservation

No high-throughput identification of dose-sensitive genes in metazoans.

Dose-sensitivity and seed match conservation

No high-throughput identification of dose-sensitive genes in metazoans.

- Huang *et al.*, 2010: prediction of the likelihood that a human gene is haplo-insufficient (heuristic).

Dose-sensitivity and seed match conservation

No high-throughput identification of dose-sensitive genes in metazoans.

- Huang *et al.*, 2010: prediction of the likelihood that a human gene is haplo-insufficient (heuristic).
- Counting the number of gene members in each gene family: measures the duplicability of ancestral genes.

Dose-sensitivity and seed match conservation

No high-throughput identification of dose-sensitive genes in metazoans.

- Huang *et al.*, 2010: prediction of the likelihood that a human gene is haplo-insufficient (heuristic).
- Counting the number of gene members in each gene family: measures the duplicability of ancestral genes.
- Liang *et al.*, 2008: estimating the structural robustness of the components of multi-protein complexes (protein "under-wrapping" leads to dose-sensitivity in invertebrates).

Dose-sensitivity and seed match conservation

Expected	Correlation coefficient
Positive correlation between haplo-insufficiency probability and site conservation	human: $+0.190$ ($p < 2.2\ \ 10^{-16}$)

Dose-sensitivity and seed match conservation

Expected	Correlation coefficient
Positive correlation between haplo-insufficiency probability and site conservation	human: $+0.190$ ($p < 2.2\ \ 10^{-16}$)
Negative correlation between family size and site conservation	worm: -0.057 ($p = 8.7\ \ 10^{-13}$) mouse: $+0.013$ ($p = 0.017$)

Dose-sensitivity and seed match conservation

Expected	Correlation coefficient
Positive correlation between haplo-insufficiency probability and site conservation	human: $+0.190$ ($p < 2.2\ \ 10^{-16}$)
Negative correlation between family size and site conservation	worm: -0.057 ($p = 8.7\ \ 10^{-13}$) mouse: $+0.013$ ($p = 0.017$)
Positive correlation between protein under-wrapping and site conservation	worm: $+0.070$ ($p = 0.64$) fly: $+46$ % ($p = 1.57\ \ 10^{-5}$)

Discriminative predictions

According to the new hypothesis:

According to the current theory:

Discriminative predictions

According to the new hypothesis:

- Pseudo-targets do not need to be coding to exhibit conserved miRNA binding sites.

According to the current theory:

Discriminative predictions

According to the new hypothesis:

- Pseudo-targets do not need to be coding to exhibit conserved miRNA binding sites $\Longrightarrow$ there could be non-coding RNAs whose only role is to sequester miRNAs.

According to the current theory:

Discriminative predictions

According to the new hypothesis:

- Pseudo-targets do not need to be coding to exhibit conserved miRNA binding sites $\Longrightarrow$ there could be non-coding RNAs whose only role is to sequester miRNAs.

According to the current theory:

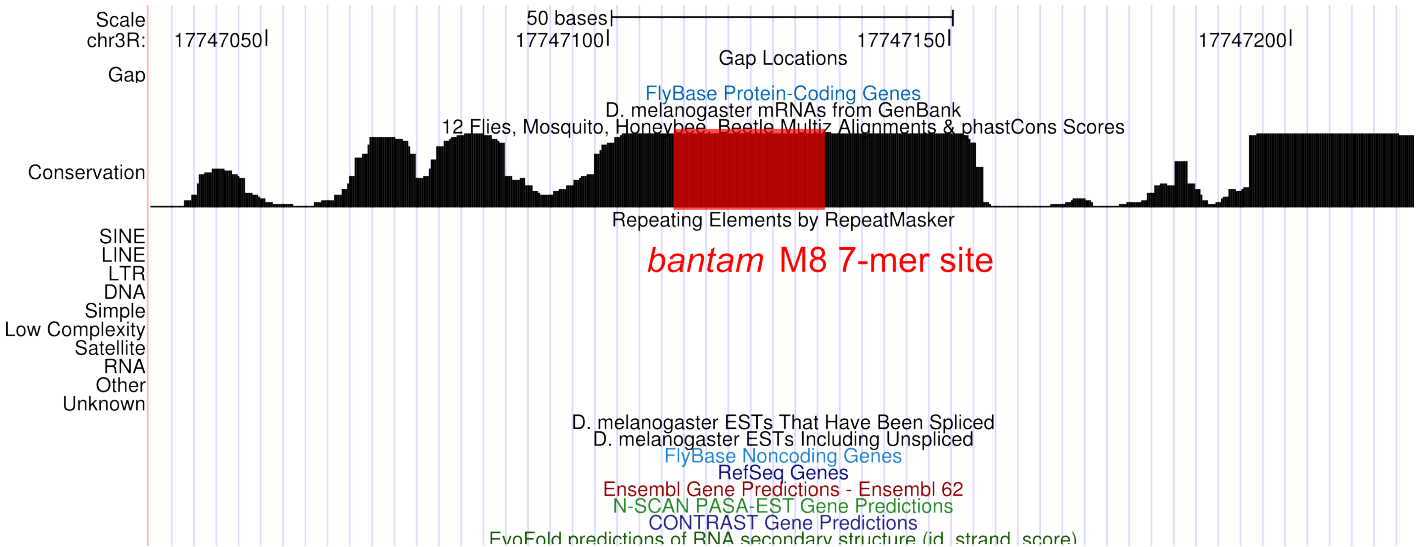
- Non-coding RNAs with conserved miRNA binding sites must have a function distinct from miRNA titration.

Non-coding pseudo-targets

Are there non-coding pseudo-targets?

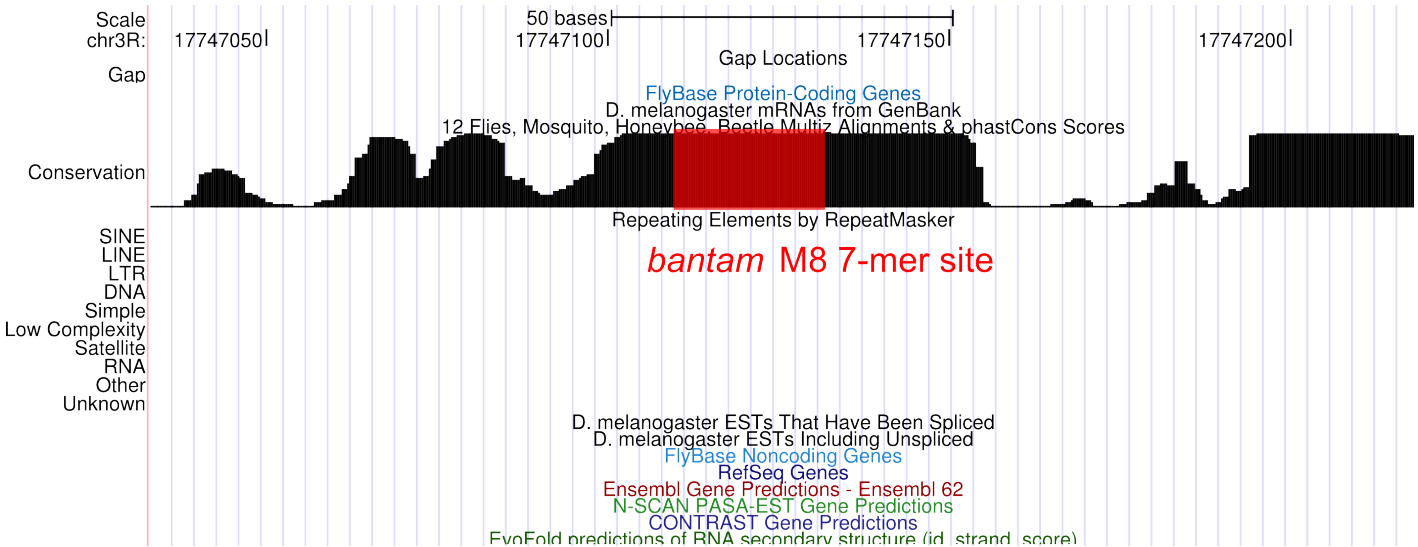
Non-coding pseudo-targets

Are there non-coding pseudo-targets?



Non-coding pseudo-targets

Are there non-coding pseudo-targets?



18 candidates identified, molecular characterization in progress (Natalia Pinzon Restrepo, Anna Sergeeva).

Discriminative predictions

According to the new hypothesis:

According to the current theory:

Discriminative predictions

According to the new hypothesis:

- Pseudo-target expression levels can fluctuate between individuals in a natural population (phenotype is robust).

According to the current theory:

Discriminative predictions

According to the new hypothesis:

- Pseudo-target expression levels can fluctuate between individuals in a natural population (phenotype is robust).

According to the current theory:

- miRNA targets are tightly regulated (inter-individual fluctuation should not exceed miRNA-guided repression).

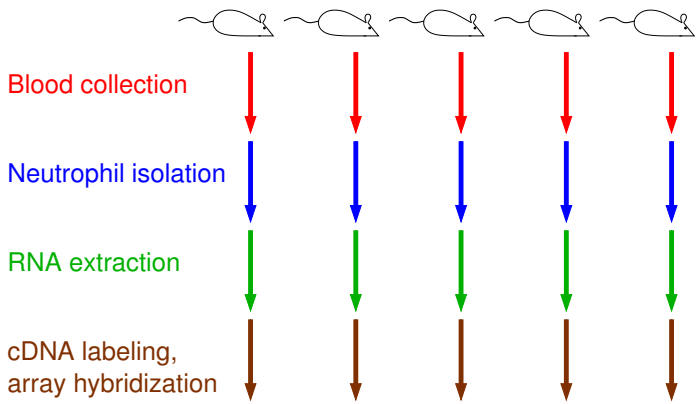
Natural variability vs. miRNA-guided repression

Natural variability vs. miRNA-guided repression

Baek *et al.*, 2008: quantification of miR-223-mediated repression in mouse neutrophils.

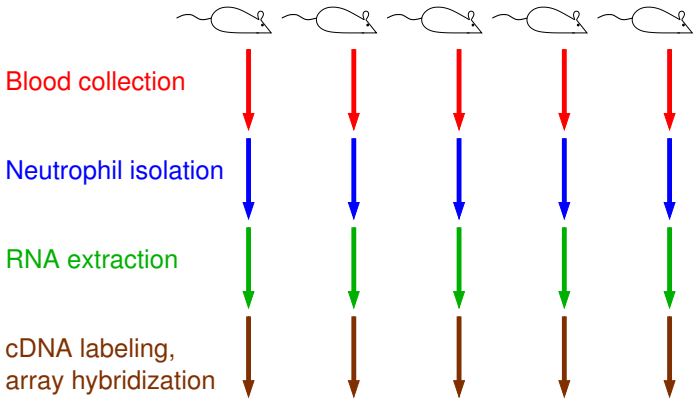
Natural variability vs. miRNA-guided repression

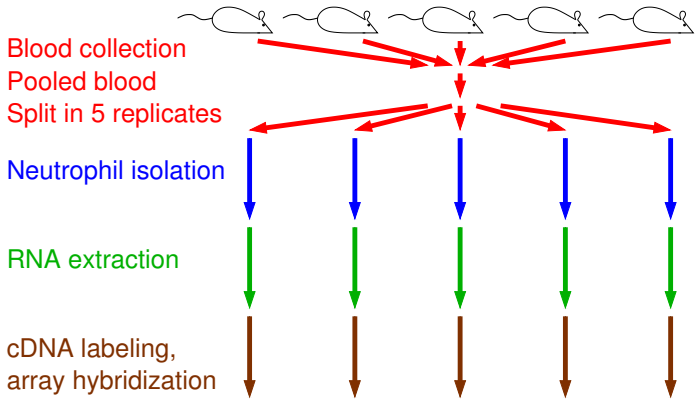
Baek *et al.*, 2008: quantification of miR-223-mediated repression in mouse neutrophils.



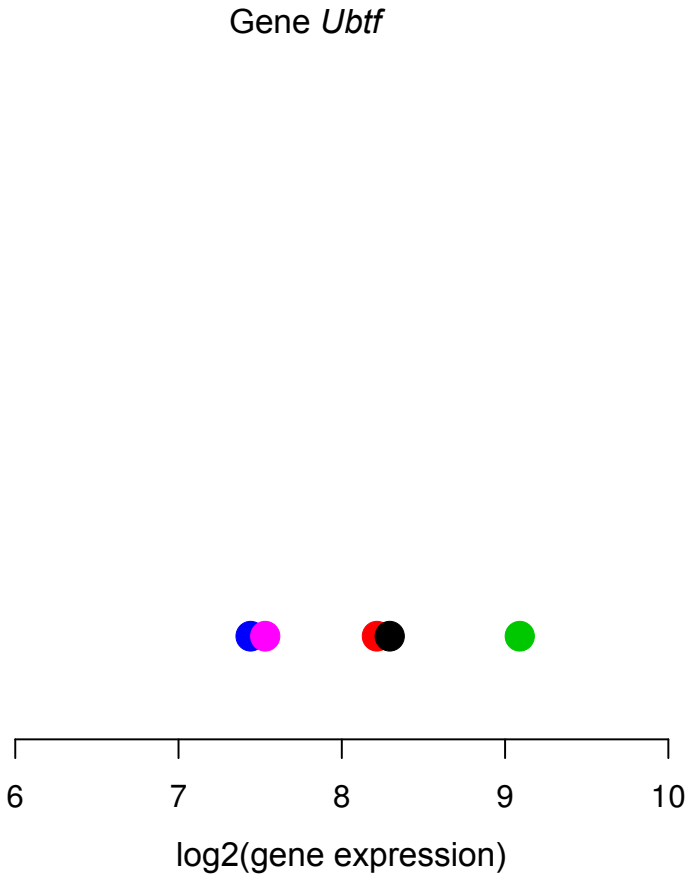
Natural variability vs. miRNA-guided repression

Baek *et al.*, 2008: quantification of miR-223-mediated repression in mouse neutrophils.

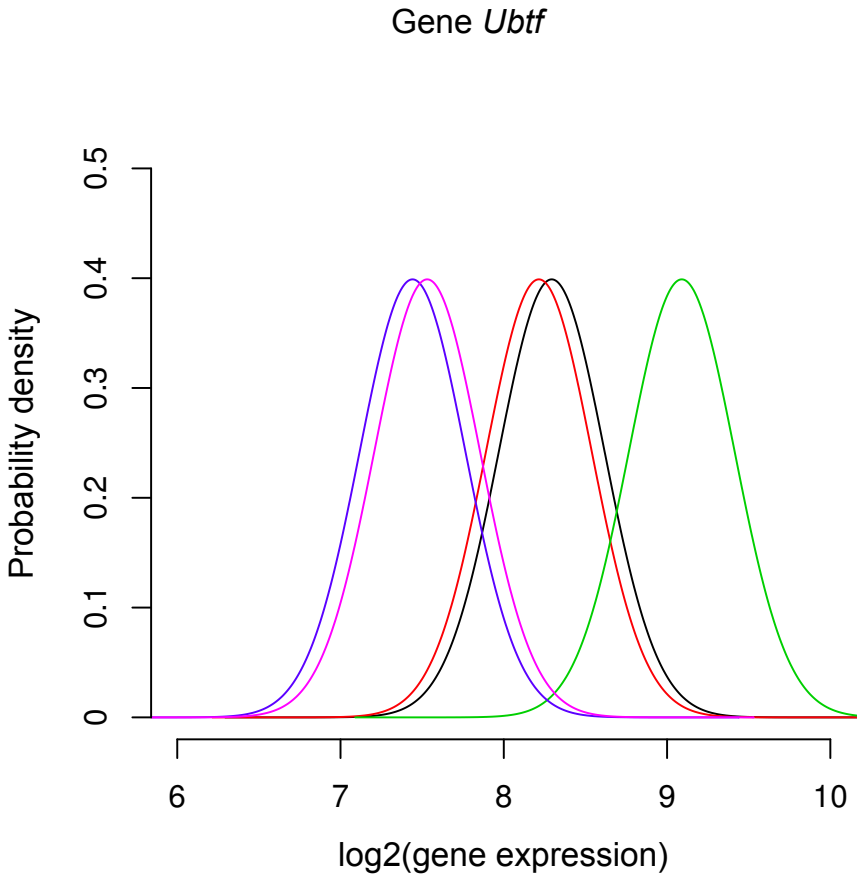




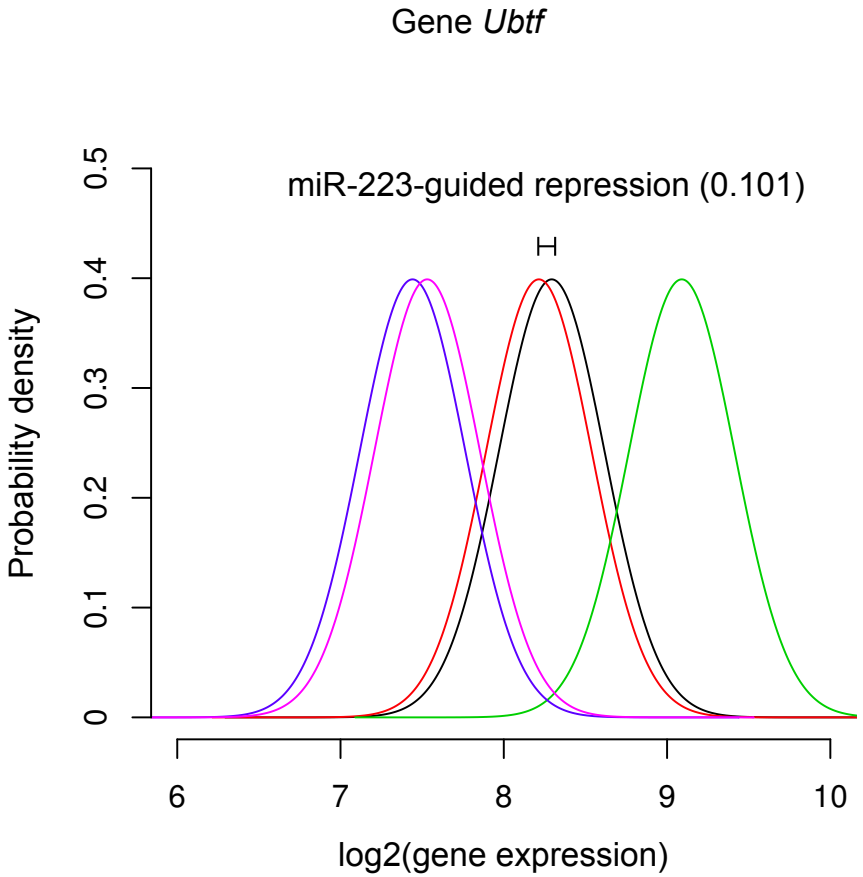
Natural variability vs. miRNA-guided repression



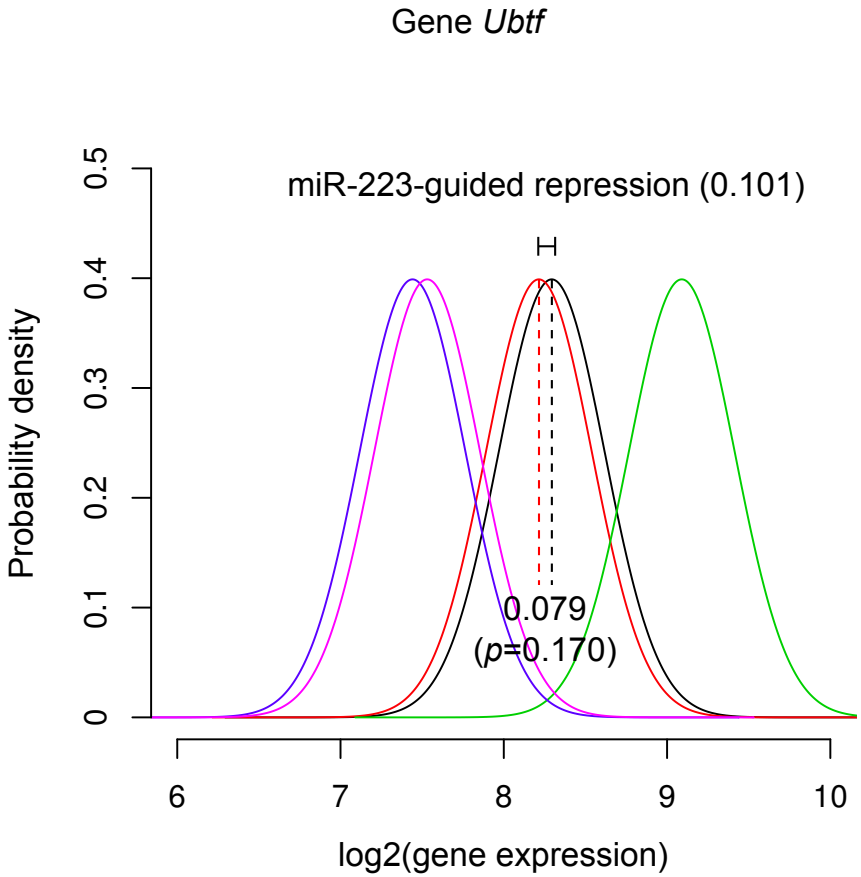
Natural variability vs. miRNA-guided repression



Natural variability vs. miRNA-guided repression

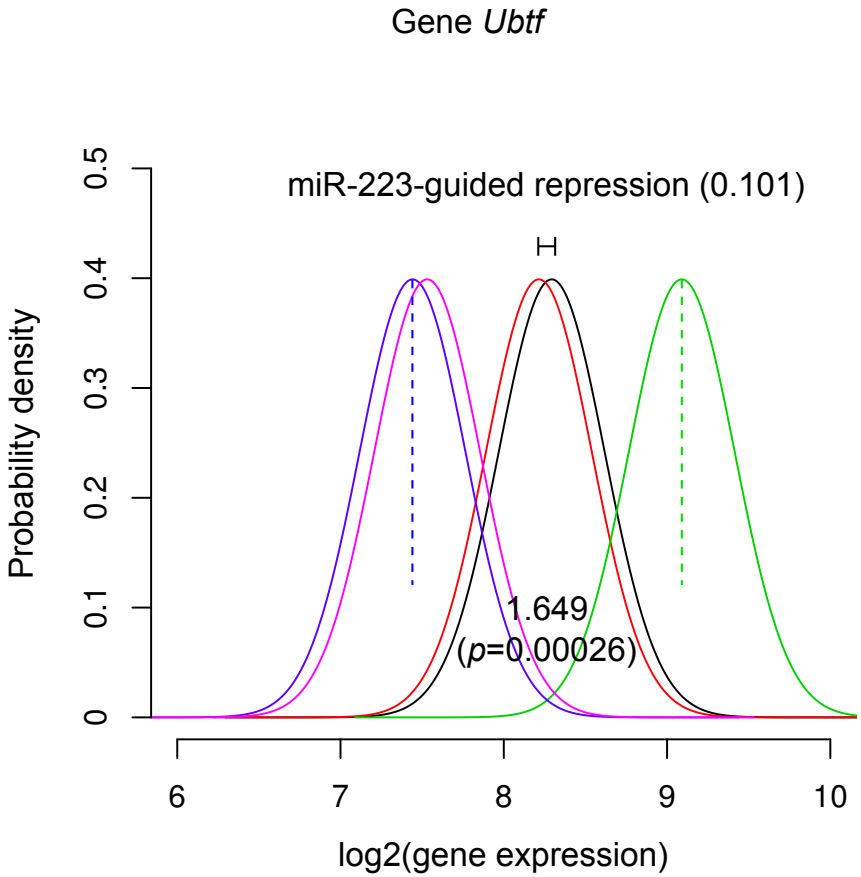


Natural variability vs. miRNA-guided repression



p: probability that the difference between two individual mice is smaller than miRNA-guided repression

Natural variability vs. miRNA-guided repression



p: probability that the difference between two individual mice is smaller than miRNA-guided repression

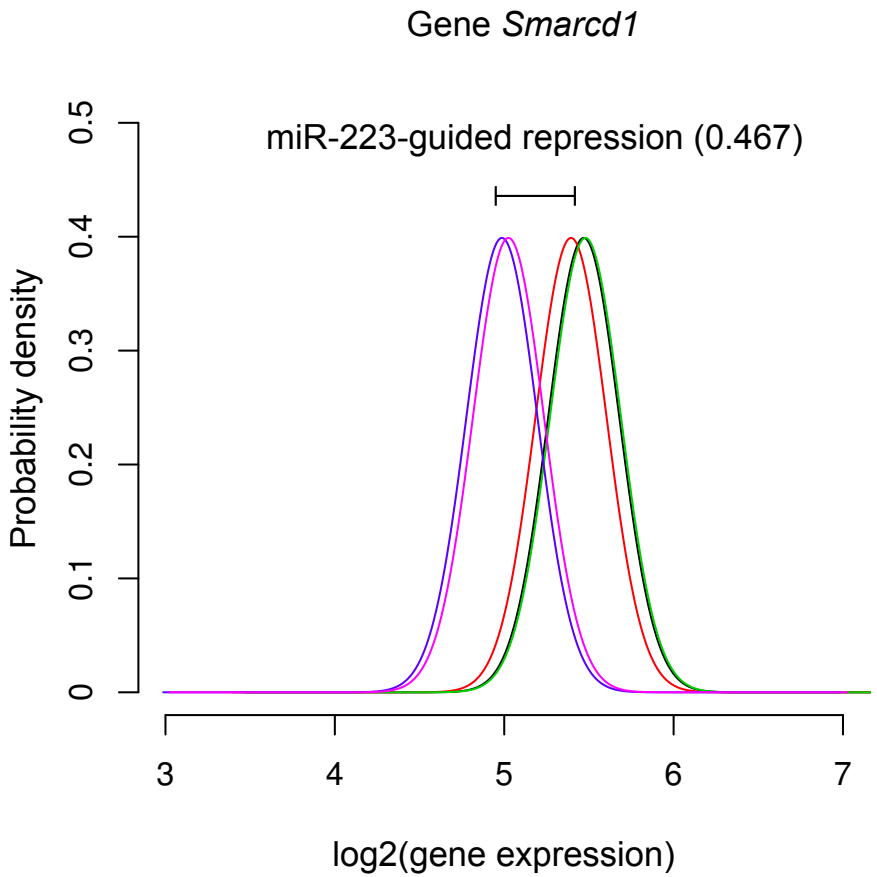
Natural variability vs. miRNA-guided repression

Natural variability vs. miRNA-guided repression

For 29 predicted targets out of 36:
inter-individual fluctuations across 5 wild-type mice exceeds miRNA-mediated regulation (p-value < 0.05).

Natural variability vs. miRNA-guided repression

For 29 predicted targets out of 36:
inter-individual fluctuations across 5 wild-type mice exceeds miRNA-mediated regulation (p-value < 0.05).



Natural variability vs. miRNA-guided repression

Gene	miR-223-guided fold-change (p-value)	Annotated function
Smarcd1	1.38 (0.283)	
Nlrp3	1.14 (0.264)	
Stim1	1.54 (0.246)	
Rasa1	1.44 (0.170)	
Carm1	0.824 (0.167)	
Acsl3	1.26 (0.116)	
Rab10	1.47 (0.0551)	

Natural variability vs. miRNA-guided repression

Gene	miR-223-guided fold-change (p-value)	Annotated function
Smarcd1	1.38 (0.283)	Chromatin structure regulation
Nlrp3	1.14 (0.264)	Component of inflammasome
Stim1	1.54 (0.246)	Ca^{2+} signal transduction
Rasa1	1.44 (0.170)	Cell motility
Carm1	0.824 (0.167)	Post-translational modification
Acsl3	1.26 (0.116)	Fatty acid metabolism
Rab10	1.47 (0.0551)	Signal transduction

Natural variability vs. miRNA-guided repression

Gene	miR-223-guided fold-change (p-value)	Annotated function
Smarcd1	1.38 (0.283)	Chromatin structure regulation
Nlrp3	1.14 (0.264)	Component of inflammasome
Stim1	1.54 (0.246)	Ca^{2+} signal transduction
Rasa1	1.44 (0.170)	Cell motility
Carm1	0.824 (0.167)	Post-translational modification
Acsl3	1.26 (0.116)	Fatty acid metabolism
Rab10	1.47 (0.0551)	Signal transduction

Generalization

The pseudo-target theory can be generalized to every regulator (transcription factors, RNA-binding proteins,...).

Generalization

The pseudo-target theory can be generalized to every regulator (transcription factors, RNA-binding proteins,...). Molecular biology and comparative genomics can identify intermolecular interactions, but they don't tell whether the function is to regulate the target, or to titrate the regulator.

Generalization

Transcription factor Zfy-overexpressing mice (Hélène Royo, Turner lab, MRC, London). Transgene is overexpressed in the testis from day 13.5, cells enter apoptosis at day 15.5. mRNA quantification by microarray at these two stages.

Generalization

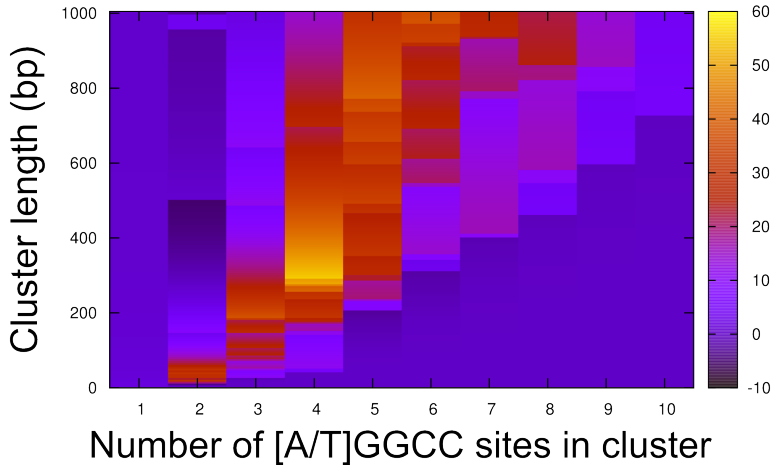
Transcription factor Zfy-overexpressing mice (Hélène Royo, Turner lab, MRC, London). Transgene is overexpressed in the testis from day 13.5, cells enter apoptosis at day 15.5. mRNA quantification by microarray at these two stages.

Zfy binds the [A/T]GGCC pentamer.

Generalization

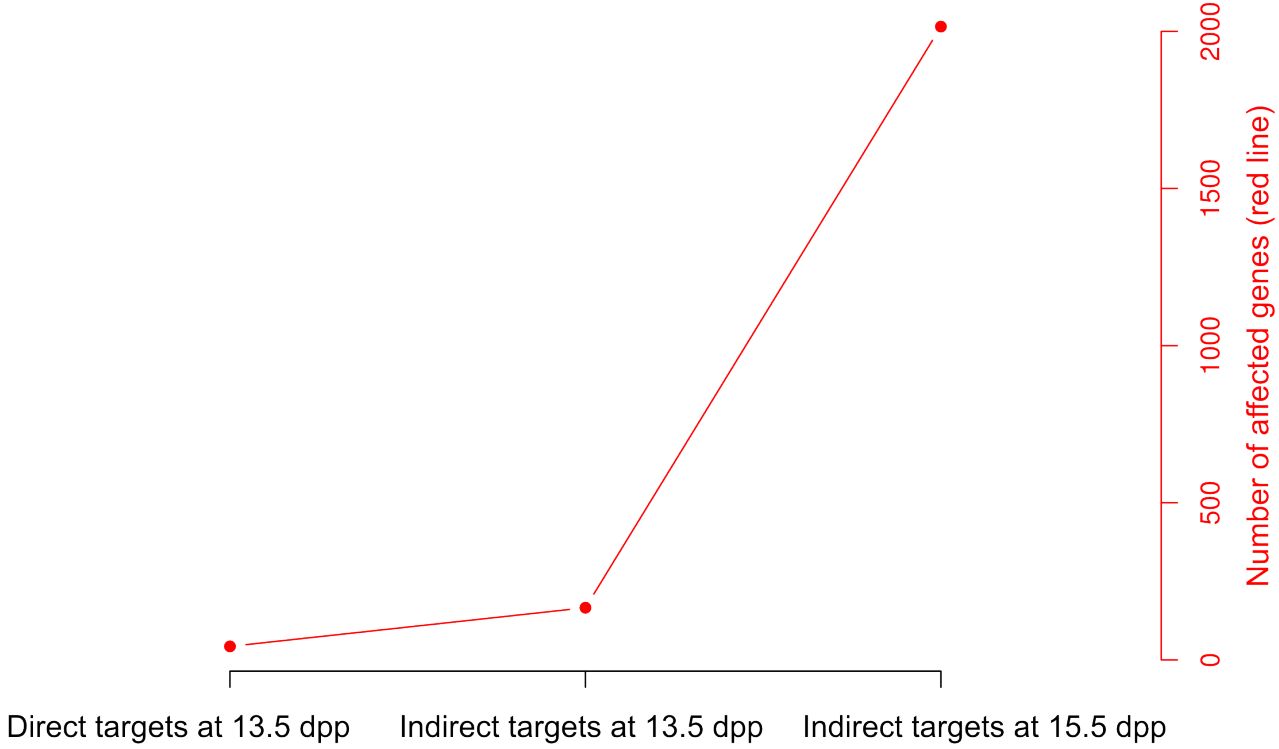
Transcription factor Zfy-overexpressing mice (Hélène Royo, Turner lab, MRC, London). Transgene is overexpressed in the testis from day 13.5, cells enter apoptosis at day 15.5. mRNA quantification by microarray at these two stages.

(% of up-regulated genes with cluster) - (% of unaffected genes with cluster)

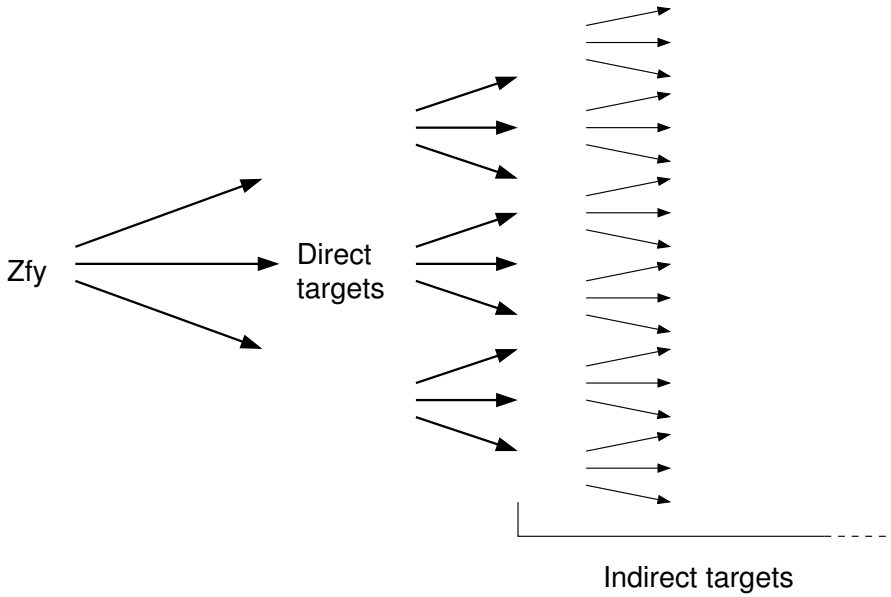


▸ Supplementary data

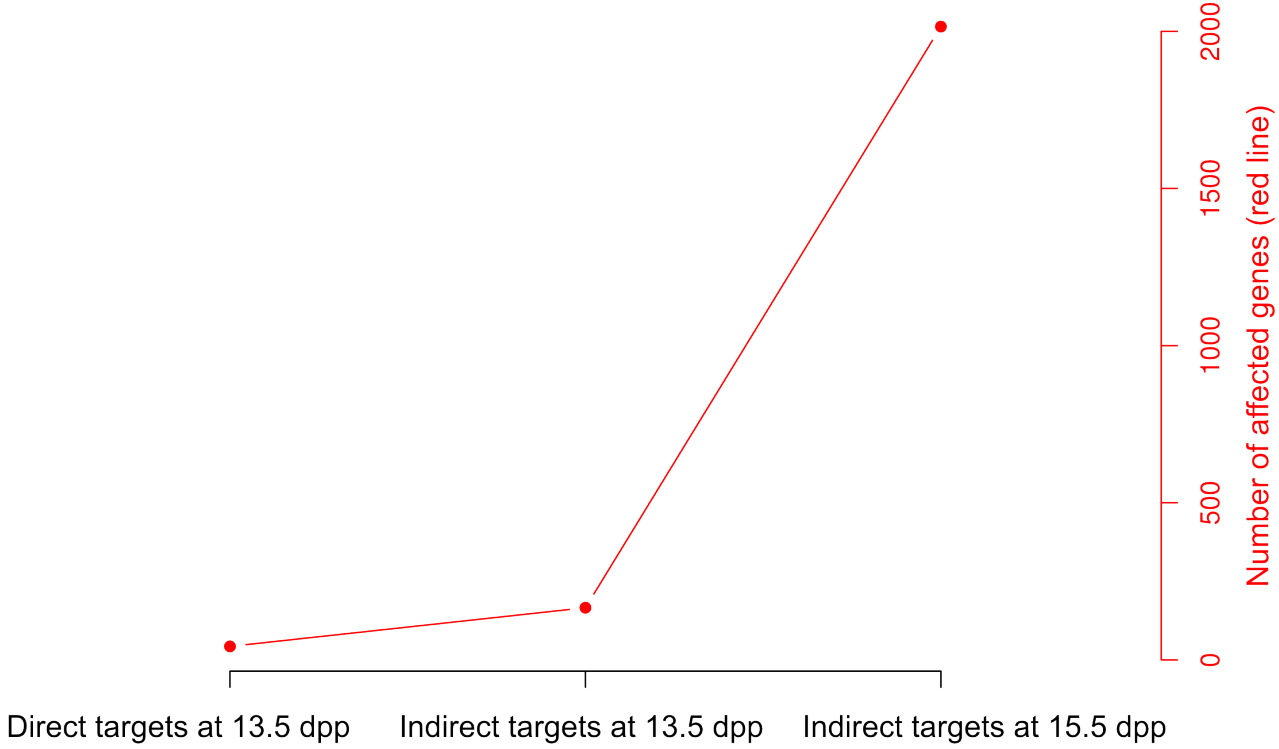
Generalization



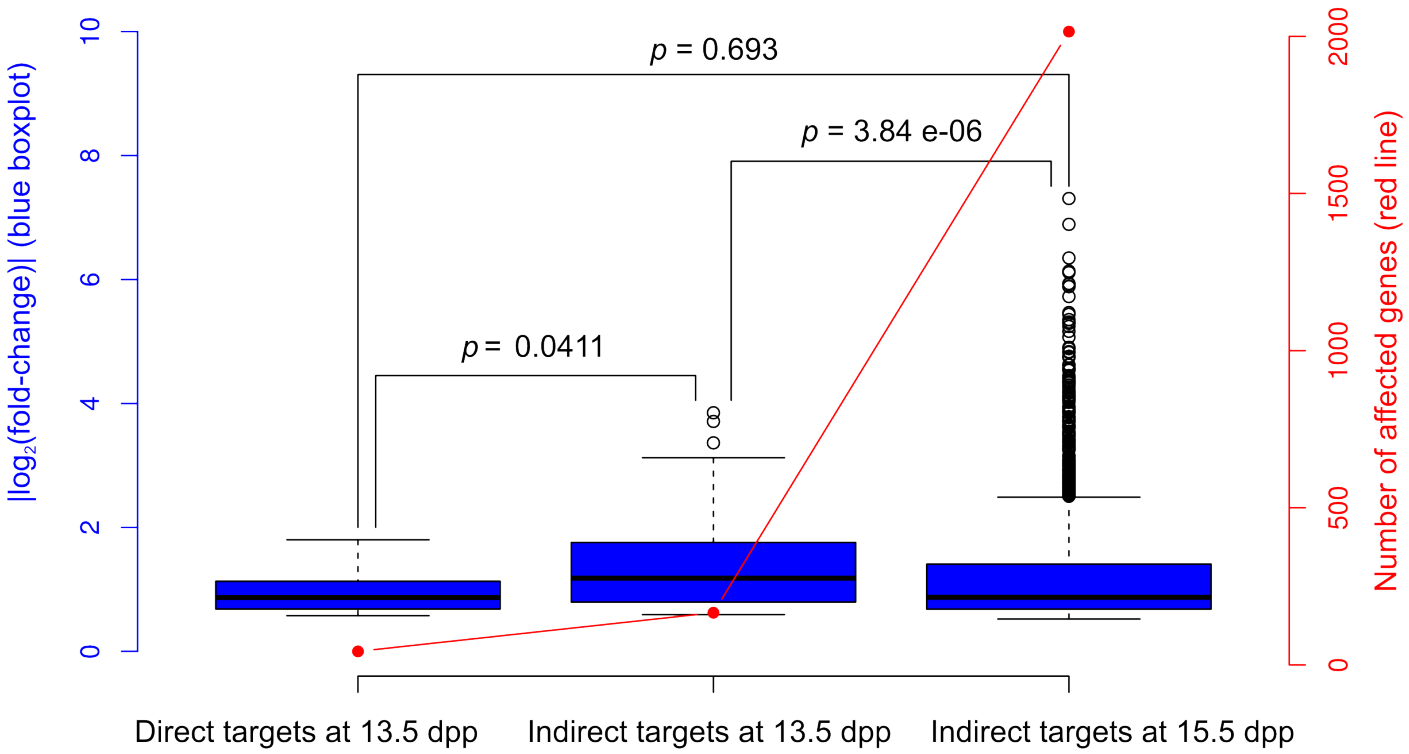
Generalization



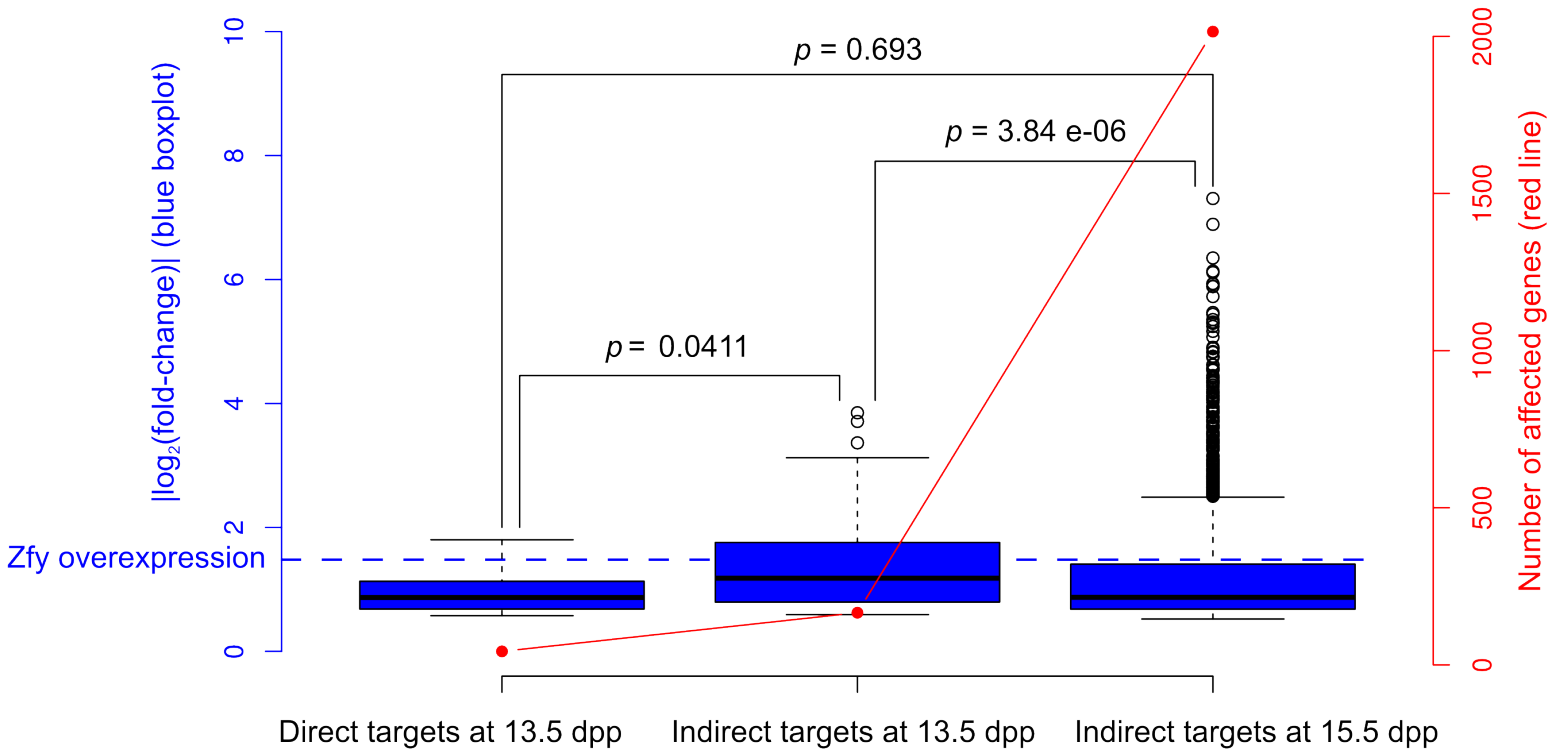
Generalization



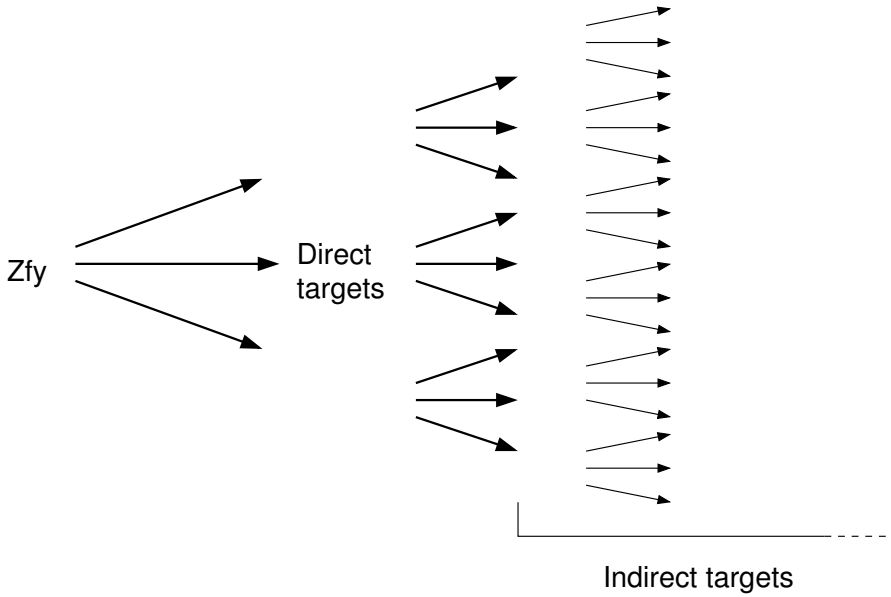
Generalization



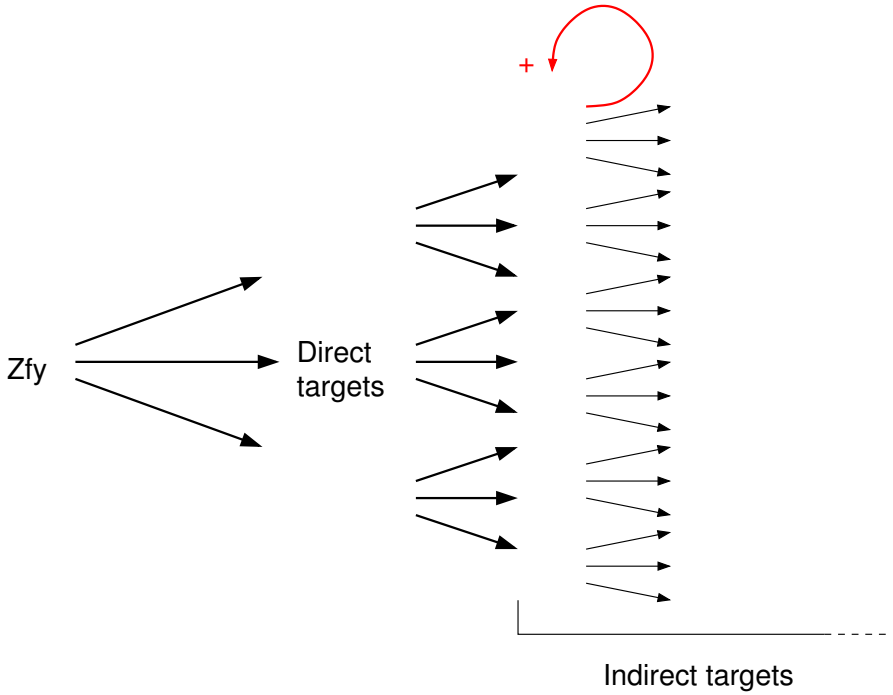
Generalization



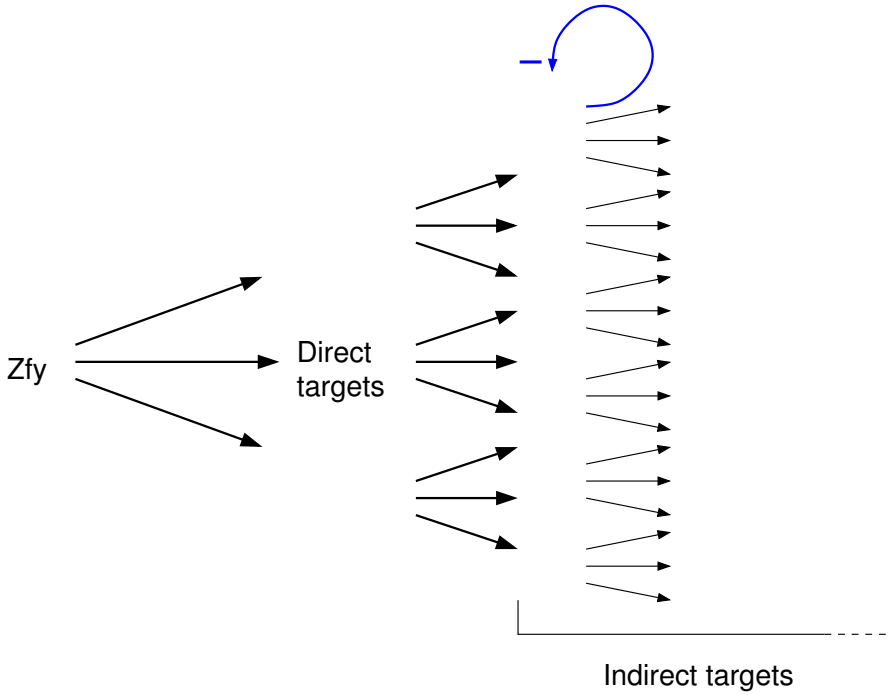
Generalization



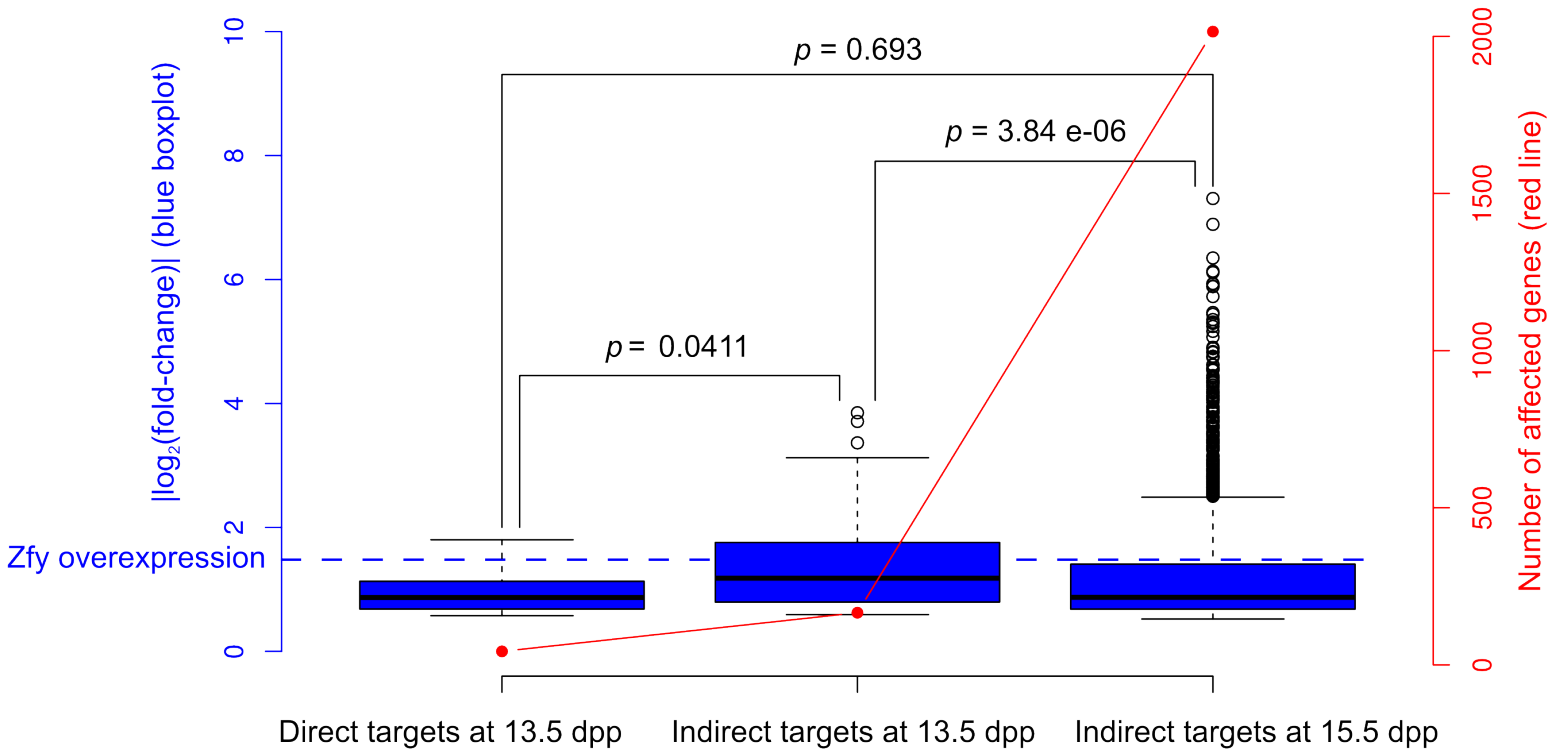
Generalization



Generalization



Generalization



Conclusion: revisiting miRNA target definition

Conclusion: revisiting miRNA target definition

Every measurable change in gene expression does not translate into a macroscopic, evolutionarily selectable phenotype.

Conclusion: revisiting miRNA target definition

Every measurable change in gene expression does not translate into a macroscopic, evolutionarily selectable phenotype.

Cells can modulate miRNA activity in a cell type-specific manner (integrating expression information from hundreds of genes on a single regulator)

Conclusion: revisiting "gene regulation" definition

High-throughput identification of transcription factor or RNA-binding protein targets: thousands of genes are bound, many of them are not under selective pressure to keep these binding sites.

Conclusion: revisiting "gene regulation" definition

High-throughput identification of transcription factor or RNA-binding protein targets: thousands of genes are bound, many of them are not under selective pressure to keep these binding sites.

Microscopic events which are neutral in evolutionary terms.

Conclusion: revisiting "gene regulation" definition

High-throughput identification of transcription factor or RNA-binding protein targets: thousands of genes are bound, many of them are not under selective pressure to keep these binding sites.

Microscopic events which are neutral in evolutionary terms.

Are there pseudo-targets for these other regulators?

Conclusion: revisiting "gene regulation" definition

High-throughput identification of transcription factor or RNA-binding protein targets: thousands of genes are bound, many of them are not under selective pressure to keep these binding sites.

Microscopic events which are neutral in evolutionary terms.

Are there pseudo-targets for these other regulators?

A possible methodology: quantification of gene expression changes after perturbation of a regulator (attenuation and amplification along the affected pathways).

Conclusion: revisiting "gene regulation" definition

High-throughput identification of transcription factor or RNA-binding protein targets: thousands of genes are bound, many of them are not under selective pressure to keep these binding sites.

Microscopic events which are neutral in evolutionary terms.

Are there pseudo-targets for these other regulators?

A possible methodology: quantification of gene expression changes after perturbation of a regulator (attenuation and amplification along the affected pathways).

A central feature of biological systems: their robustness to external insults. Hard to reconcile with the extreme sensitivity required for fine-tuning (the "butterfly effect" has probably been counter-selected).

Acknowledgements

Anna Sergeeva

Laura Martinez

Natalia Pinzon Restrepo

Acknowledgements

Anna Sergeeva

Laura Martinez

Natalia Pinzon Restrepo

Hélène Royo and James Turner (MRC, London, UK)
Jessy Presumey and Florence Apparailly (INM, Montpellier, France)

Acknowledgements

Anna Sergeeva

Laura Martinez

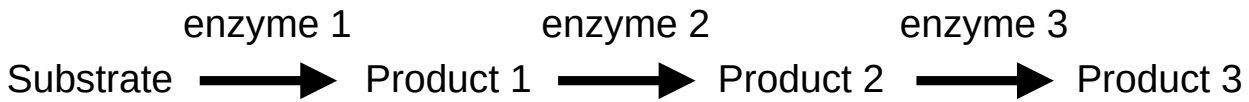
Natalia Pinzon Restrepo

Hélène Royo and James Turner (MRC, London, UK)
Jessy Presumey and Florence Apparailly (INM, Montpellier, France)

Supplementary data

- Robustness of biological pathways
- Conservative approximations in the assessment of abundance/conservation correlation
- Positive correlation between gene expression and conservation of miRNA binding sites
- Identification of Zfy binding sites

Second paradox

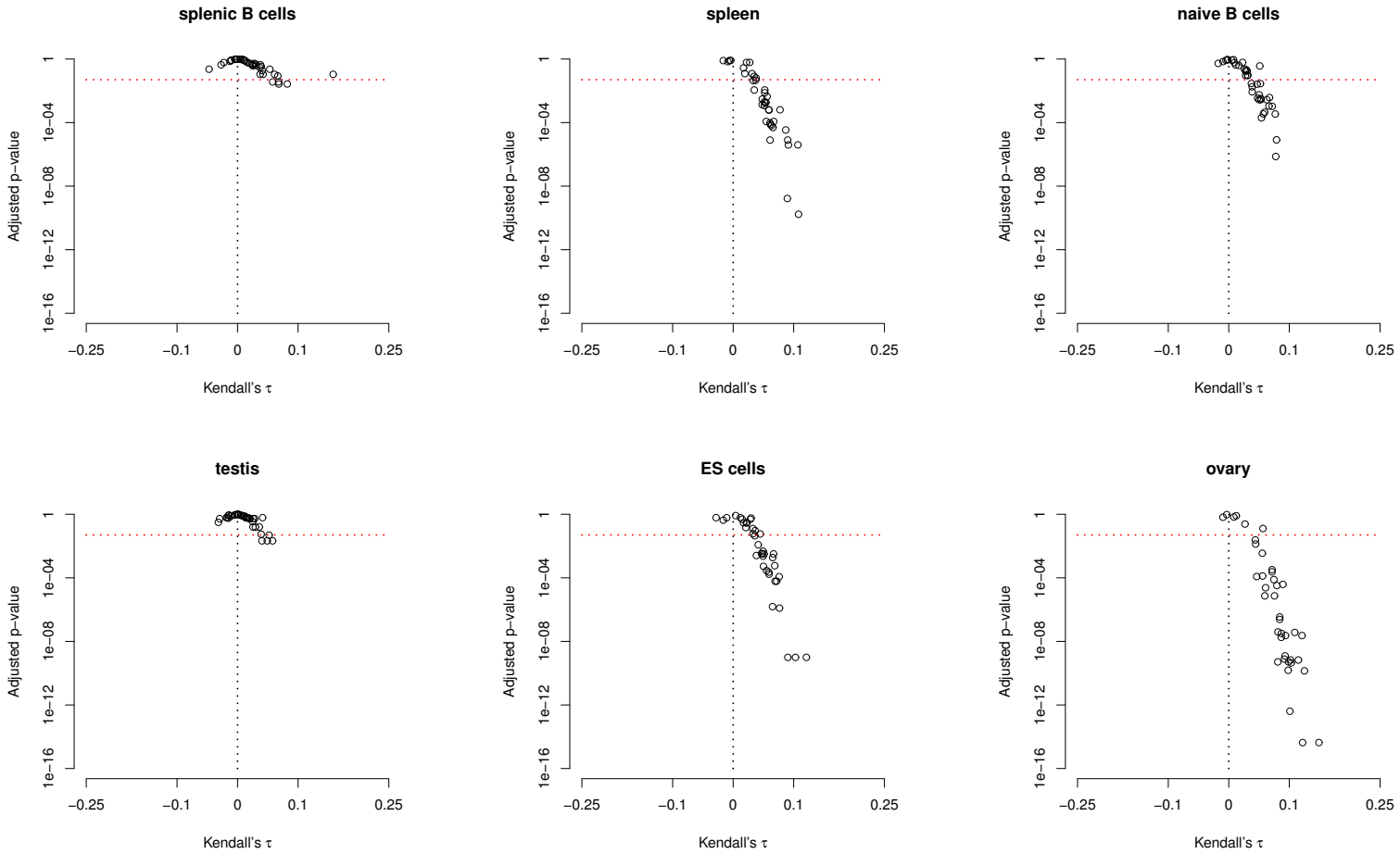


mRNA abundance and seed match conservation

- Most predicted targets are expected to be pseudo-targets (conservative approximation: we will consider every predicted target).
- miRNA binding site should correlate with that mRNA's abundance in the cells where miRNA titration is beneficial (conservative approximation: we will consider whole tissues and organs).
- Poorly abundant mRNAs (without a real titration effect on the miRNA) should not exhibit such correlation (conservative approximation: we will consider every mRNA).

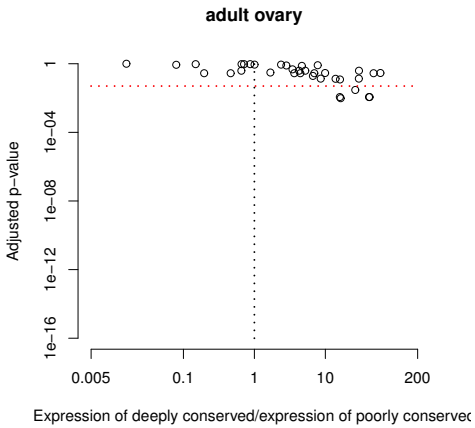
◂ Return

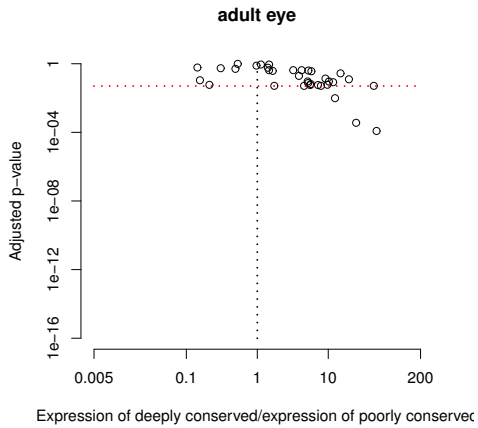
mRNA abundance and seed match conservation

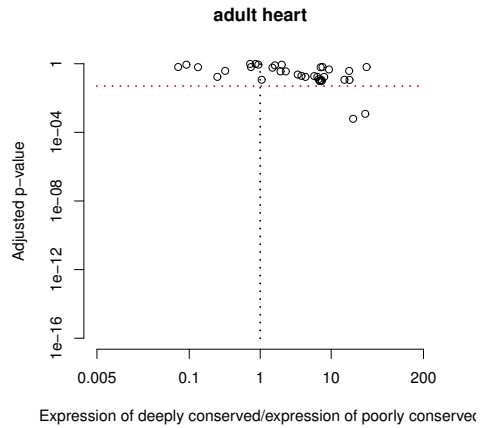


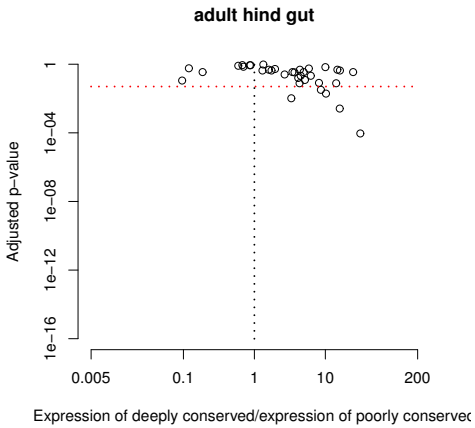
Return

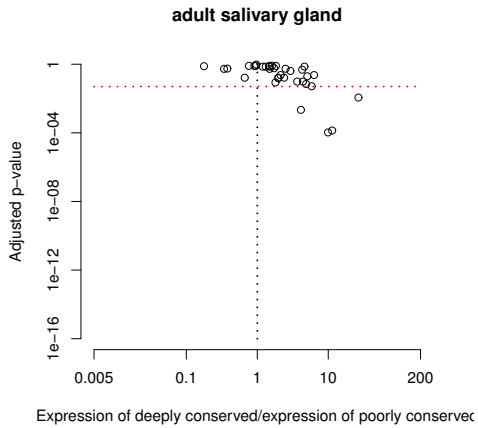
mRNA abundance and seed match conservation

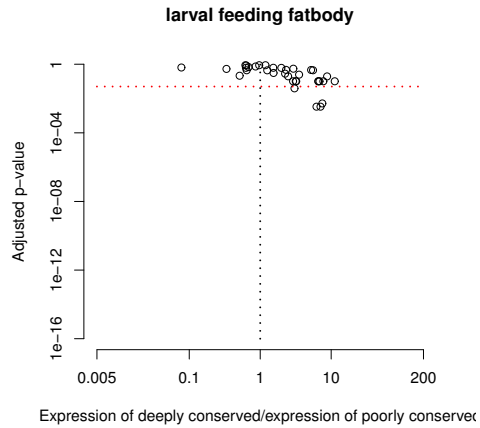








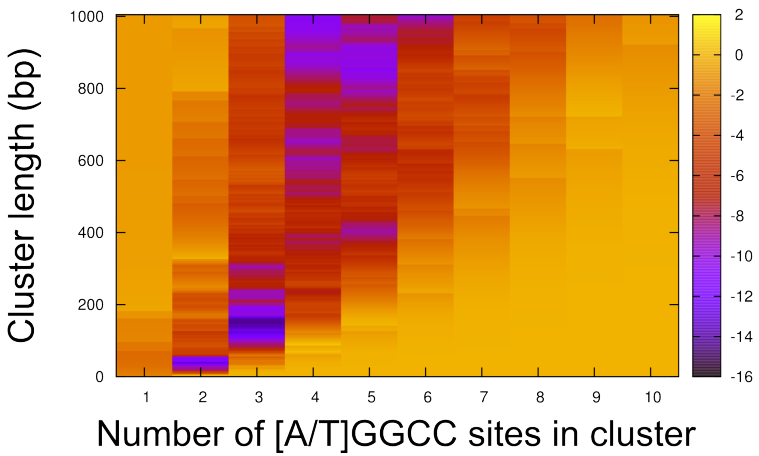




◂ Return

Generalization

(% of down-regulated genes with cluster) - (% of unaffected genes with cluster)



◂ Return