Core vertebrate long range cis-regulatory interactions revealed by zebrafish – human comparative genomics

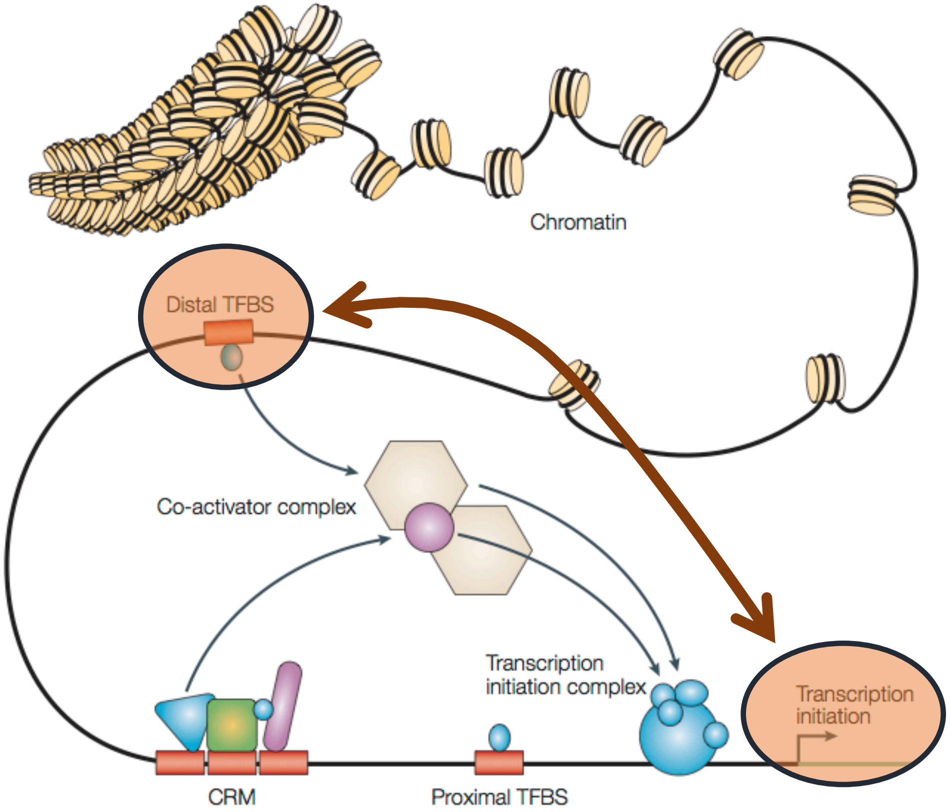
Yves Clément & Hugues Roest Crollius

GTGC Lyon 01/07/2016

IBENS
Institut de Biologie de l'Ecole Normale Supérieure

Dyogen Group

Enhancers as long range cis-regulatory regions



- Distal transcription factor binding sites = enhancers
- Our goals?
  - Identify these regions
  - Which genes they regulate
  - Which regulatory interactions are conserved: which functions are involved

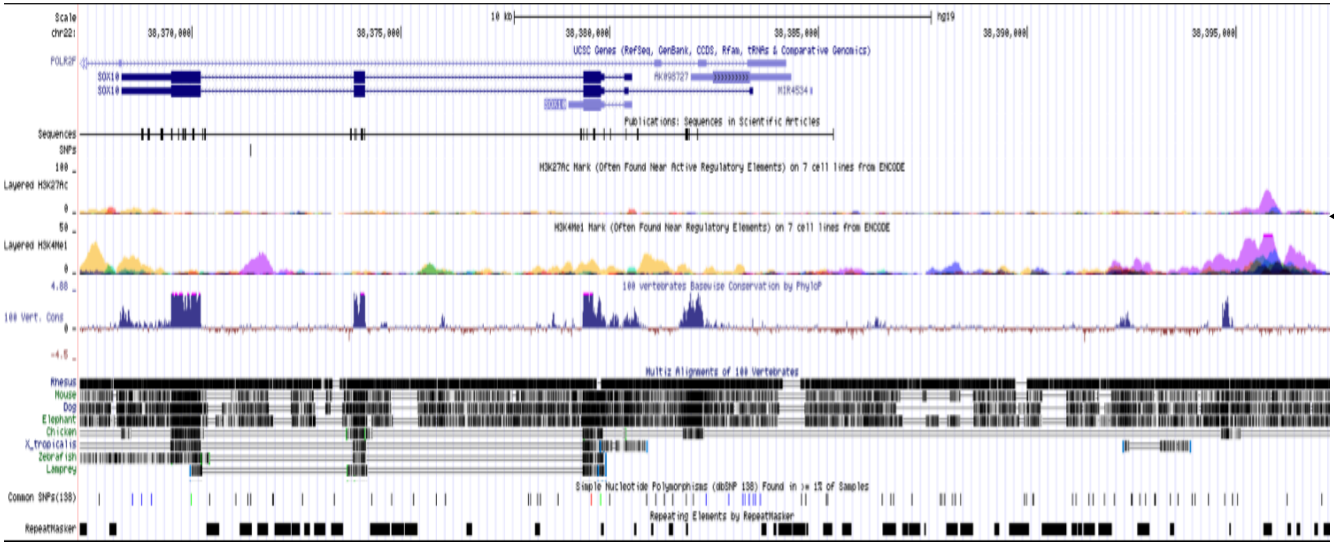
Wassermann & Sandelin, 2004

Identifying putative enhancers analyzing various types of signals

Conservation (HARs, ultra-conserved elements…)

Histone marks

- H3K27ac
- H3K4me1

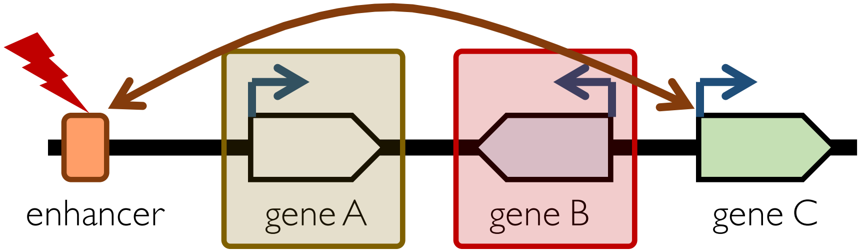


~ 30 kb around the SOX10 gene human chromosome 22

Other types of data (CAGE), integration of signals

UCSC Genome browser

Identifying target genes



Which one is the target gene?

- Direct way: mutagenesis
- Simple way: nearest gene
- Indirect ways:
    - Epigenome editing using TALE technology

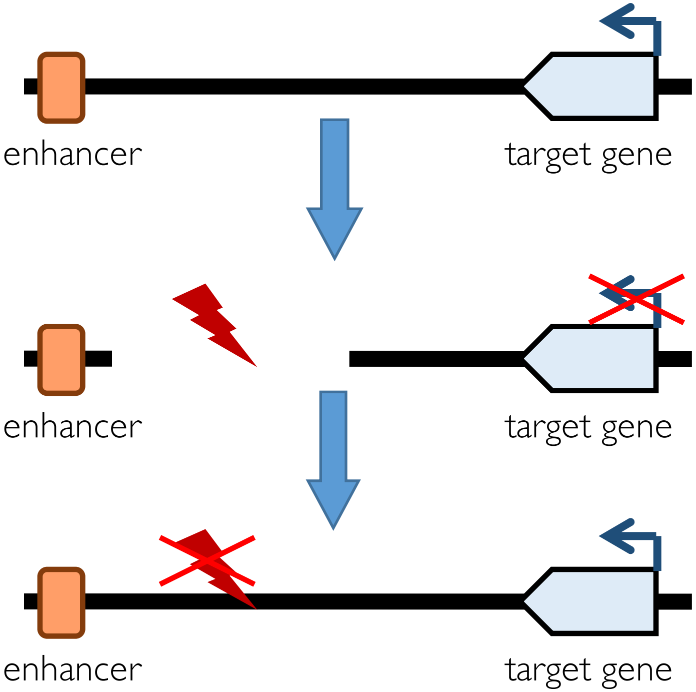
      (Mendenhall EM, et al. 2013 *Nat Biotechnol* **31**(12): 1133-1136.)
    - Mapping of coordinated chromatin state dynamics between enhancers and genes

      (Ernst J. et al. 2011 *Nature* **473**(7345): 43-49.)
    - Chromatin Capture Technologies (3C – 4C – 5C – HiC – Capture HiC)

      Splinter E.. 2012. *Methods* **58**(3): 221-230.

      Mifsud, B. et al. 2015. *Nature Genetics*, **47**(6), 598–606.

Predicting enhancer – target gene interactions by conservation of physical link



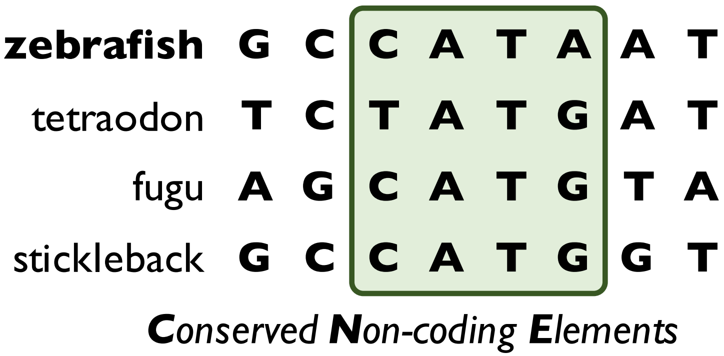
- Enhancer & target gene link is functional: mutations disrupting it will be eliminated by natural selection
- Computation of a co-segregation score for each potential target gene
- Most likely target gene = gene with most evolutionary conserved link with an enhancer

Long-range evolutionary constraints reveal *cis*-regulatory interactions on the human X chromosome

Magali Naville[1,2,3,*], Minaka Ishibashi[4,*], Marco Ferg[5,*], Hemant Bengani[6,*], Silke Rinkwitz[4], Monika Krecsmarik[7], Thomas A. Hawkins[8], Stephen W. Wilson[8], Elizabeth Manning[4], Chandra S.R. Chilamakuri[9], David I. Wilson[10], Alexandra Louis[1,2,3], F. Lucy Raymond[11], Sepand Rastegar[5], Uwe Strähle[5], Boris Lenhard[12], Laure Bally-Cuif[7], Veronica van Heyningen[6], David R. FitzPatrick[6], Thomas S. Becker[4,13,**] & Hugues Roest Crollius[1,2,3,**]

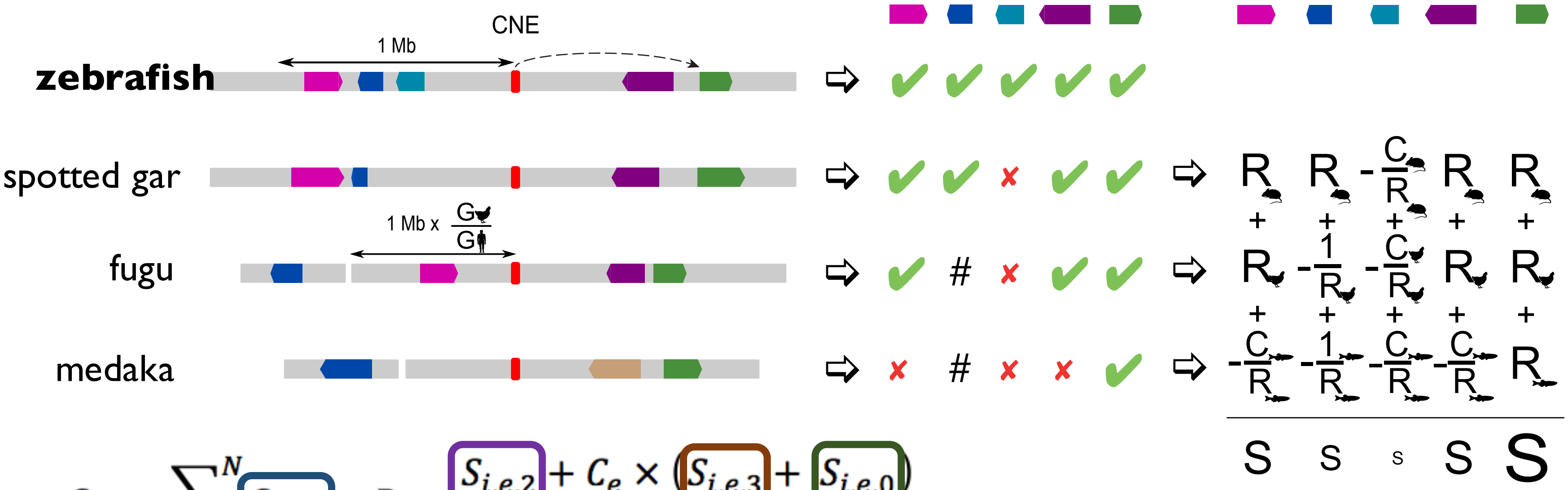
Finding the most likely target gene

1. Identification of putative enhancers: **C**onserved **N**on-coding **E**lements
2. Identification of the most likely target gene (co-segregation score)



zebrafish	G	C	C	A	T	A	A	T
tetraodon	T	C	T	A	T	G	A	T
fugu	A	G	C	A	T	G	T	A
stickleback	G	C	C	A	T	G	G	T

__C__onserved __N__on-coding __E__lements



$$S_{Ai} = \sum_1^N \boxed{S_{i,e,1}} \times R_e - \frac{\boxed{S_{i,e,2}} + C_e \times \left(\boxed{S_{i,e,3}} + \boxed{S_{i,e,0}}\right)}{R_e}$$

Co-segregation score for gene of interest

genome coverage

pair-wise rearrangement rate

sum over each species e

Gene in vicinity

Gene outside radius

Gene on ≠ chromosome

Gene absent

Goals of this study

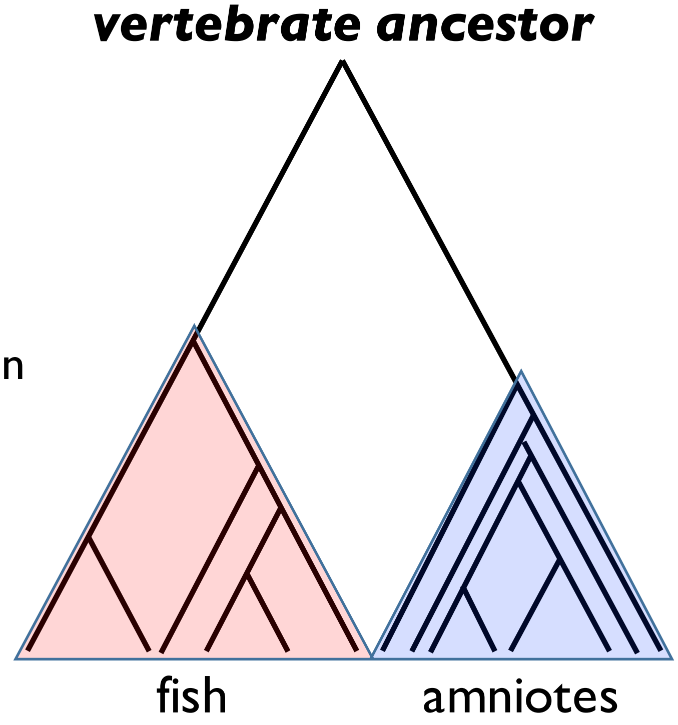
ARTICLE

Received 1 May 2014 | Accepted 12 Mar 2015 | Published 24 Apr 2015 DOI: 10.1038/ncomms7904 OPEN

Long-range evolutionary constraints reveal *cis*-regulatory interactions on the human X chromosome

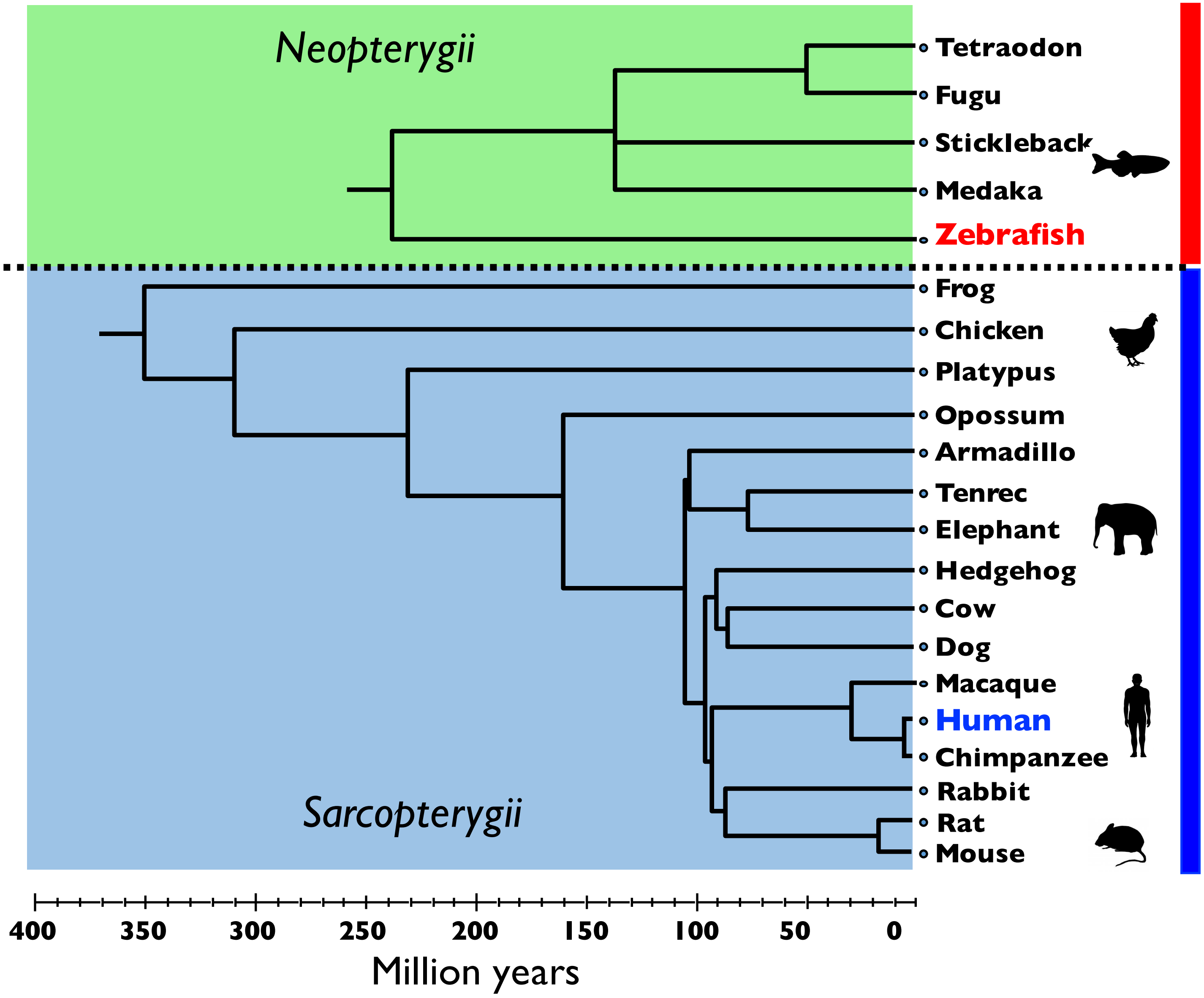
Magali Naville[1,2,3,*], Minaka Ishibashi[4,*], Marco Ferg[5,*], Hemant Bengani[6,*], Silke Rinkwitz[4], Monika Krecsmarik[7], Thomas A. Hawkins[8], Stephen W. Wilson[8], Elizabeth Manning[4], Chandra S.R. Chilamakuri[9], David I. Wilson[10], Alexandra Louis[1,2,3], F. Lucy Raymond[11], Sepand Rastegar[5], Uwe Strähle[5], Boris Lenhard[12], Laure Bally-Cuif[7], Veronica van Heyningen[6], David R. FitzPatrick[6], Thomas S. Becker[4,13,**] & Hugues Roest Crollius[1,2,3,**]

1. Identify functional enhancer – target gene interactions in zebrafish
2. Identify ancestral gene regulation circuits in vertebrates
3. Gain insights on enhancer evolution in vertebrates



Can we identify functional enhancer – target gene interactions?

Analysis of two independent sets of species for human and zebrafish



zebrafish

7 species
Home-made multiple alignments (multiz)

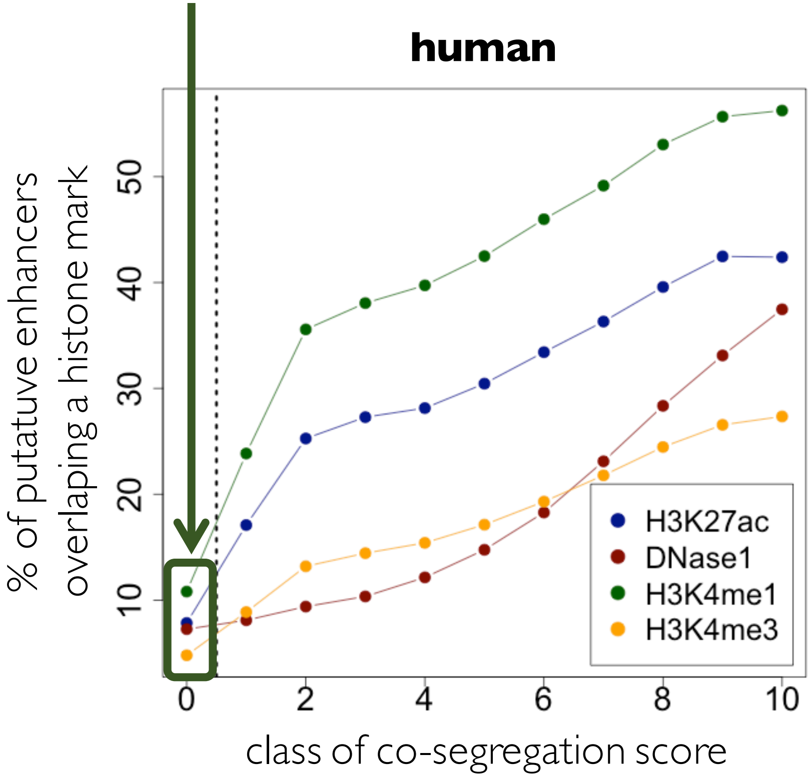
human

35 species
UCSC multiple alignments

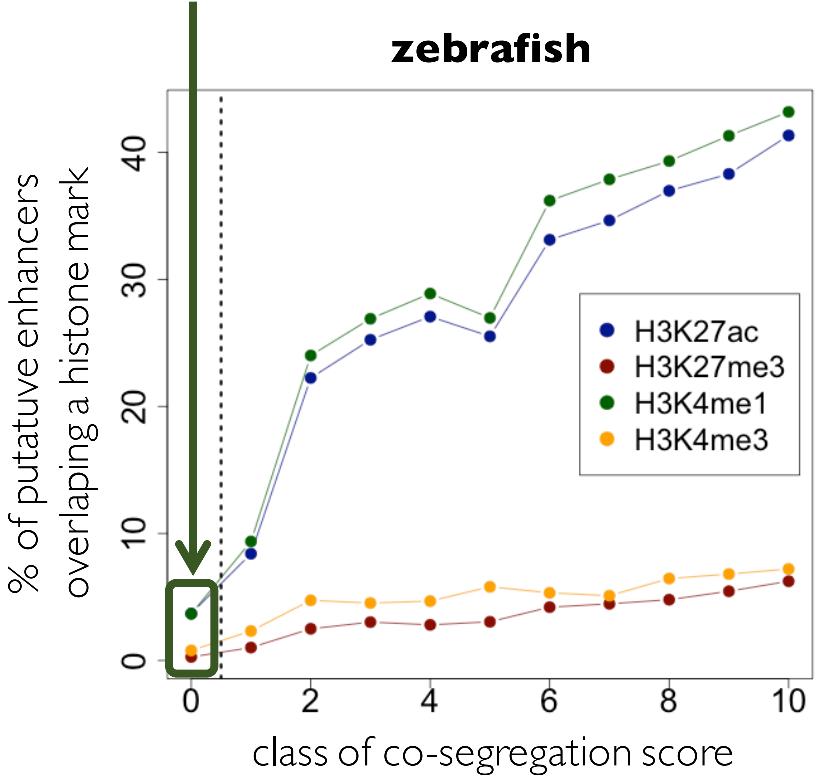
Blanchette et al 2004

Putative enhancers are marked with functional features

	# of enhancers with a predicted target gene	*# of target genes*
human	1,331,659	18,344
zebrafish	77,560	20,201



Putative enhancer with no target gene



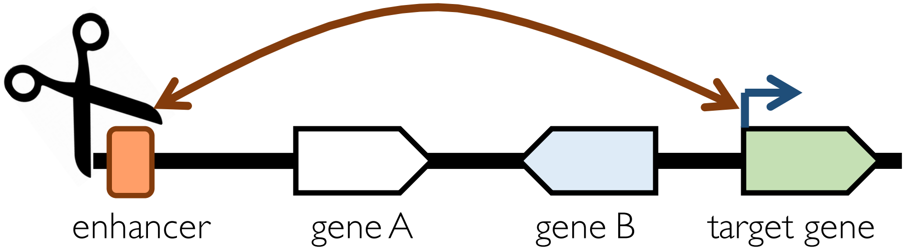
Putative enhancer with no target gene

Bogdanovic et al. 2012

ENCODE project

Testing functional interactions with CRISPR-Cas9

Collaboration with Patrick Torbey & Pascale Gilardi (Patrick Charnay team, IBENS)



IBENS
Institut de Biologie de l'Ecole Normale Supérieure

Testing enhancers in zebrafish
Enhancers with signals of:

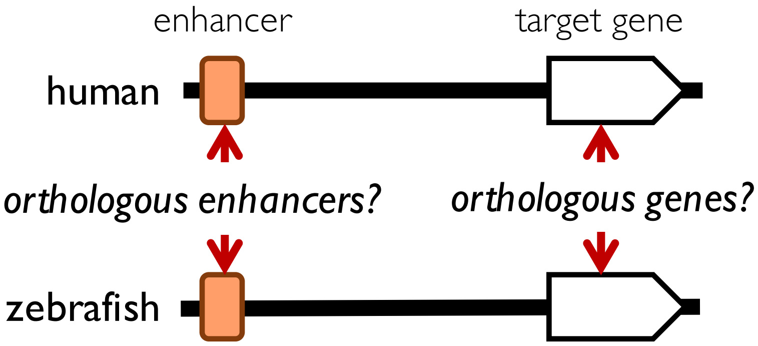
- Active chromatin (H3K4me1, H3K27ac)
- Sequence conservation (phyloP)
- Orthologous regulation with human

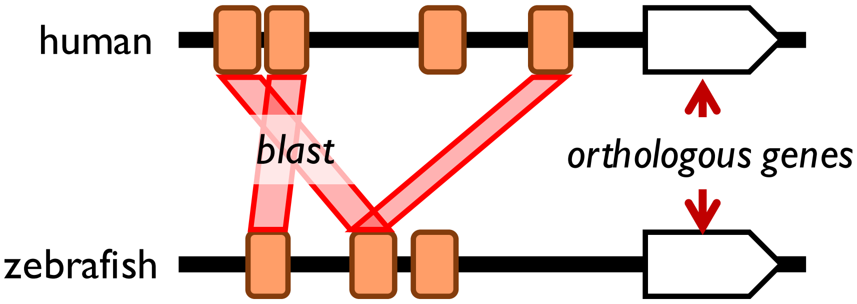
First-hand test of conservation of synteny-based method

gene	enhancer	pathology	rank
gria3b	chr14_1222	mental retardation	intronic
pcdh7b	chr7_8050		intronic
irx1b	chr19_2678	macular dystrophy	neighboring
hey1	chr19_3005		neighboring
cx35b/Gjb3	chr20_813	deafness	jumping
robo1/robo2	chr15_4050		

What are the conserved regulatory interactions in vertebrates?

Finding orthologous enhancers and target genes





	# of orthologous enhancers with orthologous target genes		# of orthologous target genes
human	397,524	150,011	11,053
zebrafish		53,100	13,299

What functions do these orthologous target genes of orthologous enhancers have?

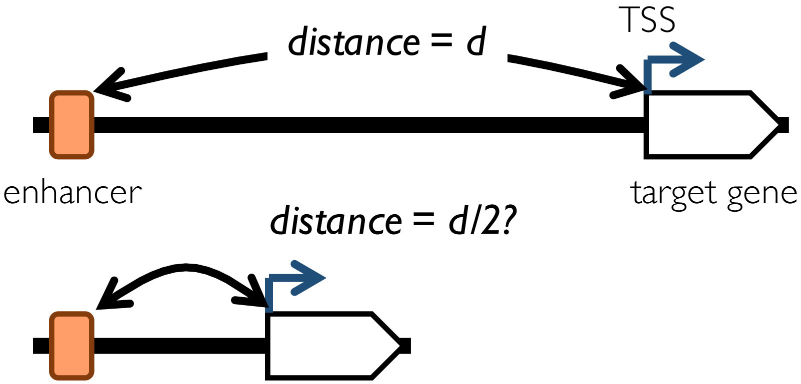
GO term enrichment analysis

human		**zebrafish**	
GO term	**rank**	**GO term**	**rank**
chondrocyte development	4	neuron recognition	1
dorsal/ventral axis specification	8	axon guidance	4
brain morphogenesis	9	neuron projection guidance	7
organ induction	12	retina development in camera-type eye	8
axonal fasciculation	18	embryonic viscerocranium morphogenesis	14
myoblast fusion	19	determination of bilateral symmetry	17
neuron recognition	21	specification of symmetry	18
dendrite morphogenesis	22	determination of left/right symmetry	19
neuromuscular junction development	24	neuron differentiation	27
cardiac chamber morphogenesis	25	tissue development	28

Orthologous genes with conserved regulatory regions are enriched for developmental functions

GOrilla: Eden et al. 2009

Enhancer – TSS interaction distances in human and zebrafish



human (genome size ≈ 3Gb)

zebrafish (genome size ≈ 1.5 GB)

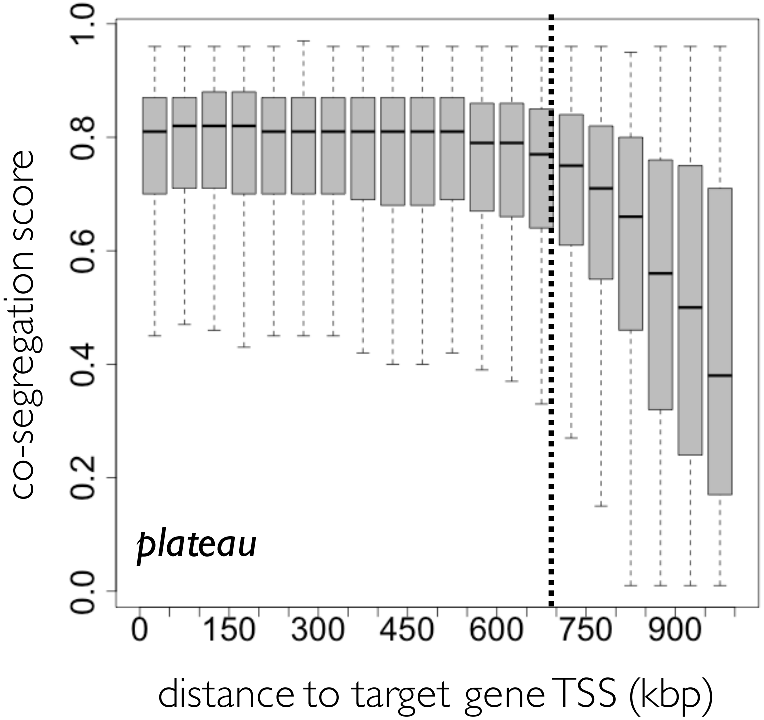
- Large scale characterization of enhancer – TSS interaction domains and distances in human (Capture Hi-C)
- Such data unavailable in zebrafish

- *Are enhancer – TSS interaction distances linked with genome size?*

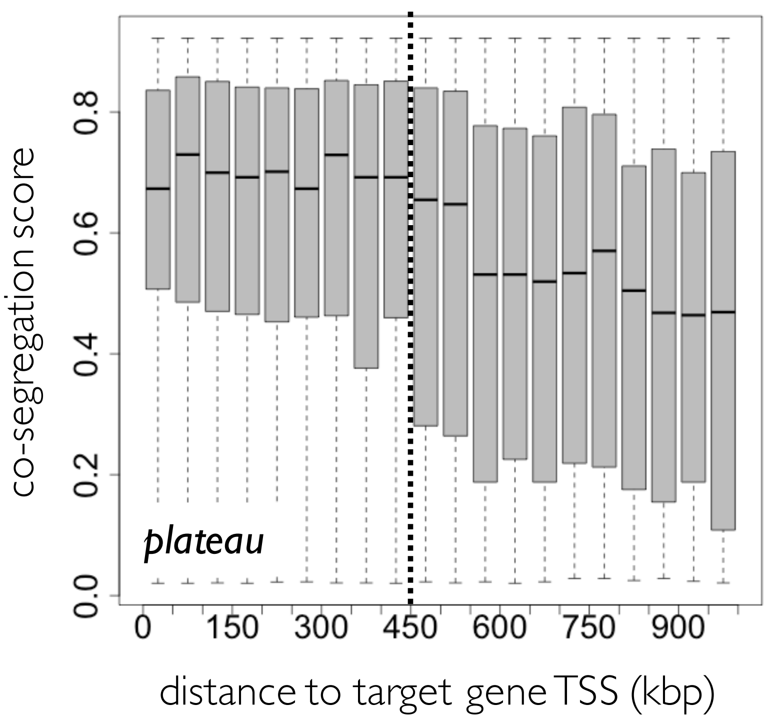
Mifsud et al. 2015

Linking co-segregation score and distance to target gene

human



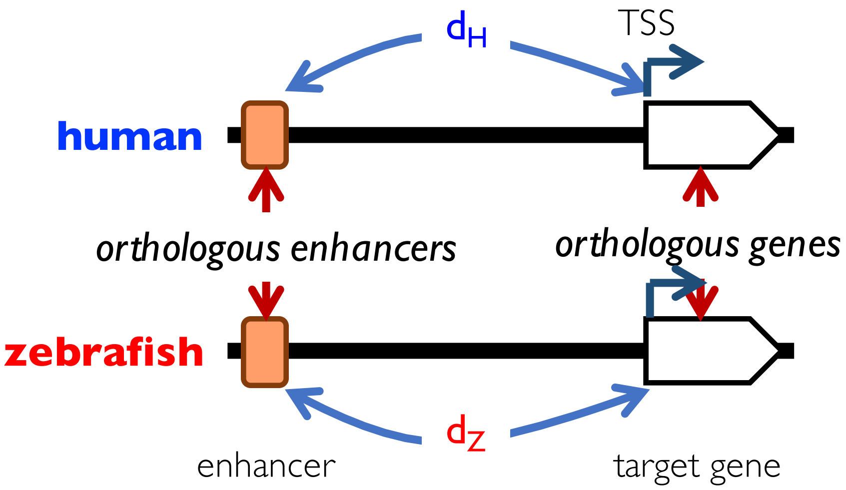
zebrafish



$\mathit{plateau}_{human} \sim 1.5 * \mathit{plateau}_{zebrafish}$

- Link between distance to target gene and co-segregation score
- Suggests a link between genome size and enhancer – TSS interaction distances

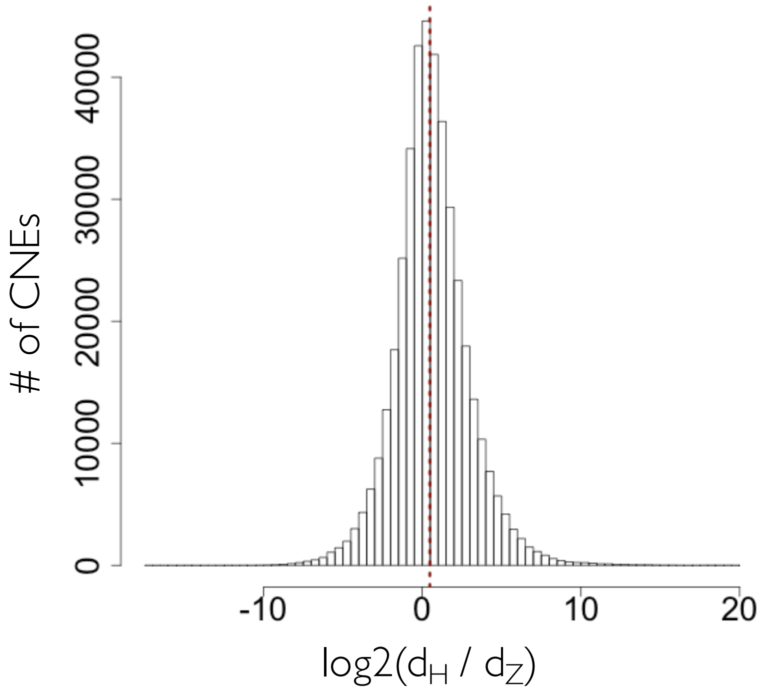
Link between genome size and enhancer – TSS interaction distance



$\log2(\mathbf{d_H} / \mathbf{d_Z})$

median $\log2(d_H / d_Z) \approx 0.48$

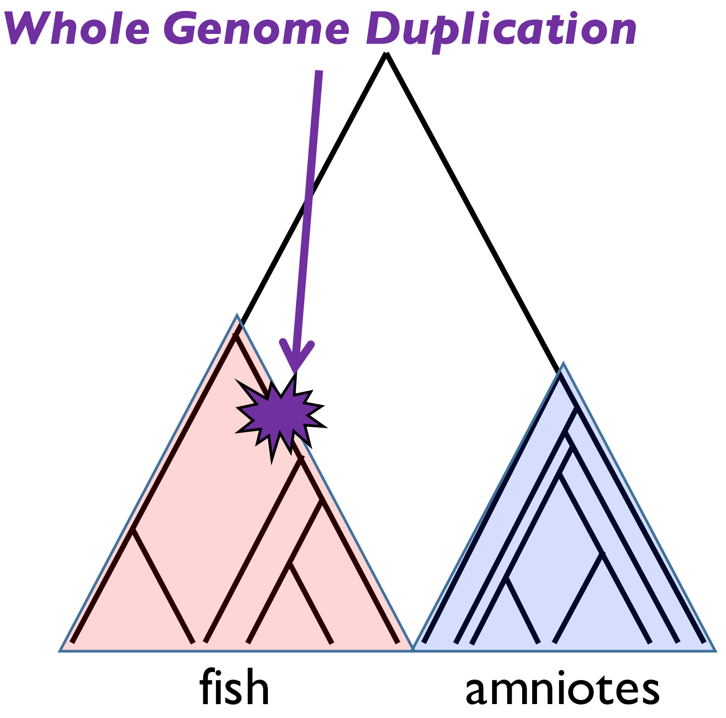
median $d_H / d_Z \approx 1{,}4$

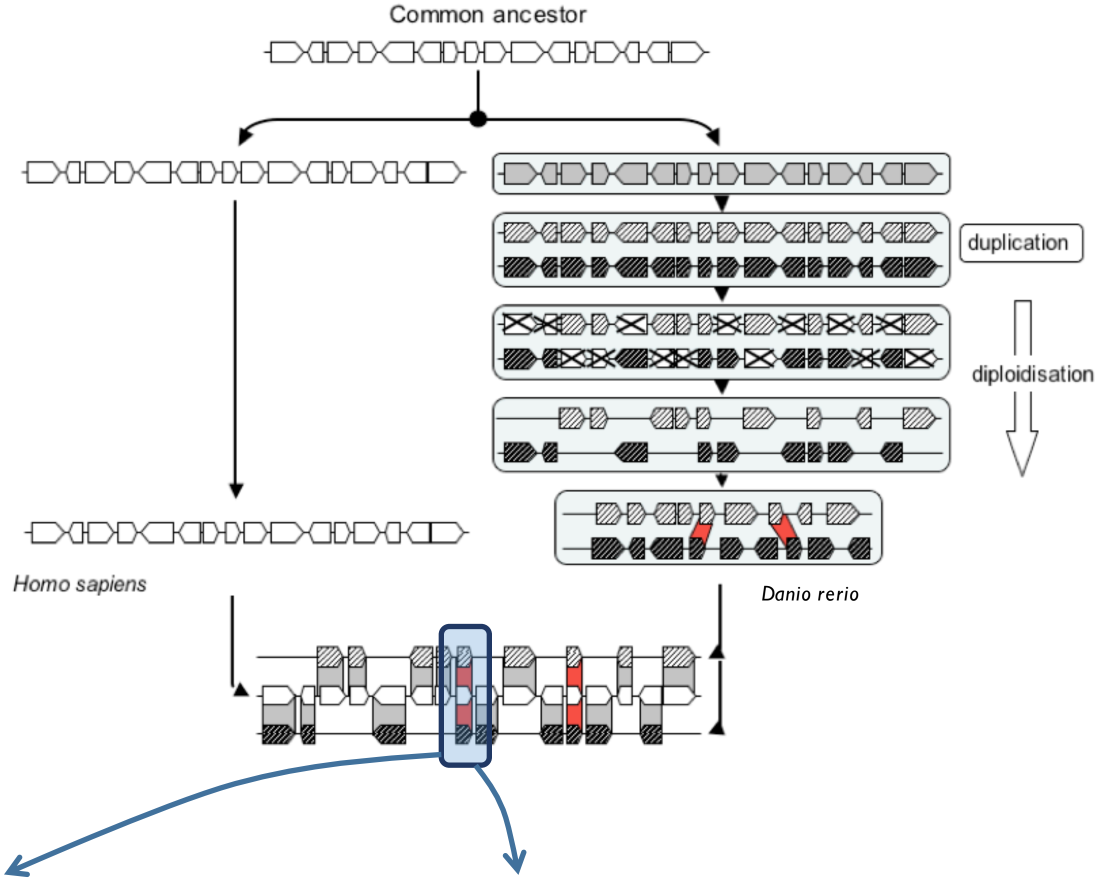


- Enhancer – TSS interactios distances scale with genome size

Fate of regulatory regions following a whole genome duplication

Focusing on the whole genome duplication in fish

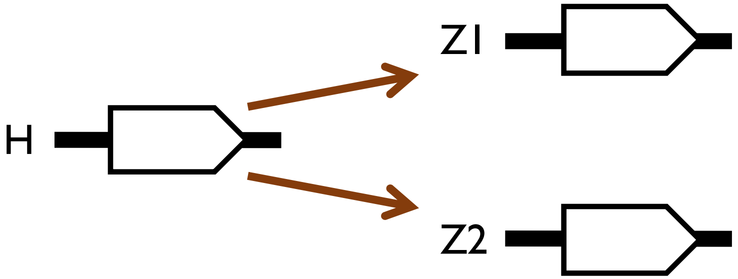




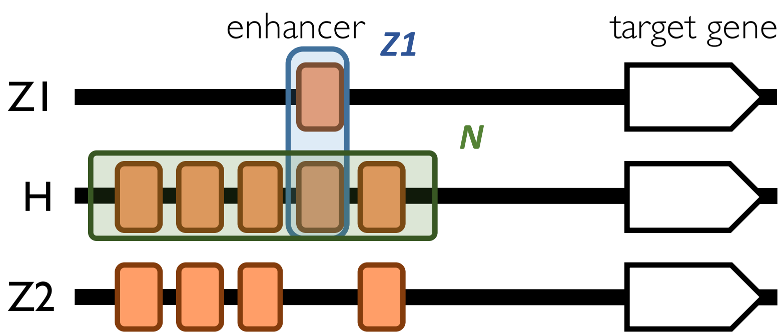
Fate of regulatory regions?
- Conservation?
- Sub-functionalization?
- Neo-functionalization?

Link with gene expression?

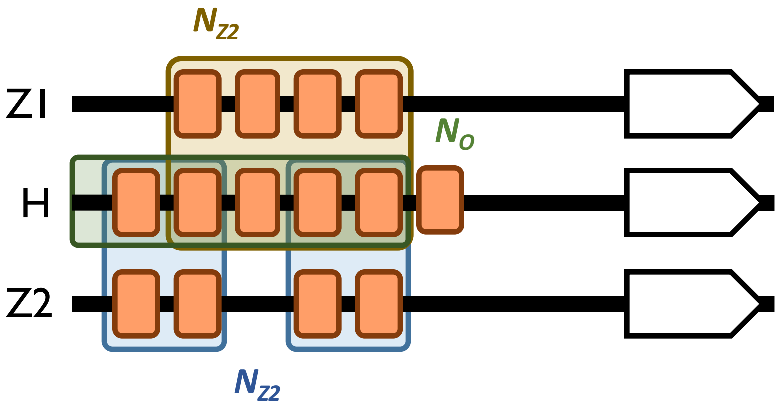
Possible fate for enhancers following WGD



Ohnolog "Trios": two zebrafish ohnologs (courtesy of Camille Berthelot) and their human ortholog



Biased retention of enhancers on one copy?



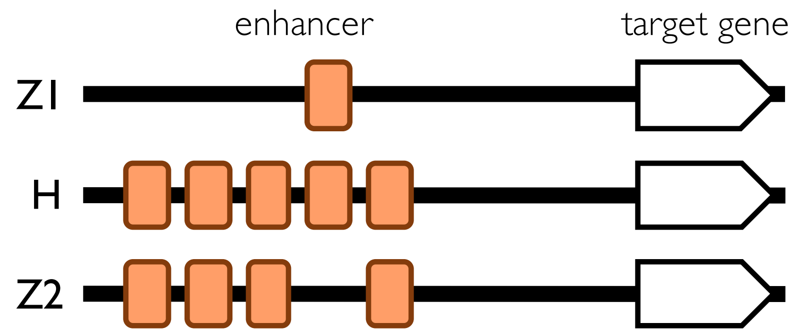
Conservation of enhancers on both copies?

probability of observed numbers caused by chance?

$$P(N_{Z1}|N) = \sum_{i=0}^{N_{Z1}} \frac{\binom{N}{i}}{2^N}$$

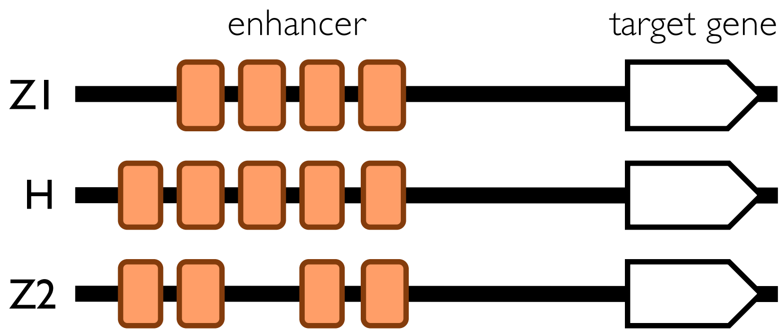
$$P(N_O|N_{Z1},N_{Z2},N_T) = \frac{\binom{N_{Z2}}{N_O} * \binom{N_T-N_{Z2}}{N_{Z1}-N_O}}{\binom{N_T}{N_{Z1}}}$$

Possible fate for enhancers following WGD



Biased retention of enhancers on one copy?

median probability	*# of cases w/ $p < 0.05$*
0.03125	971 / 1794



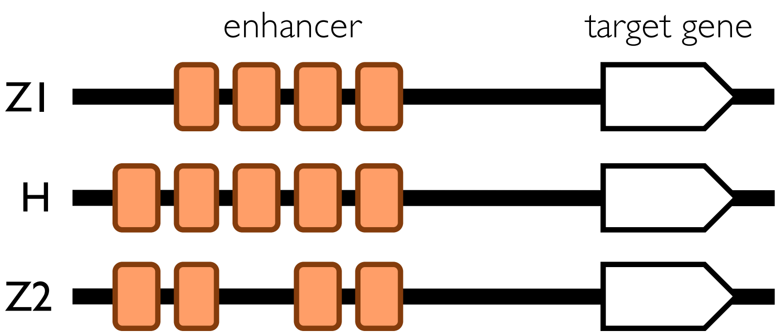
Conservation of enhancers on both copies?

median probability	*# of cases w/ $p < 0.05$*
0.275	355 / 1215

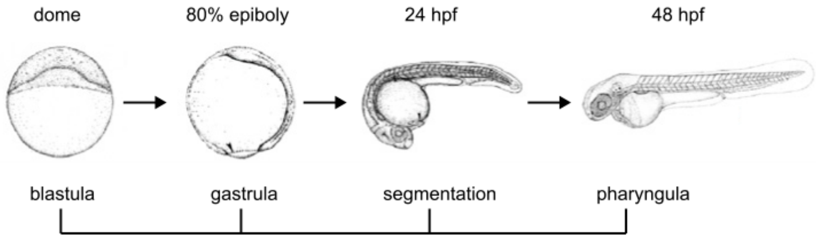
		biased retention	
		$p \geq 0.05$	$p < 0.05$
both copies	$p \geq 0.05$	434	425
	$p < 0.05$	86	267

Retention of enhancers following WGD is mostly random or biased on one copy.

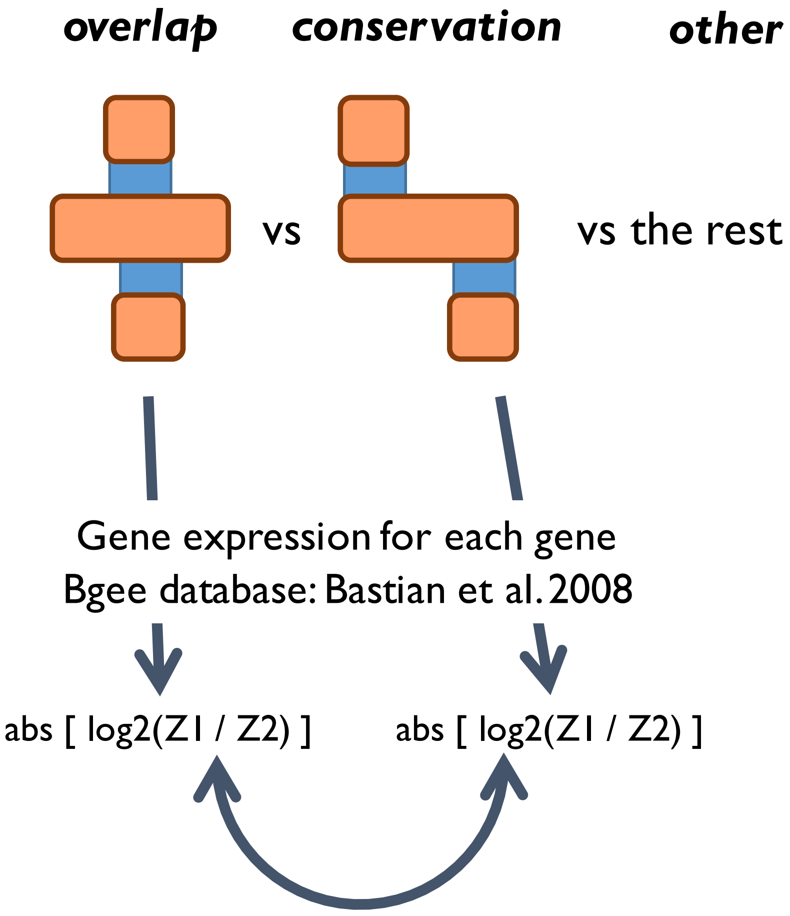
Link between regulation conservation and expression conservation



Conservation of enhancers on both copies



Independent analysis for different developmental stages



overlap ***conservation*** ***other***

vs

vs the rest

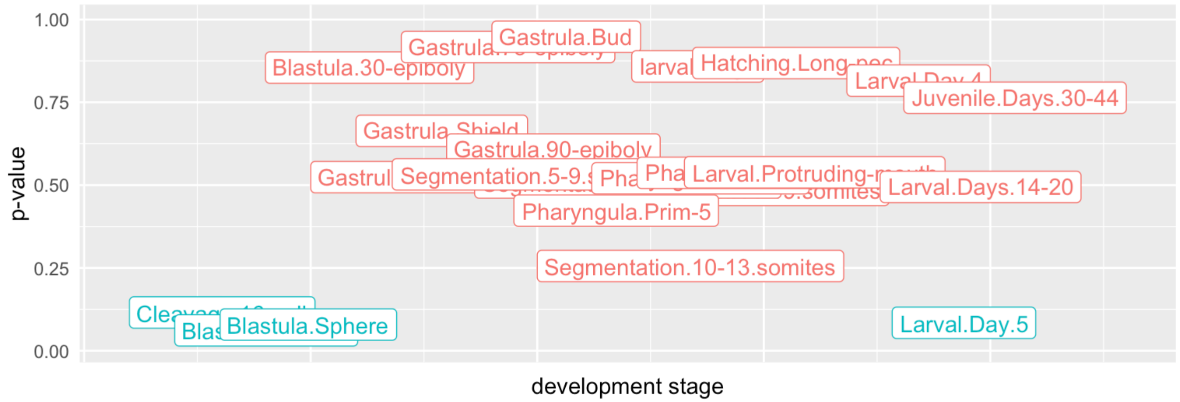
Gene expression for each gene
Bgee database: Bastian et al. 2008

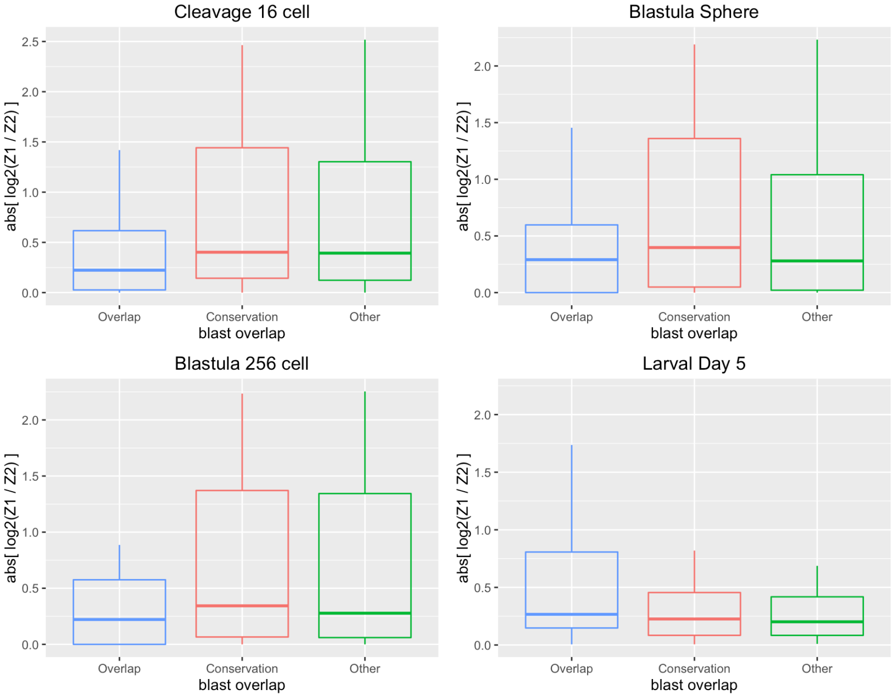
abs [log2(Z1 / Z2)]

abs [log2(Z1 / Z2)]

Compare distributions (T-test)

Early developmental stages show conservation of expression





- Ohnologous genes with overlapping enhancers show signals of conserved expression
- Conserved transcription factor binding sites?
- This feature is true for early developmental stages in zebrafish.

Conclusions

- Large-scale predictions of long range cis-regulatory interaction in two vertebrate genomes
- Predicted elements overlap functional marks

- Conserved regulatory interaction in vertebrate are linked with developmental functions.
- Enhancer – TSS interaction distances scale with genome size

- Random or biased conservation towards one ohnolog of enhancers following WGD

- Conservation of expression for ohnologs with conserved regulatory regions for genes expressed in early developmental stages

Hugues Roest Crollius
Magali Naville
Camille Berthelot
Alexandra Louis
Joseph Lucas
Lambert Moyon
Amélie Peres
Christine Sacerdot

IBENS
Institut de Biologie de l'Ecole Normale Supérieure

Dyogen Group

cnrs
dépasser les frontières
www.cnrs.fr