Data set integration for defining comprehensive models of gene regulation

Joel Smith

MBL, Woods Hole, MA

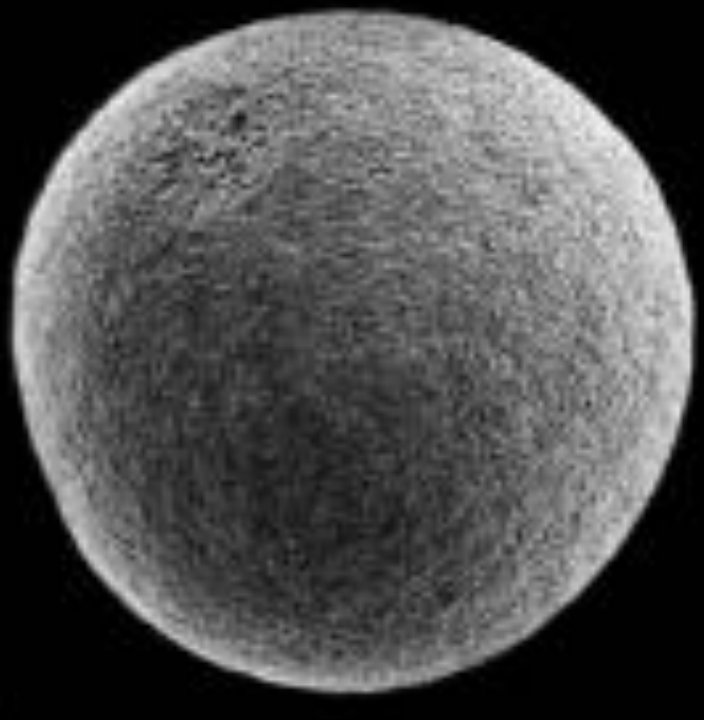
- GRN methodology for developmental systems

- [Optional] The results, examples from the sea urchin

- GRN methodology in cancer systems

20,000-30,000
2,000-3,000
200-300
20-30
2-3

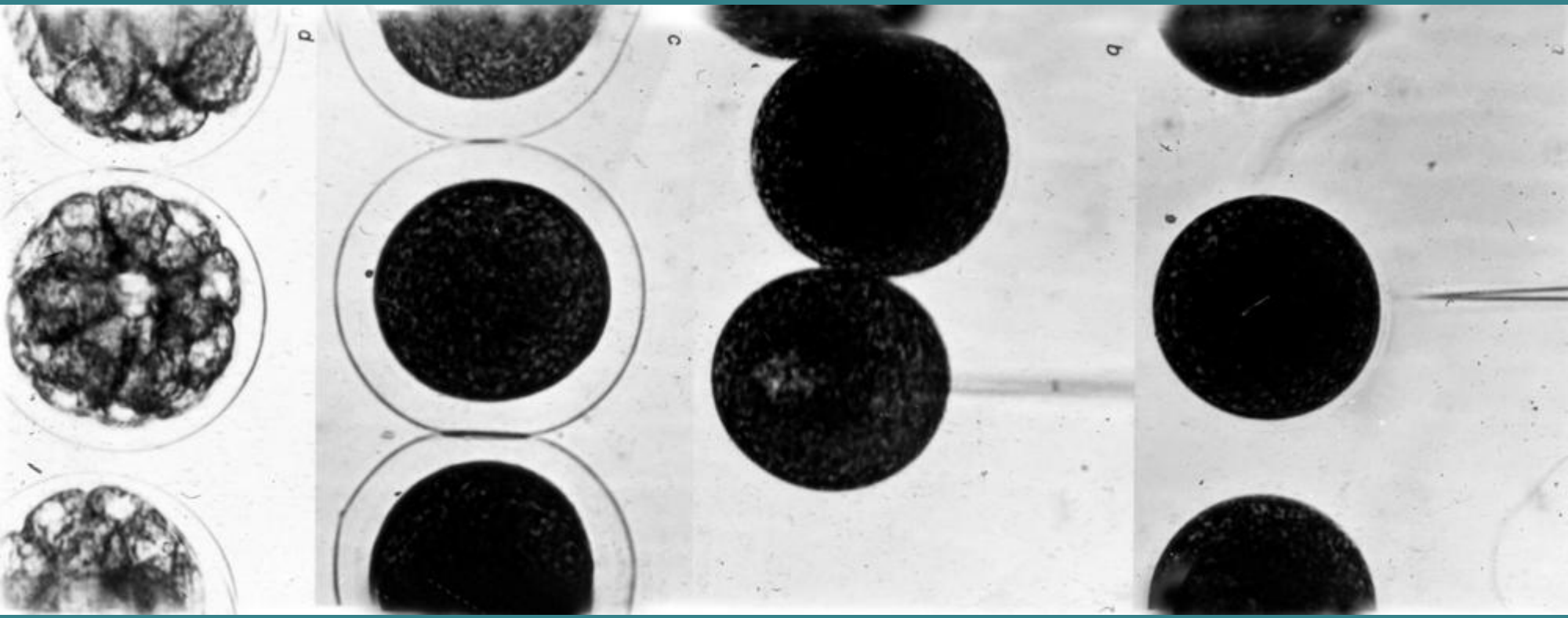
0.001
0.01
100
1,000-10,000



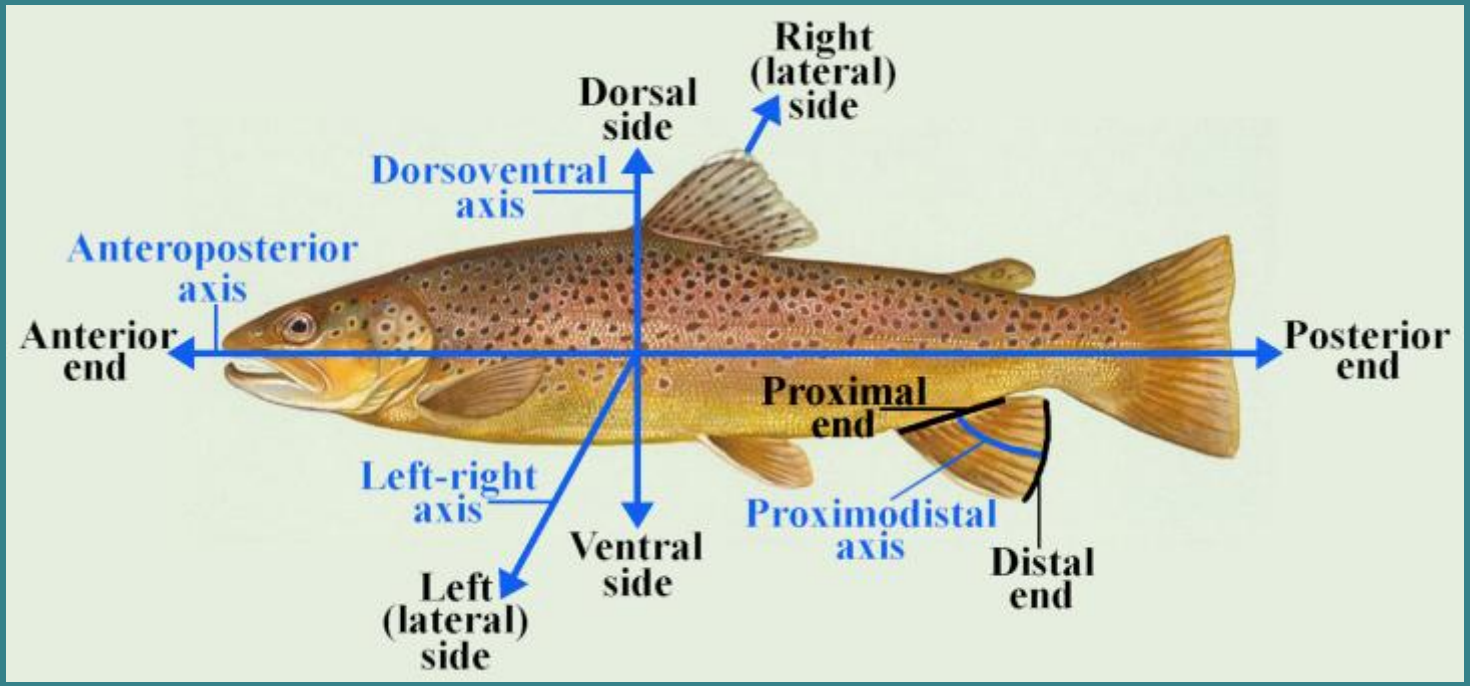
The starlet sea anemone

"Stella"

-Insights into basic animal development
-Tractable

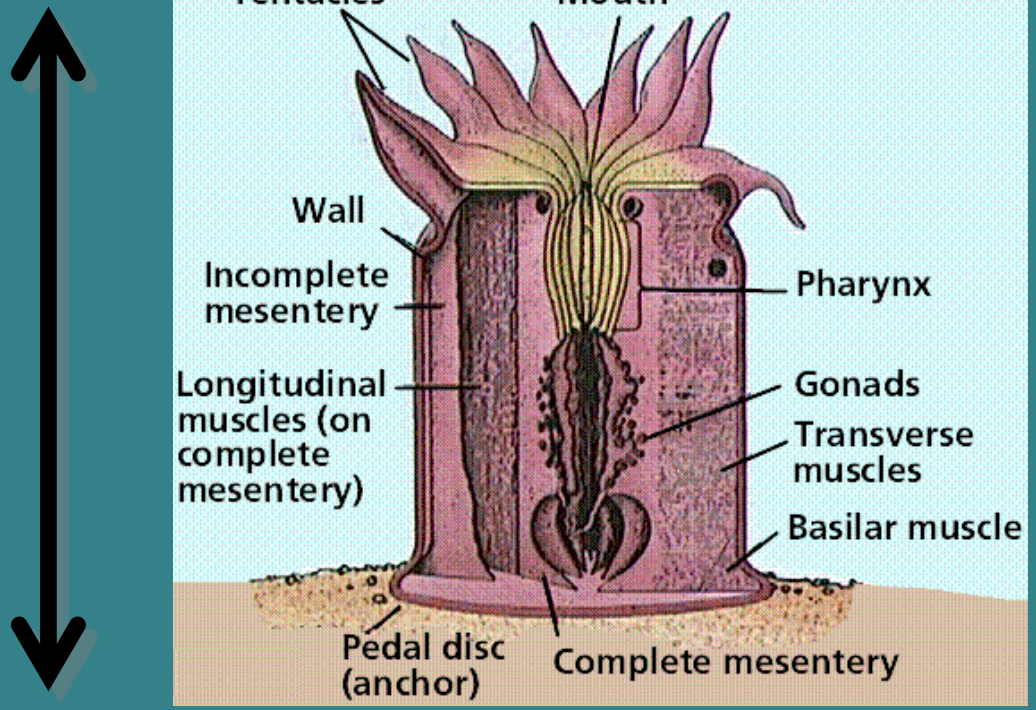


Orthogonal body axes and germ layers underly the bilaterian body plan



Coralscience.org

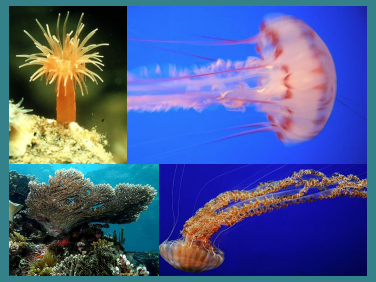
"Radial" and "Simple" Animals



The cnidarian body plan bears the "bilaterian"molecular signature

Wnt and TGFβ pattern 1° / 2° axes

Cohort of genes driving mesoderm specification and EMT



Wnt's and TGFβ's expressed along 1° / 2° axes

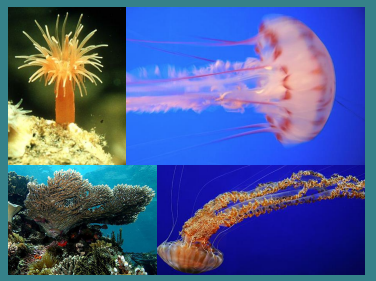
Canonical "mesoderm" genes expressed

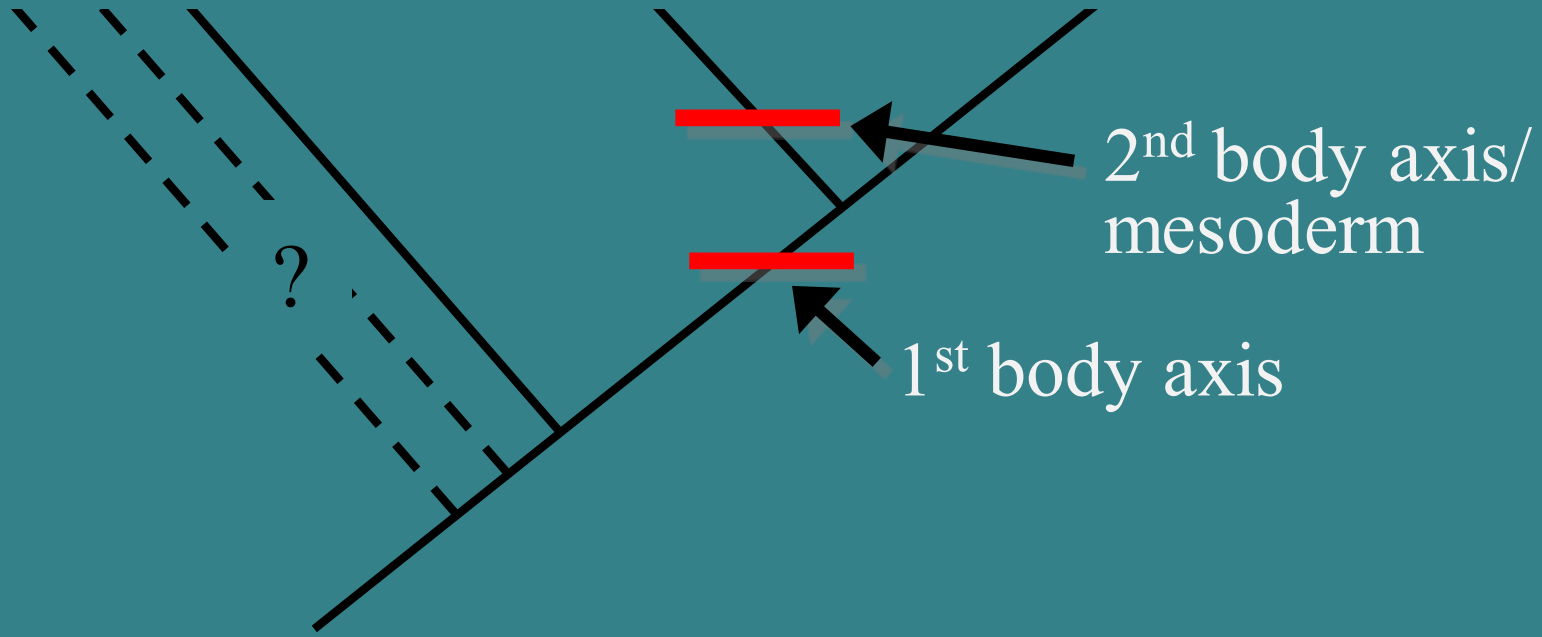
Cnidarians and Bilaterians are sister groups

Porifera

Bilateria

Cnidaria



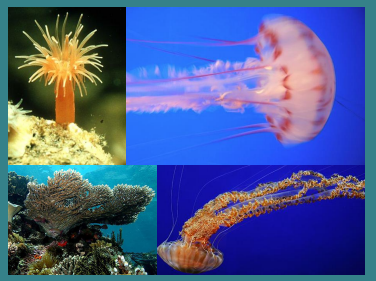


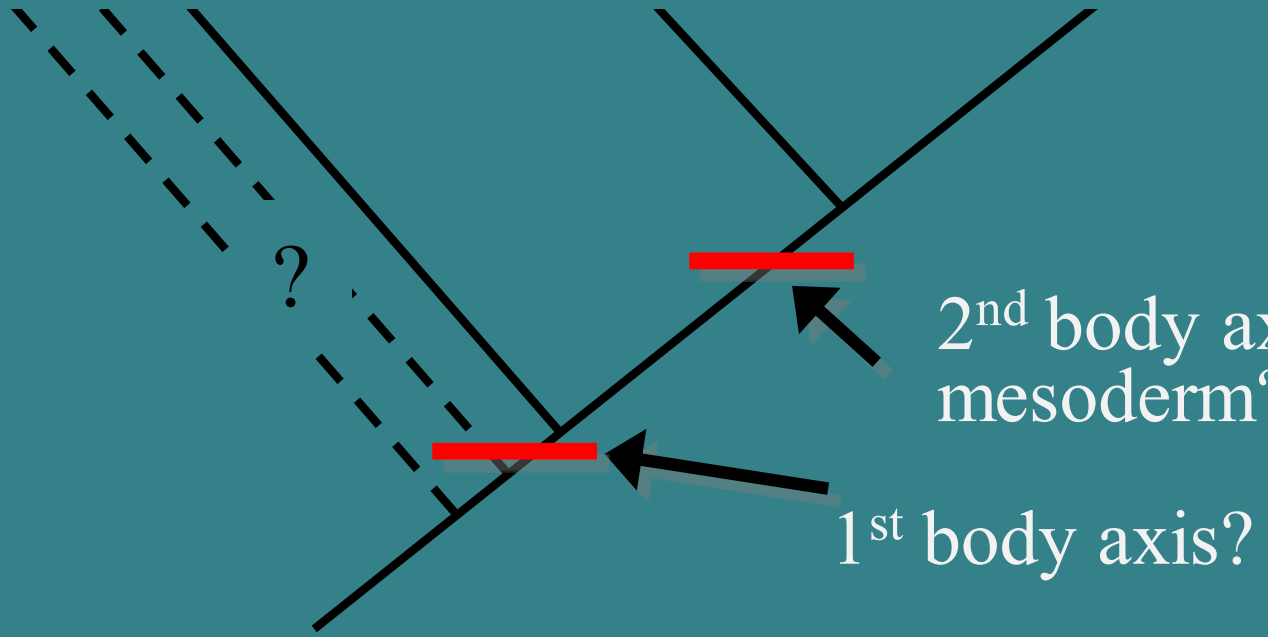
Cnidarians and Bilaterians are sister groups

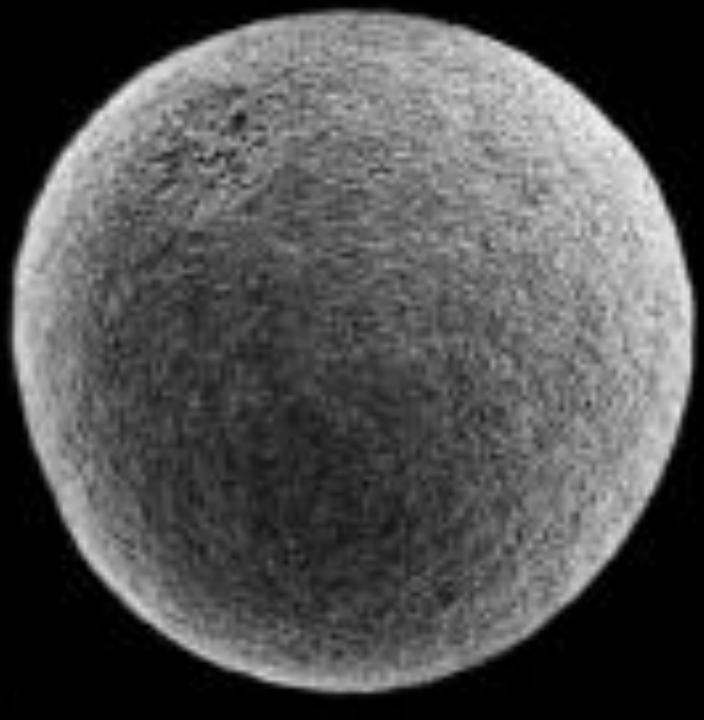
Porifera

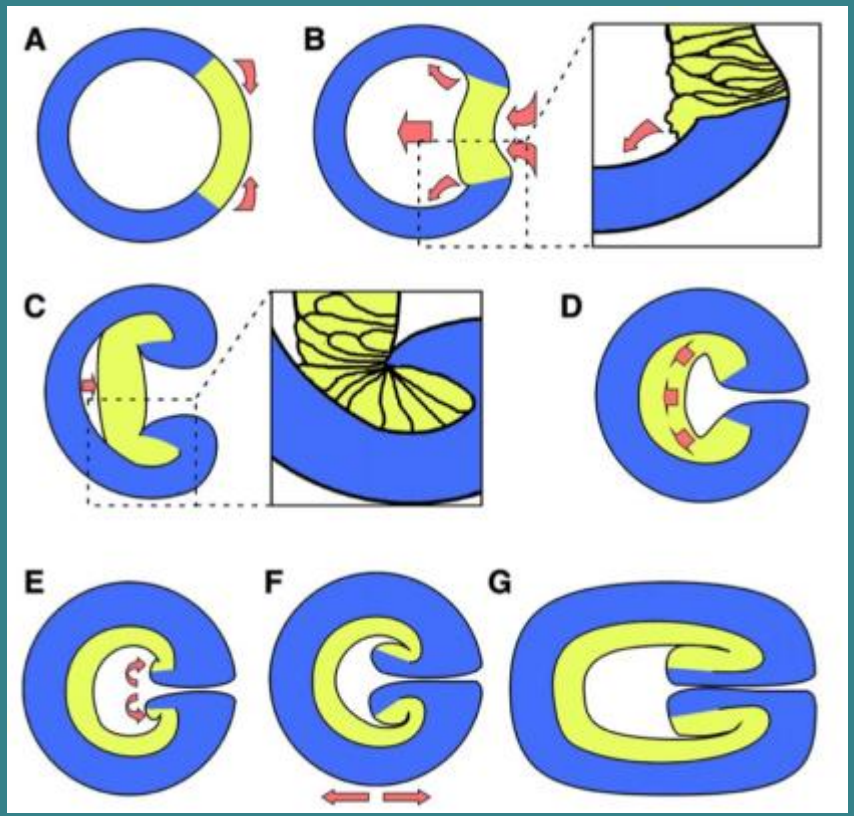
Bilateria

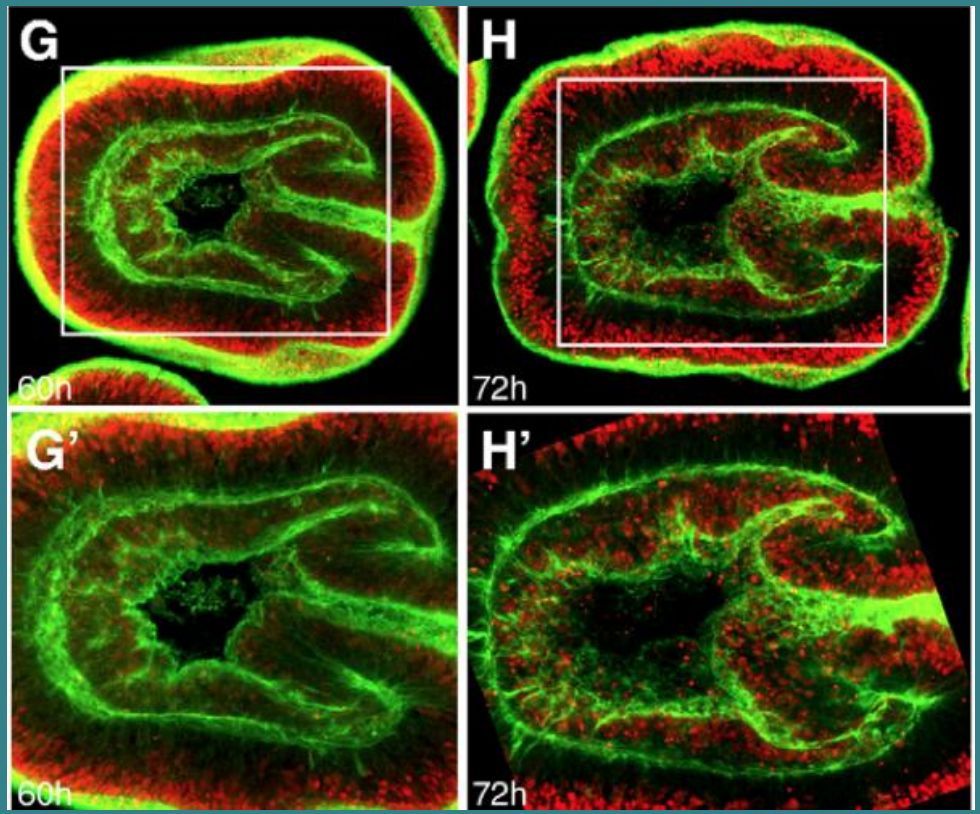
Cnidaria



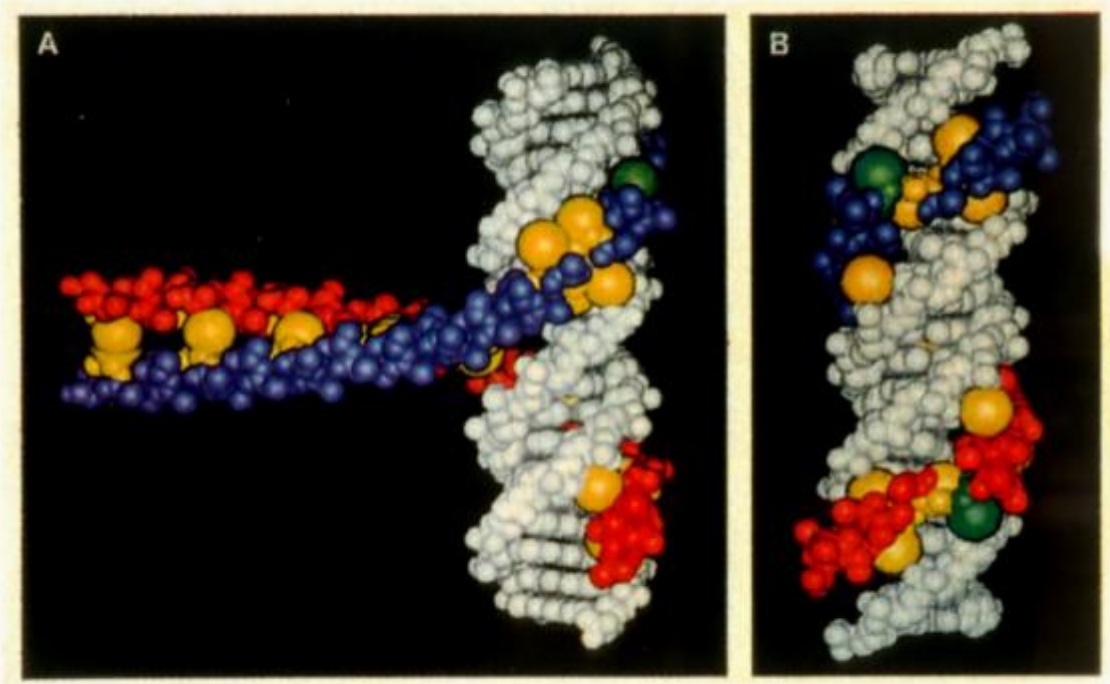


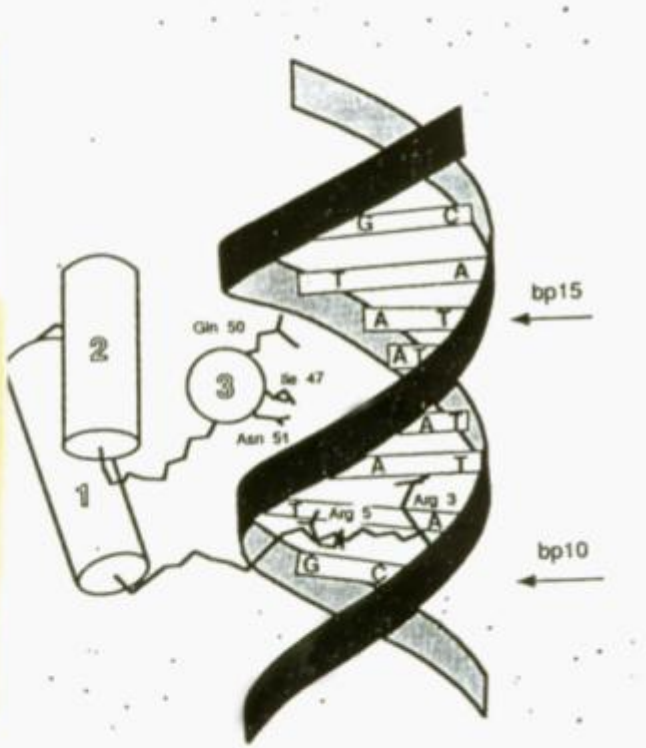






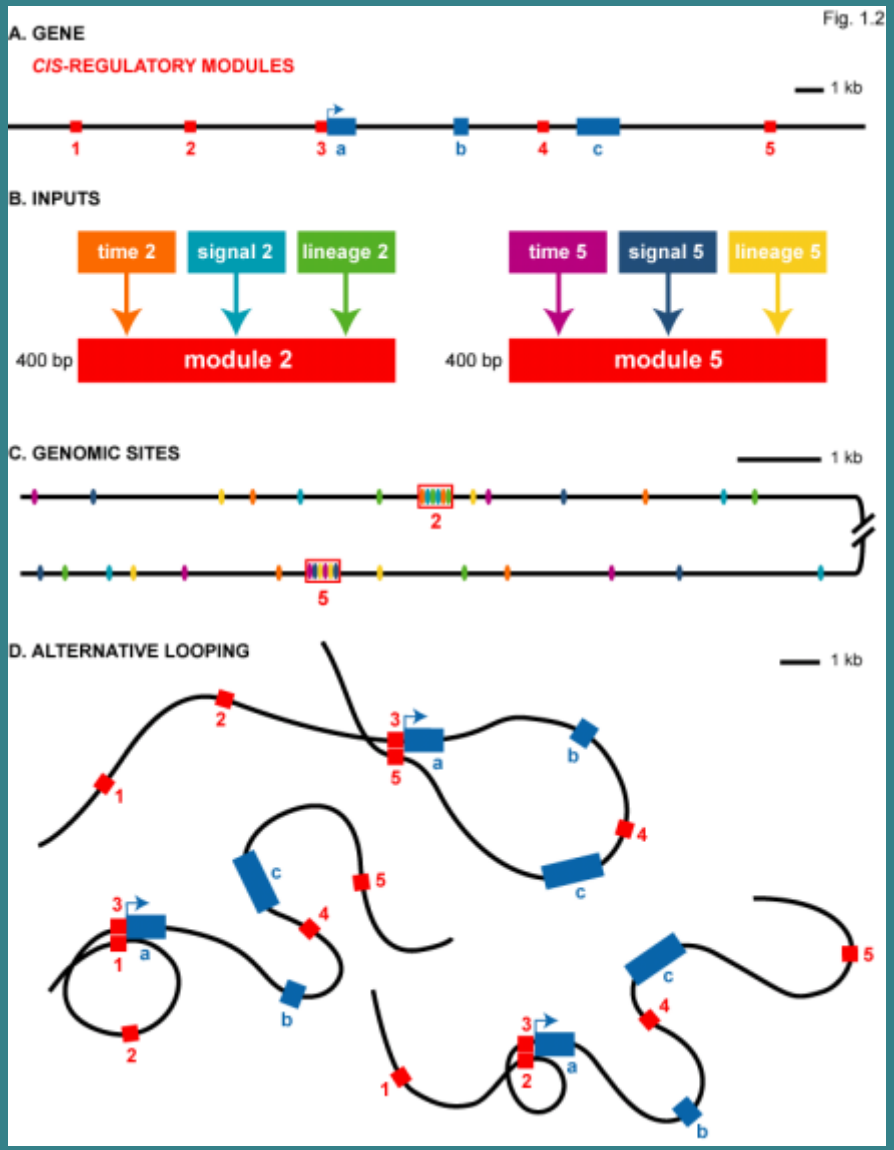
Magie et al., 2007

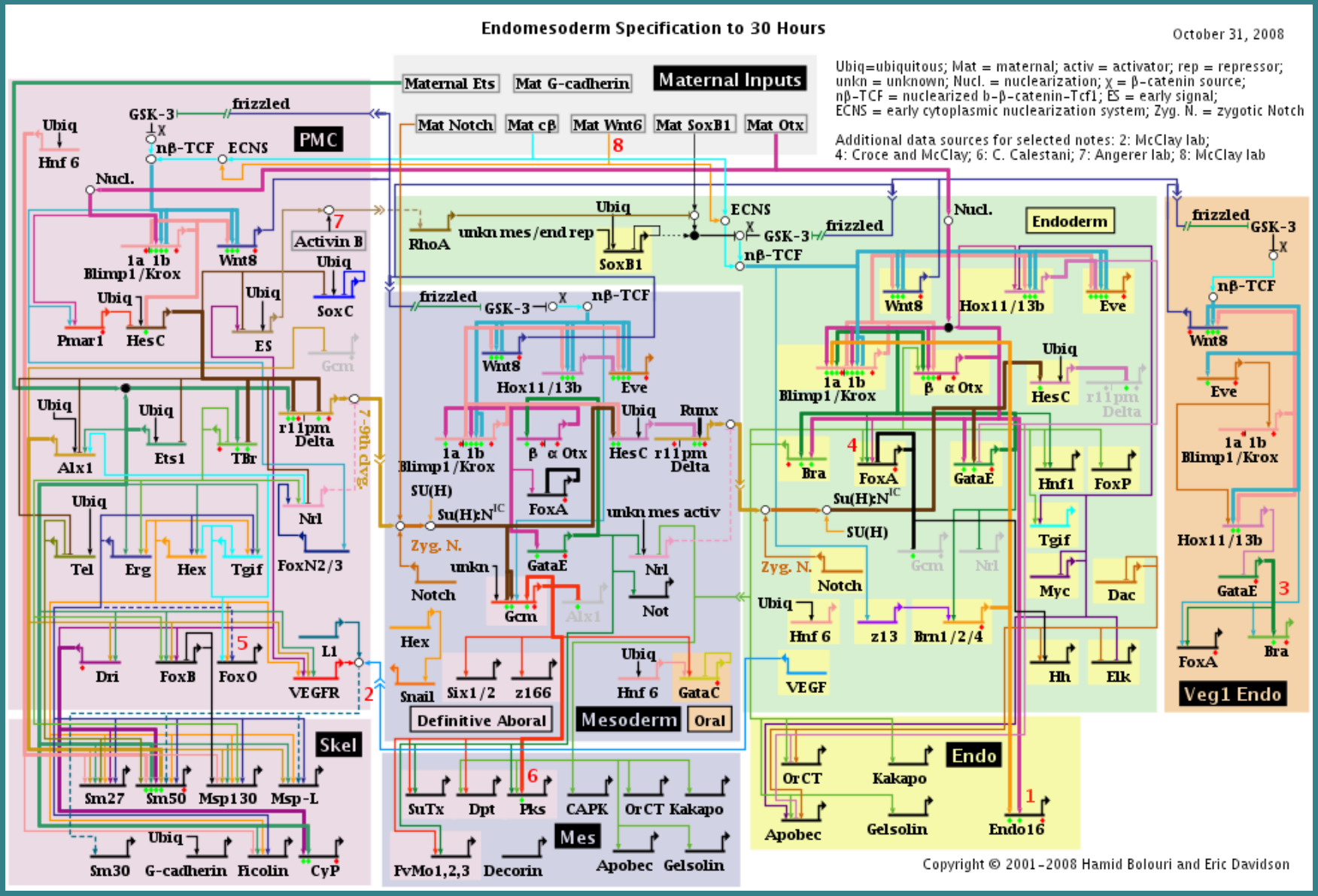


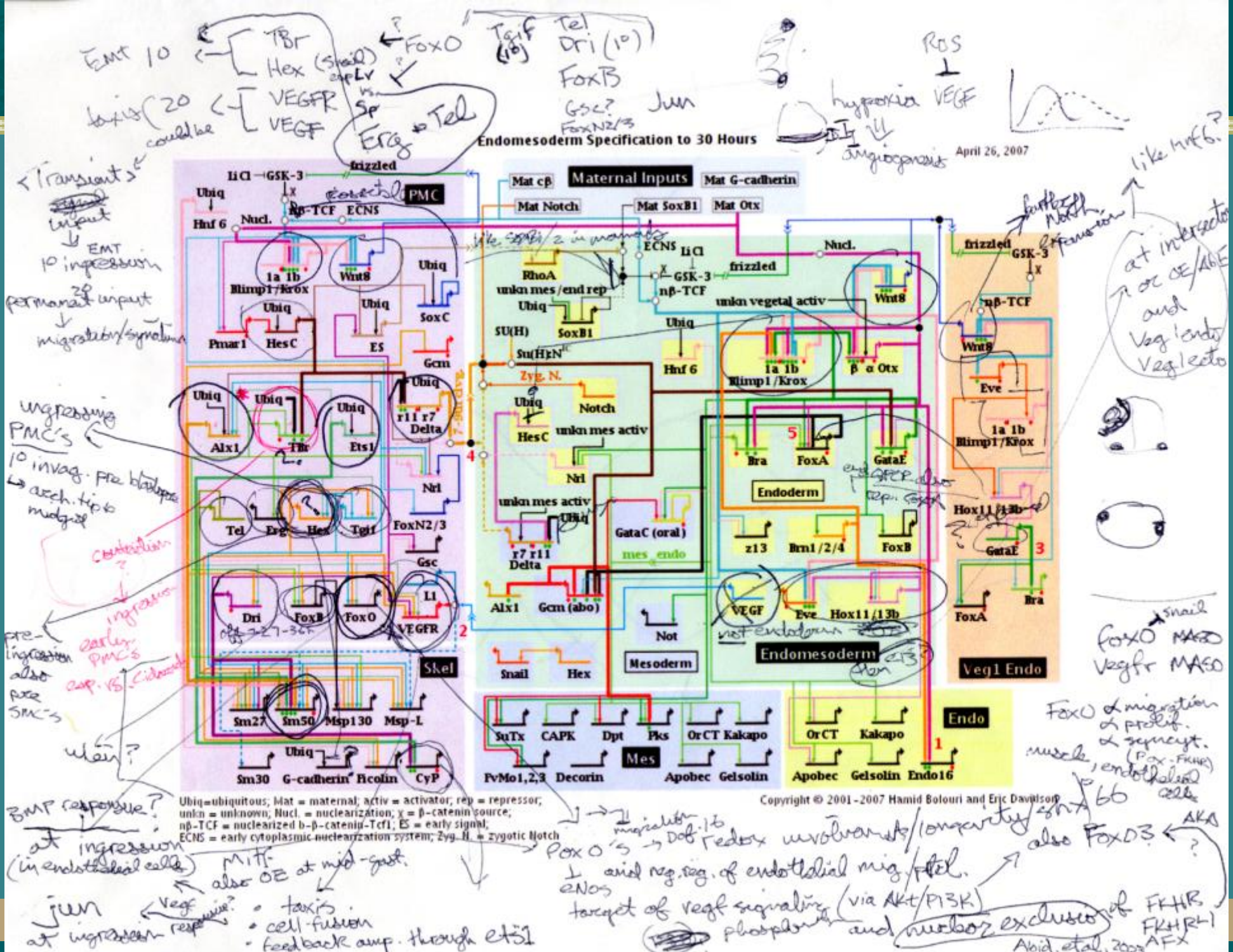


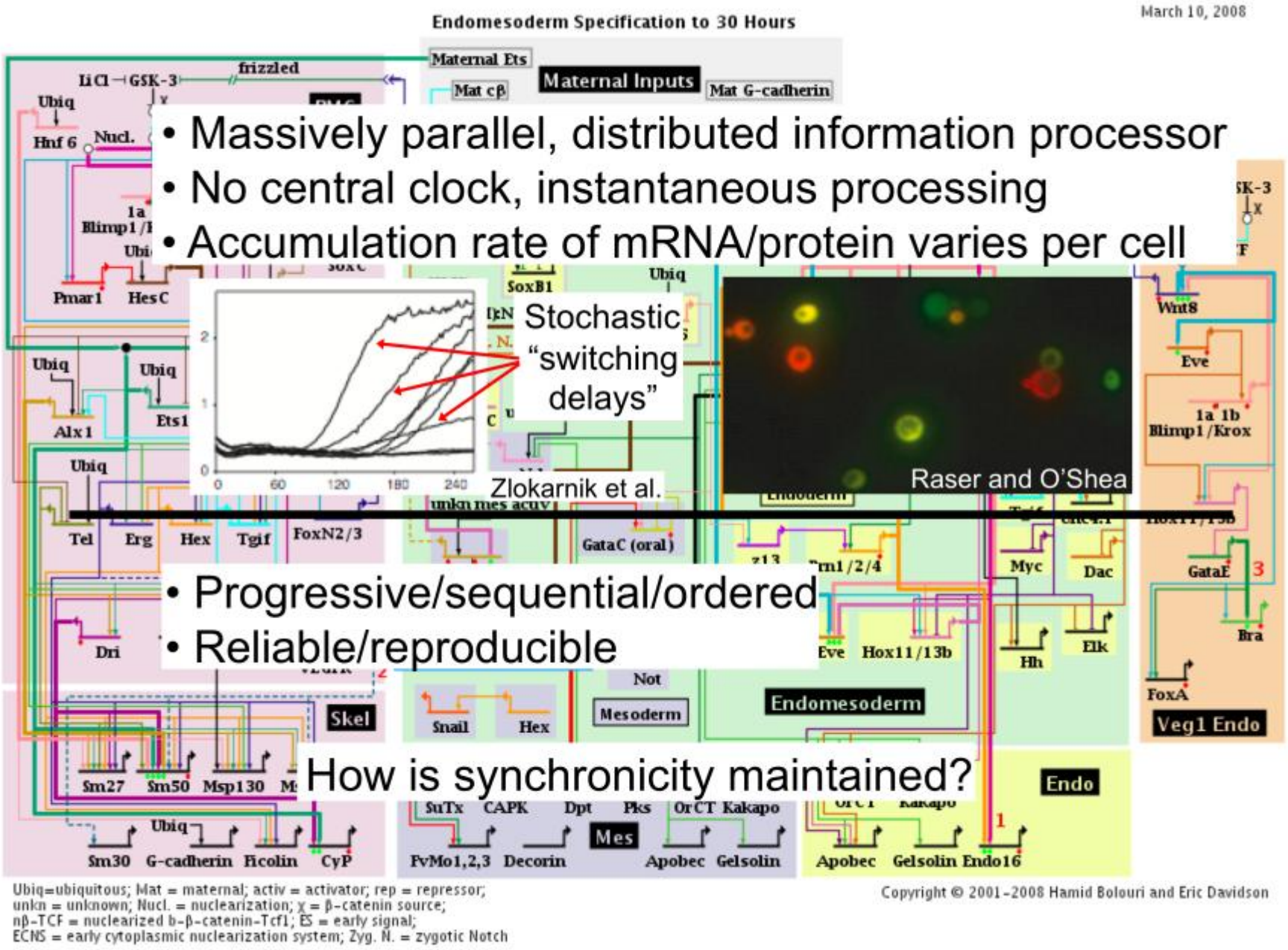
Combinations of different transcription factors

multiple “inputs” +
multiple “outputs”
= network






Endomesoderm Specification to 30 Hours
April 26, 2007
Maternal Inputs
Mat cβ
Mat G-cadherin
Mat Notch
Mat SoxB1
Mat Otx
PMC
Skel
Mesoderm
Endoderm
Endomesoderm
Endo
Mes
Veg1 Endo
Ubiq
Nucl.
nβ-TCF
ECNS
LiCl
GSK-3
frizzled
Hnf 6
1a 1b
Blimp1/Krox
Wnt8
SoxC
Pmar1
HesC
ES
Gcm
r11 r7
Delta
Alx1
Tbr
Ets1
Nrl
Tel
Erg
Hex
Tgif
FoxN2/3
Gsc
Dri
FoxB
FoxO
Li
VEGFR
Sm27
Sm50
Msp130
Msp-L
Sm30
G-cadherin
Ficolin
CyP
RhoA
unkn mes/end rep
SoxB1
SU(H)
Su(H)N
Zyg. N.
Notch
unkn mes activ
r7 r11
Delta
GataC (oral)
mes endo
Alx1
Gcm (abo)
Snail
Hex
Not
unkn vegetal activ
β α Otx
Bra
FoxA
GataE
z13
Brn1/2/4
FoxB
VEGF
Eve
Hox11/13b
SuTx
CAPK
Dpt
Pks
OrCT
Kakapo
PvMo1,2,3
Decorin
Apobec
Gelsolin
Endo16
Ubiq=ubiquitous; Mat = maternal; activ = activator; rep = repressor;
unkn = unknown; Nucl. = nuclearization; χ = β-catenin source;
nβ-TCF = nuclearized b-β-catenin-Tcf1; ES = early signal;
ECNS = early cytoplasmic nuclearization system; Zyg. N. = zygotic Notch
Copyright © 2001-2007 Hamid Bolouri and Eric Davidson


Endomesoderm Specification to 30 Hours
March 10, 2008
Maternal Inputs
Maternal Ets
Mat cβ
Mat G-cadherin
frizzled
LiCl
GSK-3
Ubiq
Hnf 6
Nucl.
Pmar1
HesC
SoxB1
Wnt8
Eve
1a 1b
Blimp1/Krox
Alx1
Ets1
Tel
Erg
Hex
Tgif
FoxN2/3
GataC (oral)
Myc
Dac
GataE
Bra
Elk
Hh
FoxA
Veg1 Endo
Dri
Skel
Sm27
Sm50
Msp130
Sm30
G-cadherin
Ficolin
CyP
Not
Mesoderm
Endomesoderm
Snail
Hex
Eve
Hox11/13b
Endo
SuTx
CAPK
Dpt
Pks
OrCT
Kakapo
Mes
FvMo1,2,3
Decorin
Apobec
Gelsolin
Endo16
Stochastic "switching delays"
Zlokarnik et al.
Raser and O'Shea
• Massively parallel, distributed information processor
• No central clock, instantaneous processing
• Accumulation rate of mRNA/protein varies per cell
• Progressive/sequential/ordered
• Reliable/reproducible
How is synchronicity maintained?
Ubiq=ubiquitous; Mat = maternal; activ = activator; rep = repressor;
unkn = unknown; Nucl. = nuclearization; χ = β-catenin source;
nβ-TCF = nuclearized b-β-catenin-Tcf1; ES = early signal;
ECNS = early cytoplasmic nuclearization system; Zyg. N. = zygotic Notch
Copyright © 2001-2008 Hamid Bolouri and Eric Davidson

Multi-scale problem

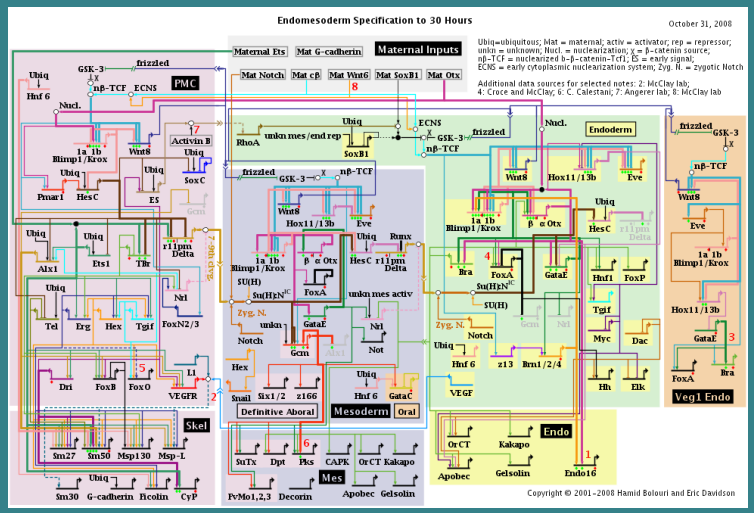
20,000-30,000 genes

2,000-3,000 “regulatory” genes

200-300 genes in control system

20-30 genes in spec. subcircuits

2-3 genes/interactions critical



Channels/gradients - milliseconds

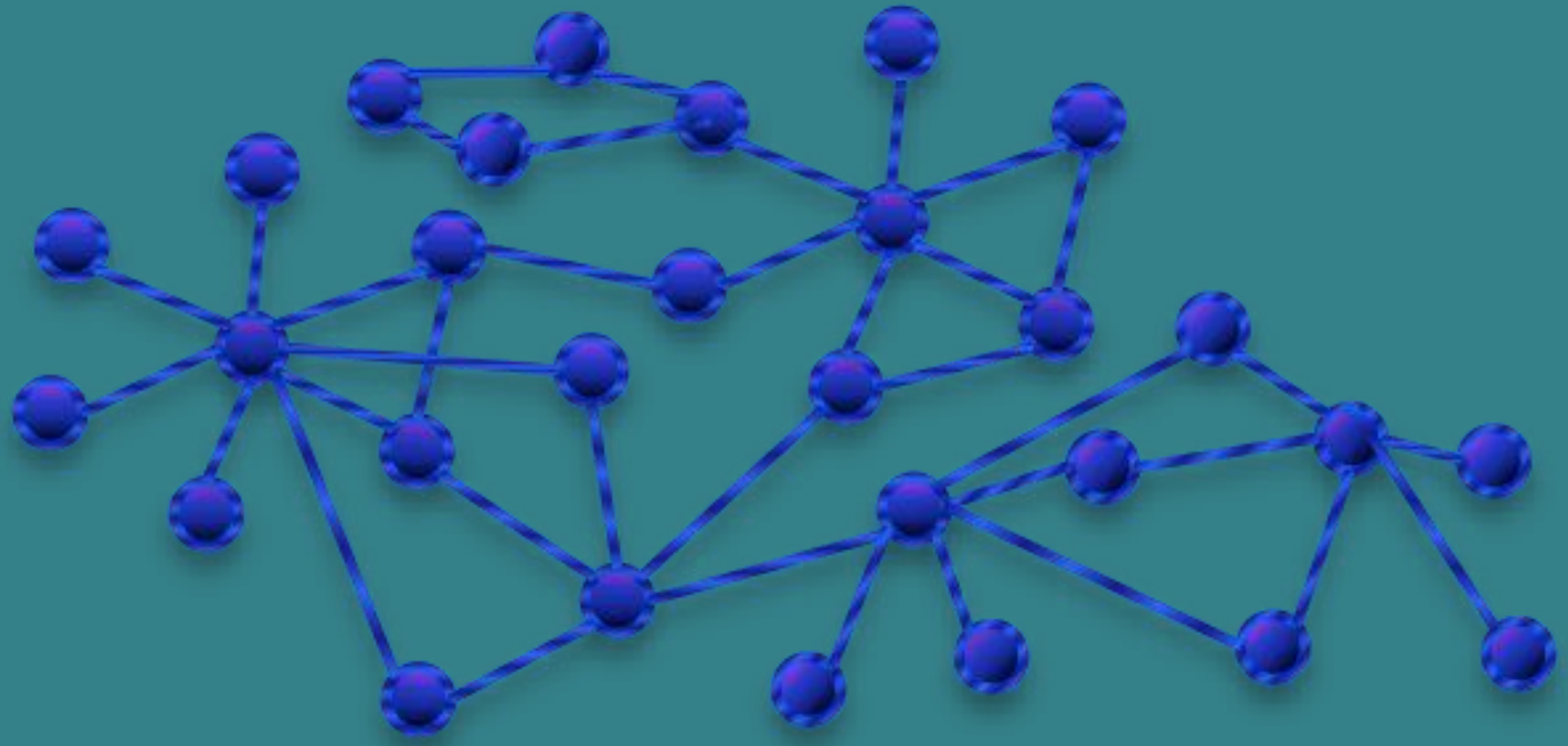
Signaling cascades - milliseconds to minutes

Transcription - minutes to hours

Scale-integration

Which genes are relevant?	Transcriptome to Interactome
How are they functionally related?	Interactome to Network
Which interactions drive specific outcomes?	Network to Subcircuit

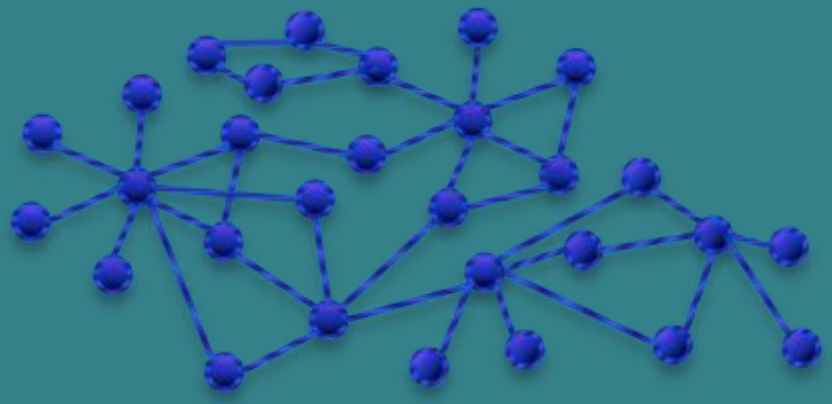
Who regulates the regulators?

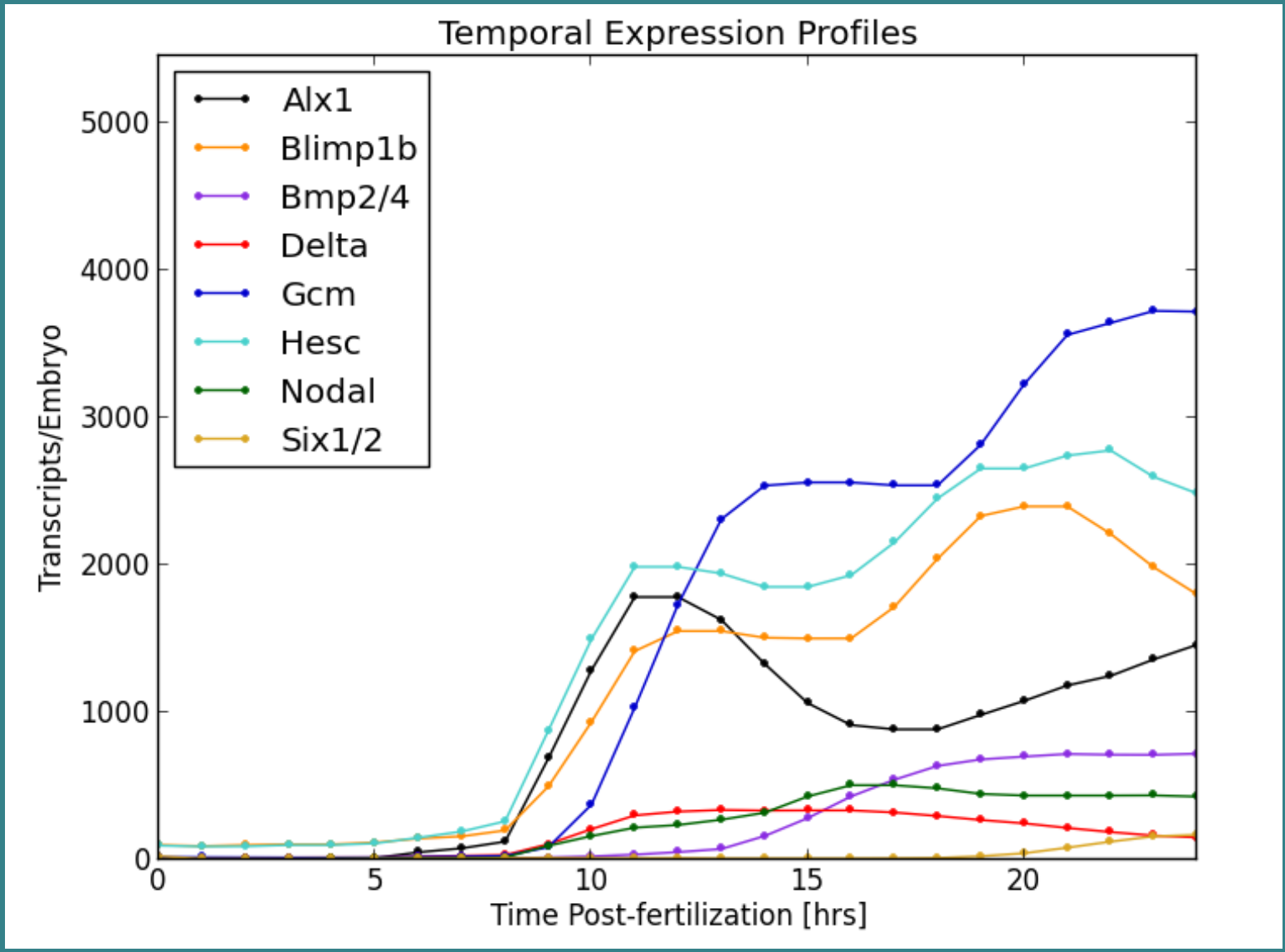


Who regulates the regulators?

Interactome “reverse-engineering”:

- the statistical dependency of each gene with every other gene
- requires hundreds of replicates
- successful with human disease biology

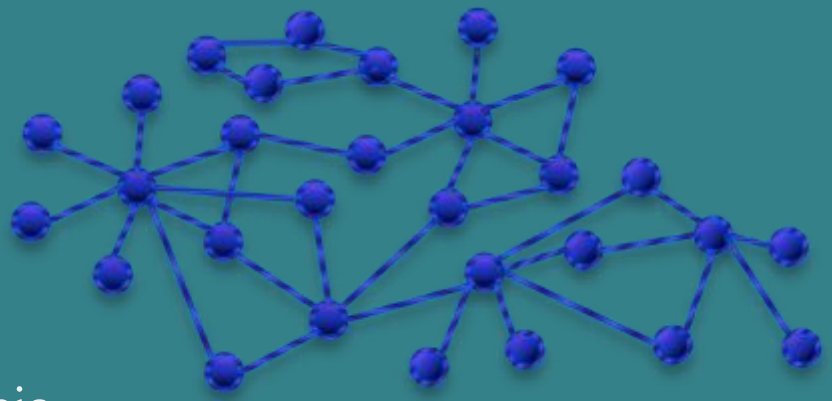




Who regulates the regulators?

Interactome "reverse-engineering":

- the statistical dependency of each gene with every other gene
- requires hundreds of replicates
- successful with human disease biology

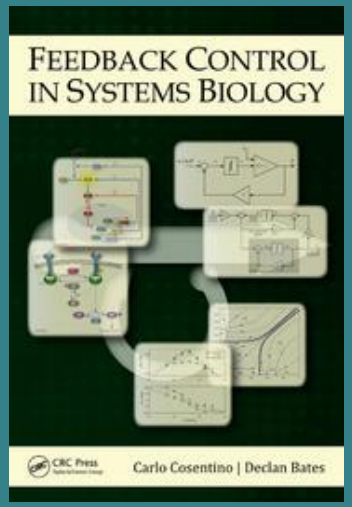


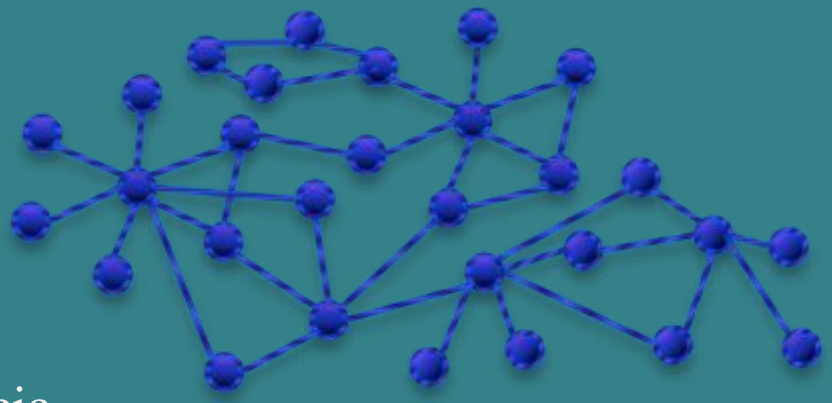
Carlo Cosentino
University of Magna Graecia,
Catanzaro, Italy

Who regulates the regulators?

Interactome "reverse-engineering":

- the statistical dependency of each gene with every other gene
- requires hundreds of replicates
- successful with human disease biology



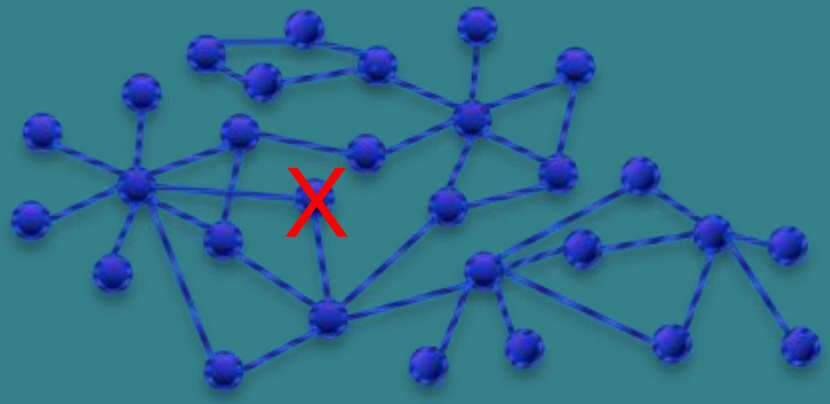


Carlo Cosentino
University of Magna Graecia,
Catanzaro, Italy

Scale-integration II: perturbations

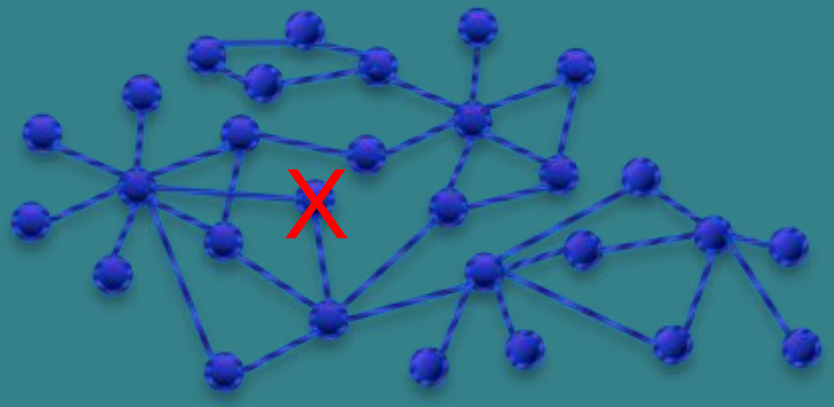
Which genes are relevant? Transcriptome to Interactome

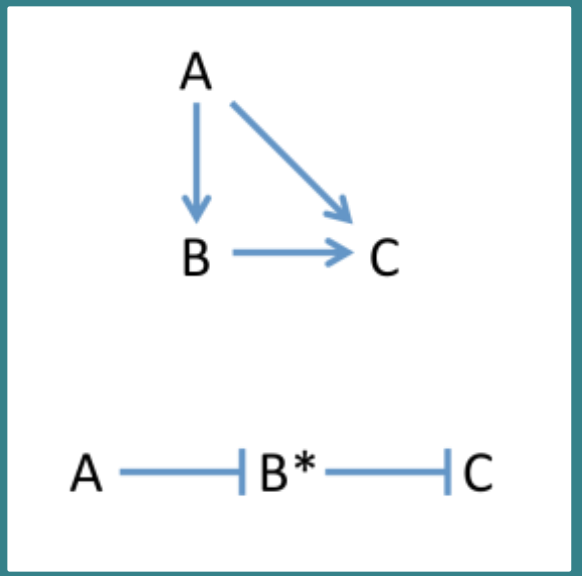
How are they functionally related? Systematic perturbations



Scale-integration II: perturbations

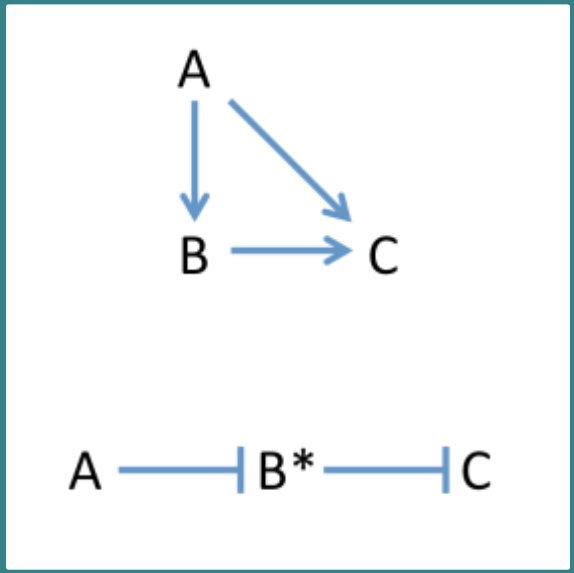
Even exhaustive perturbations fail to resolve common, functionally important network structures

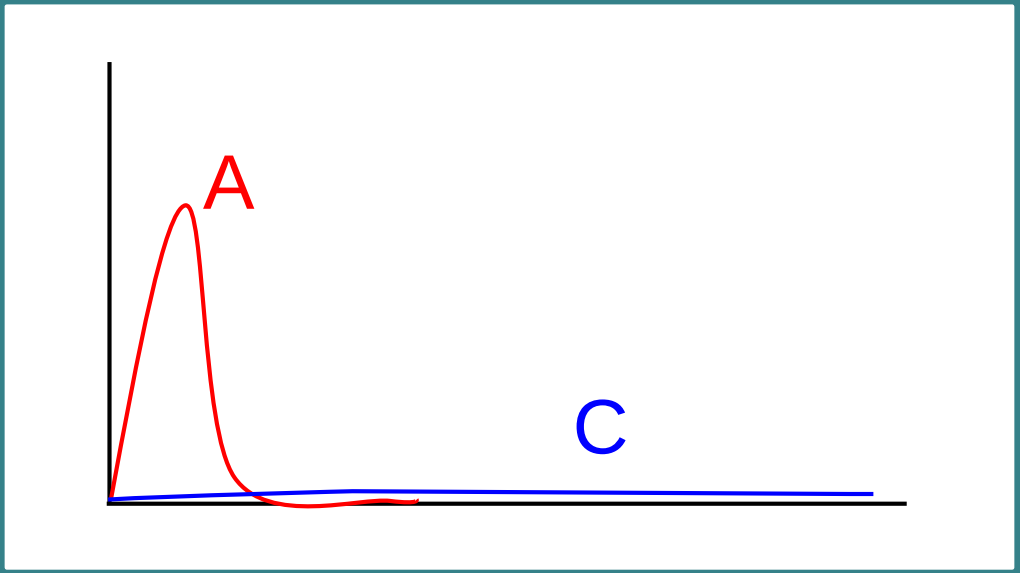




Scale-integration II: perturbations

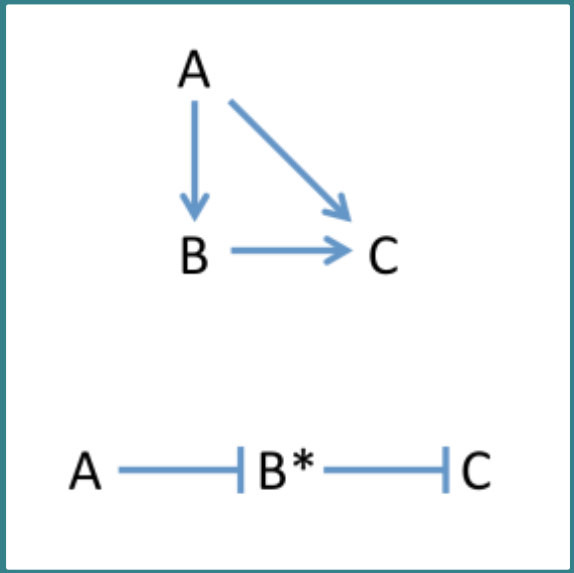
Perturbation analysis failure

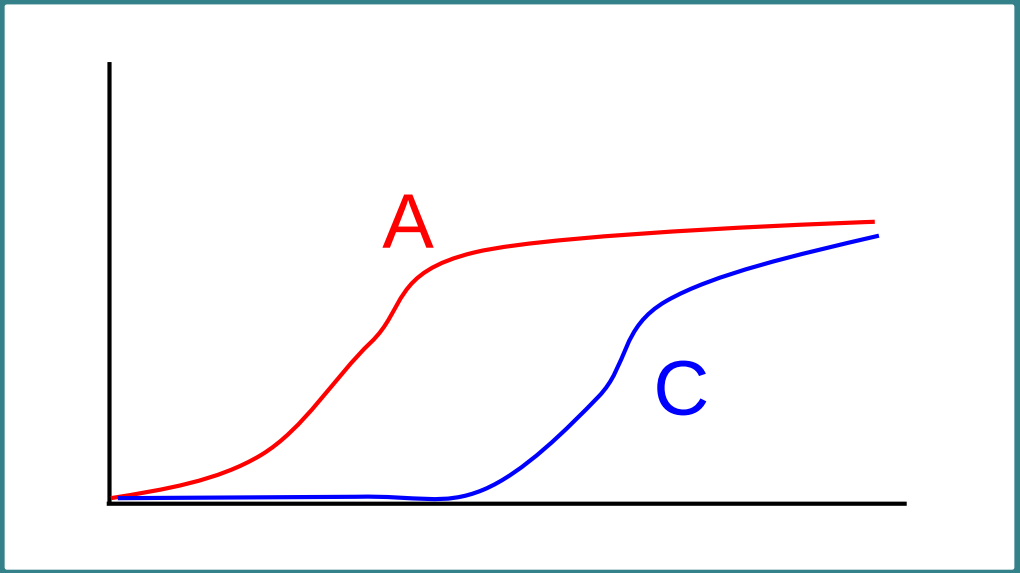




Scale-integration II: perturbations

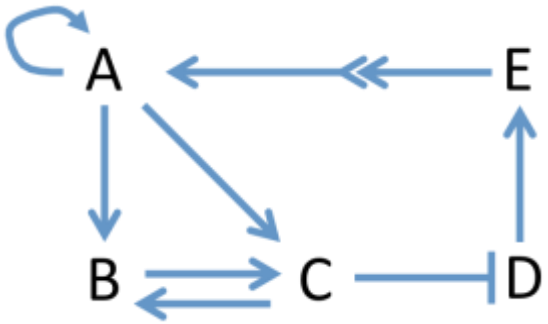
Perturbation analysis failure

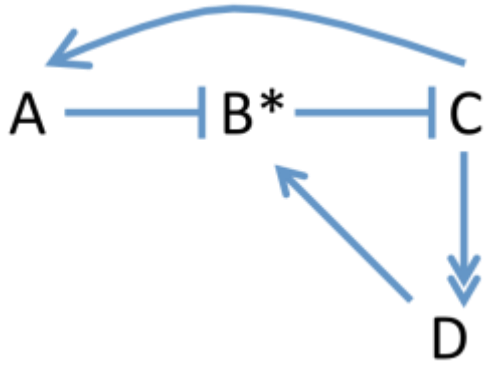


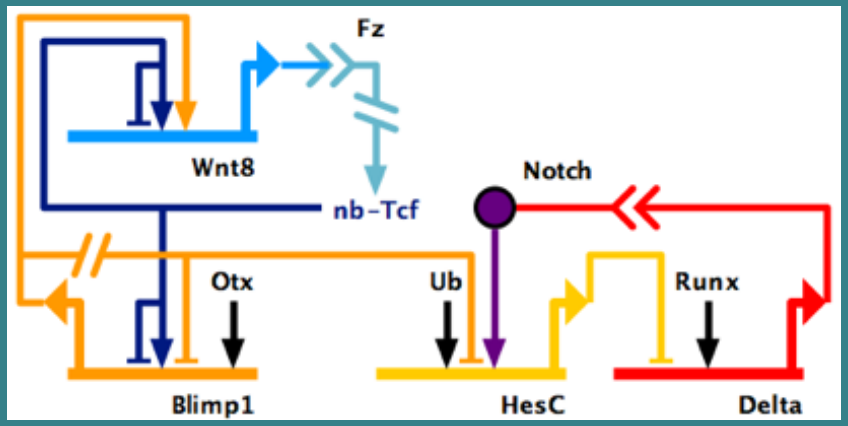


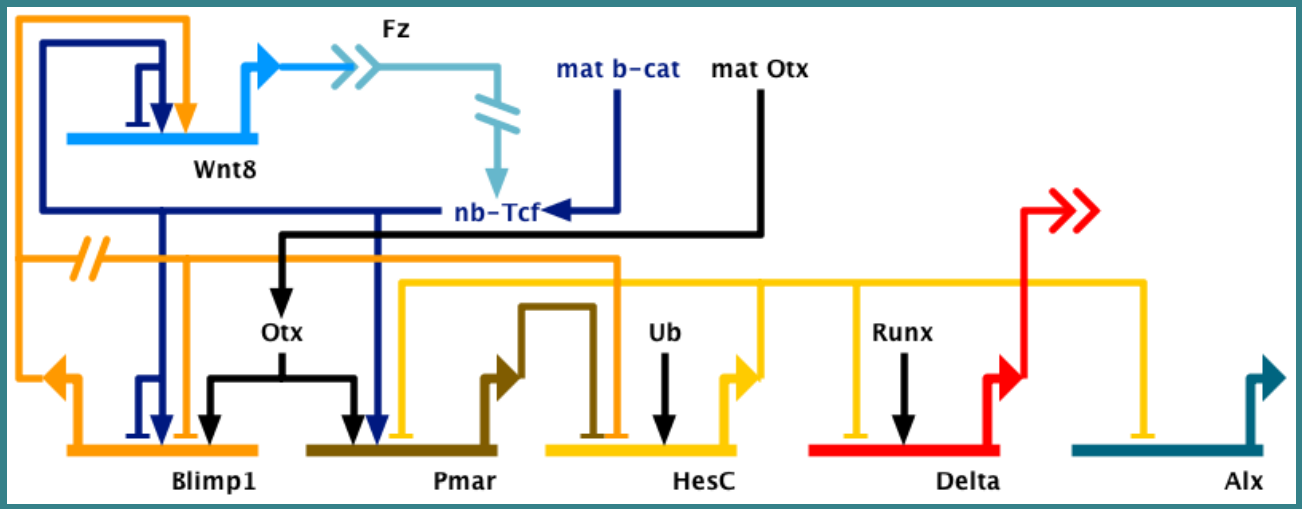
Scale-integration II: perturbations

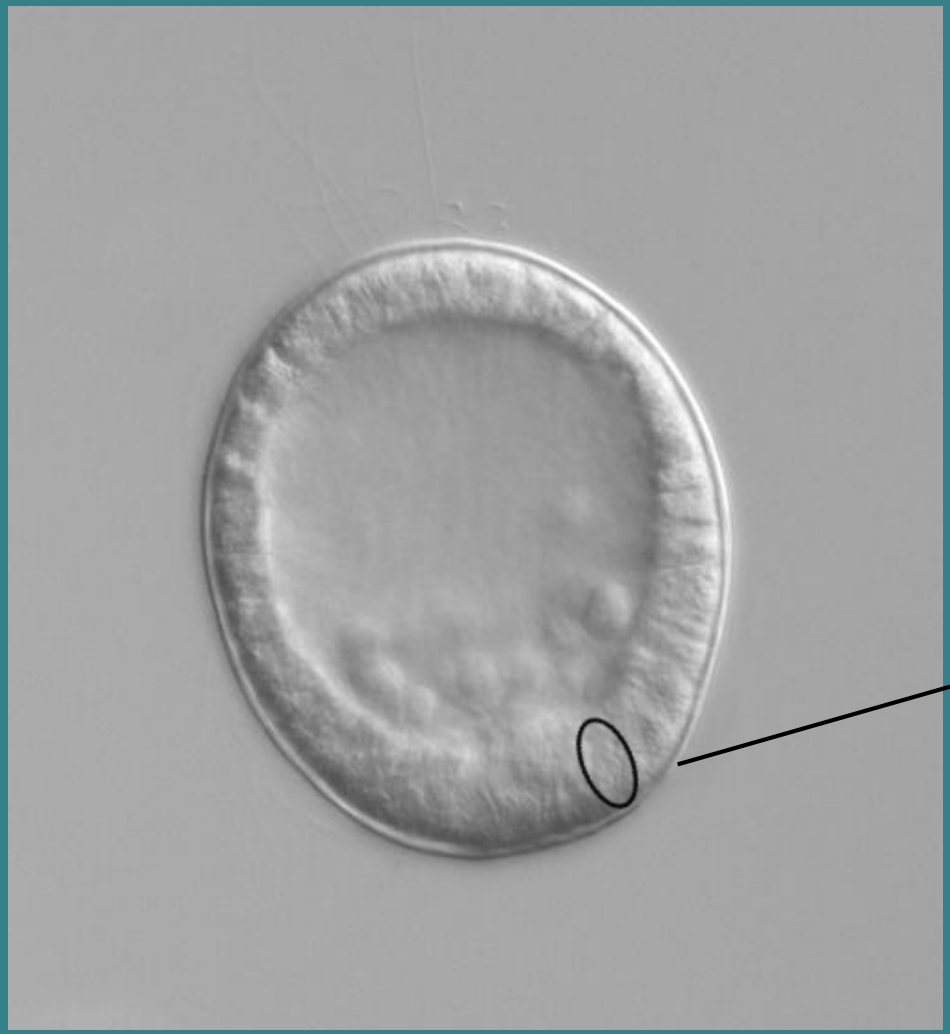
Perturbation analysis failure

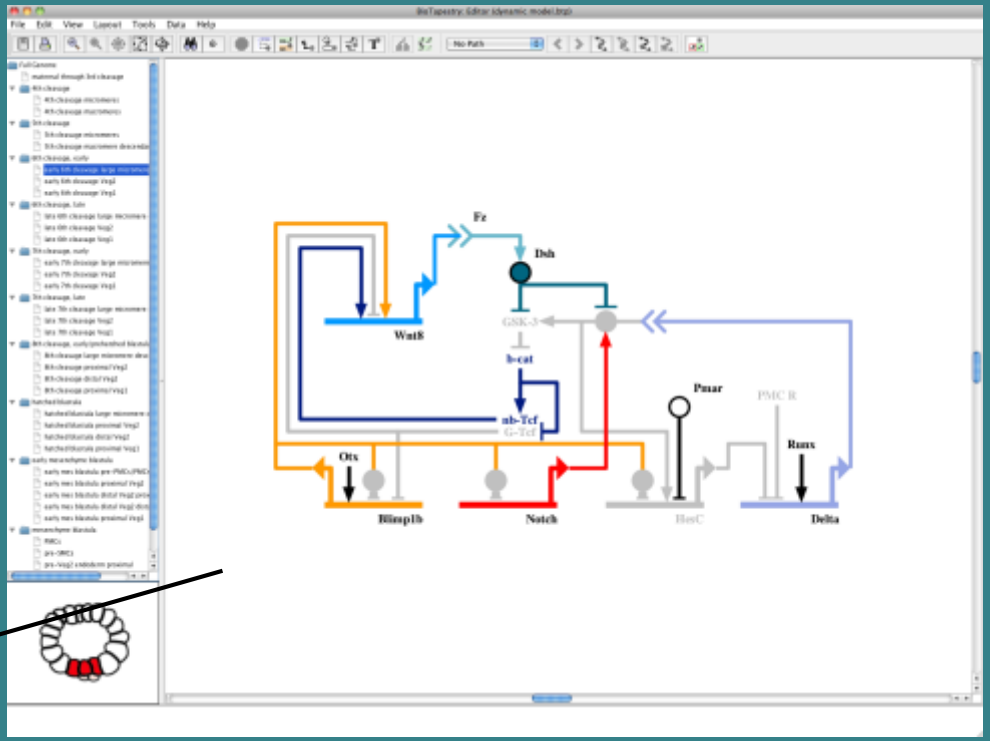






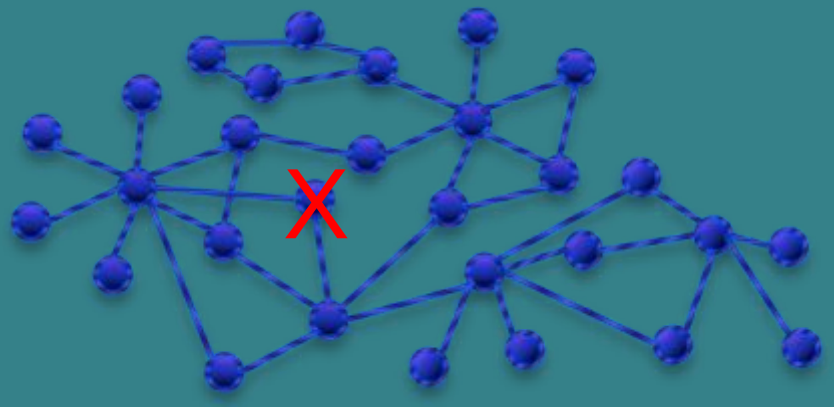






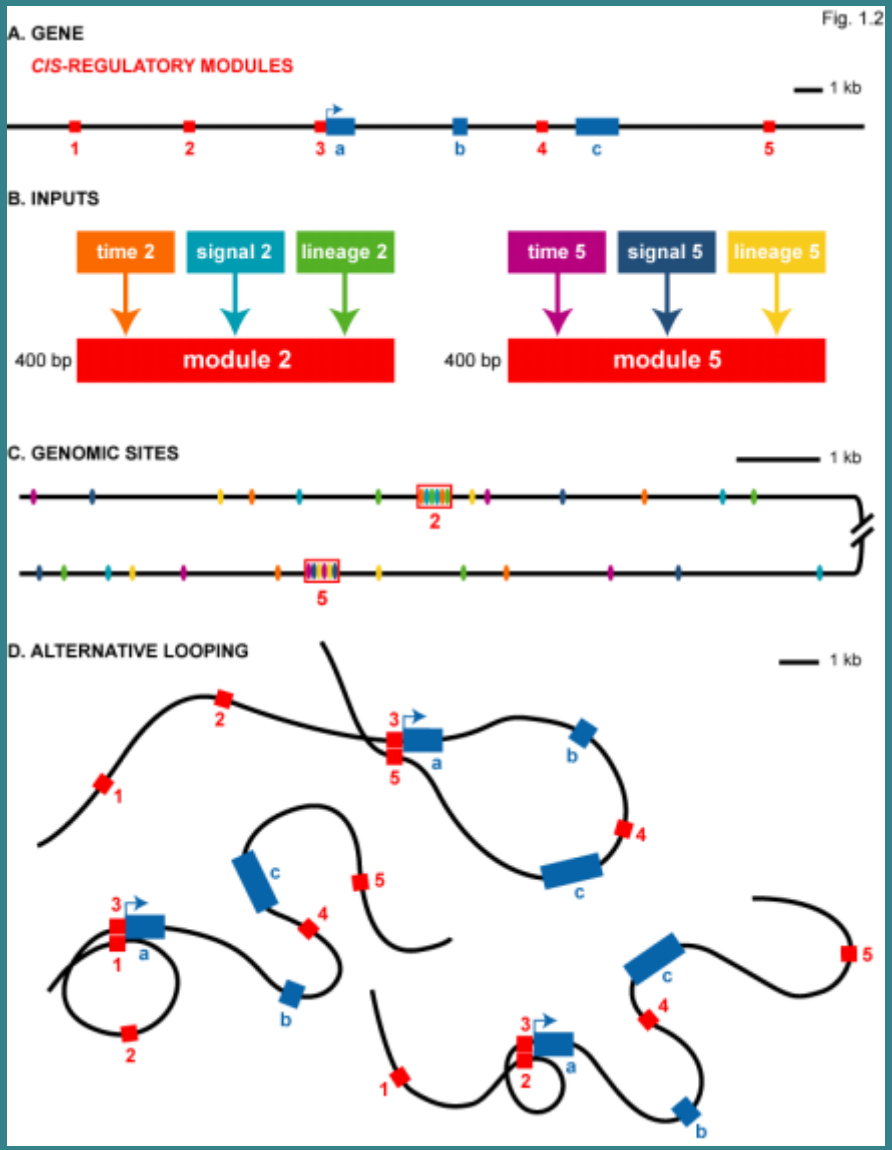
Scale-integration II: perturbations

Even exhaustive perturbations fail to resolve common, functionally important network structures



+

Complementary data sets

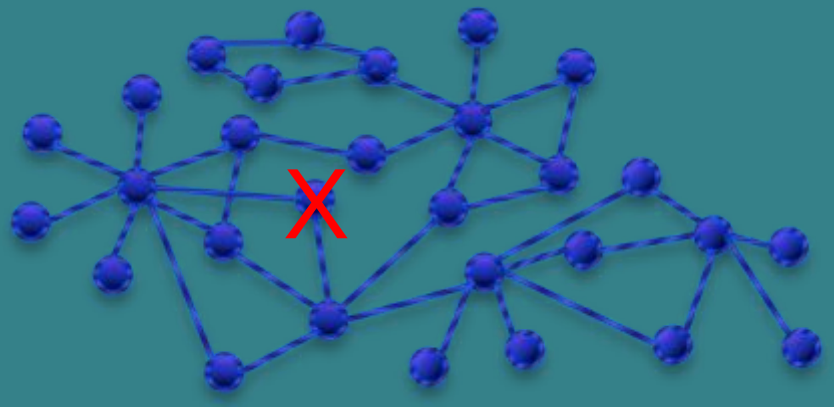


Genomic *cis*-regulatory elements

TF binding preferences

Scale-integration II: perturbations

Even exhaustive perturbations fail to resolve common, functionally important network structures

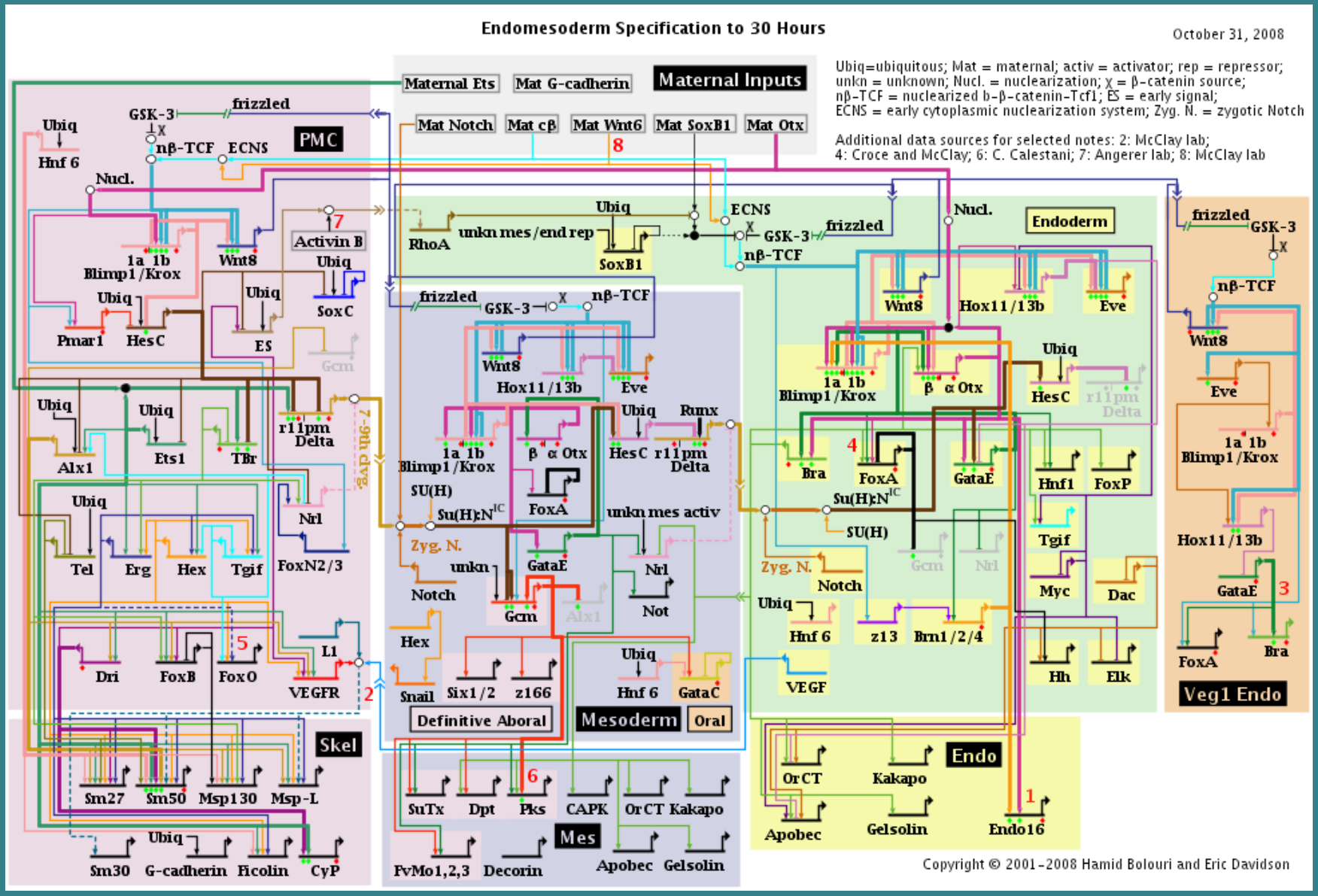


+

Complementary data sets

Scale-integration III: testing network switches

Which genes are relevant?	Transcriptome to Interactome
How are they functionally related?	Interactome to Network
Which interactions drive specific outcomes?	Network to Subcircuit

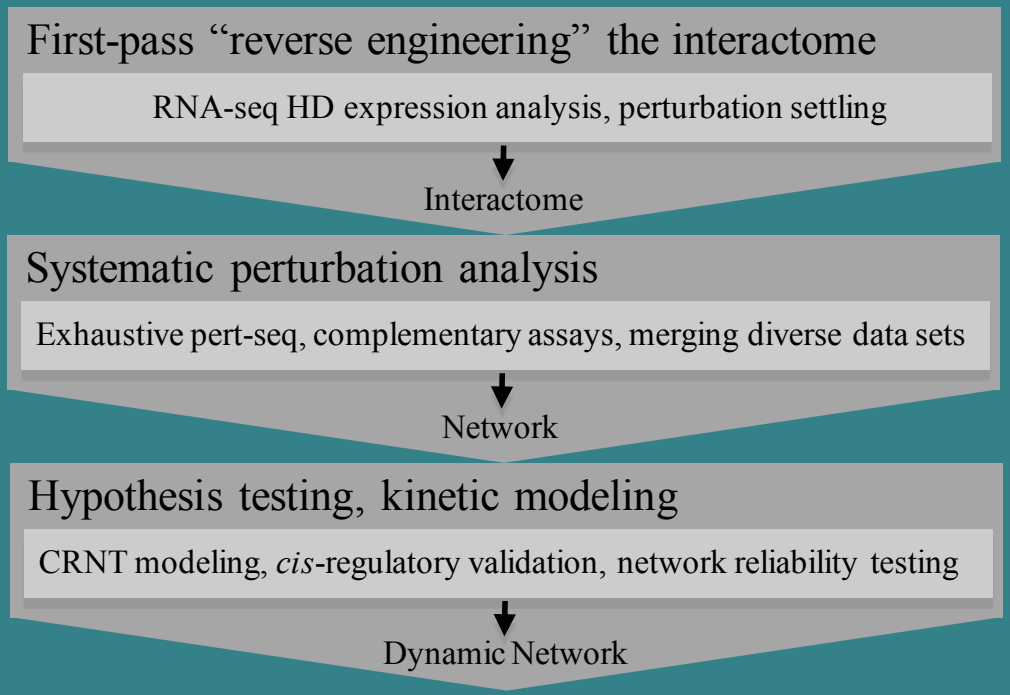


Scale-integration approach

Which genes are relevant?

How are they functionally related?

Which interactions drive specific outcomes?

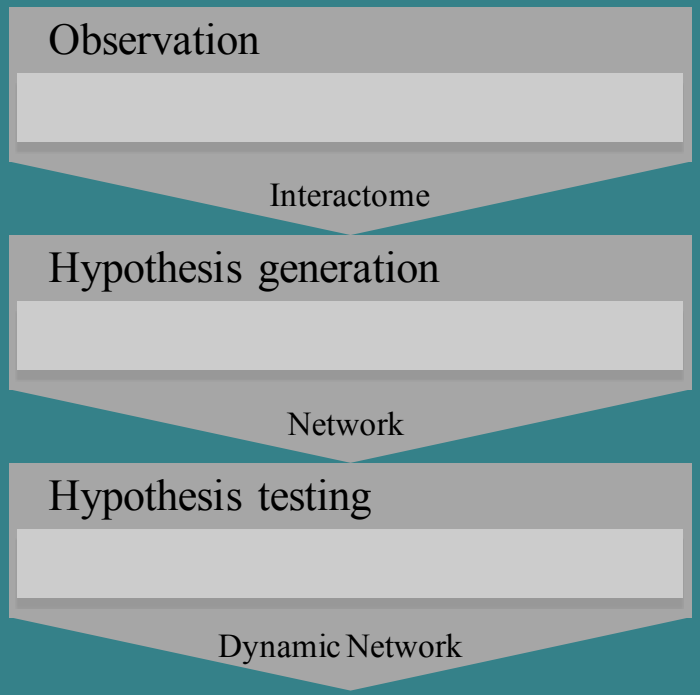


Scale-integration approach

Which genes are relevant?

How are they functionally related?

Which interactions drive specific outcomes?



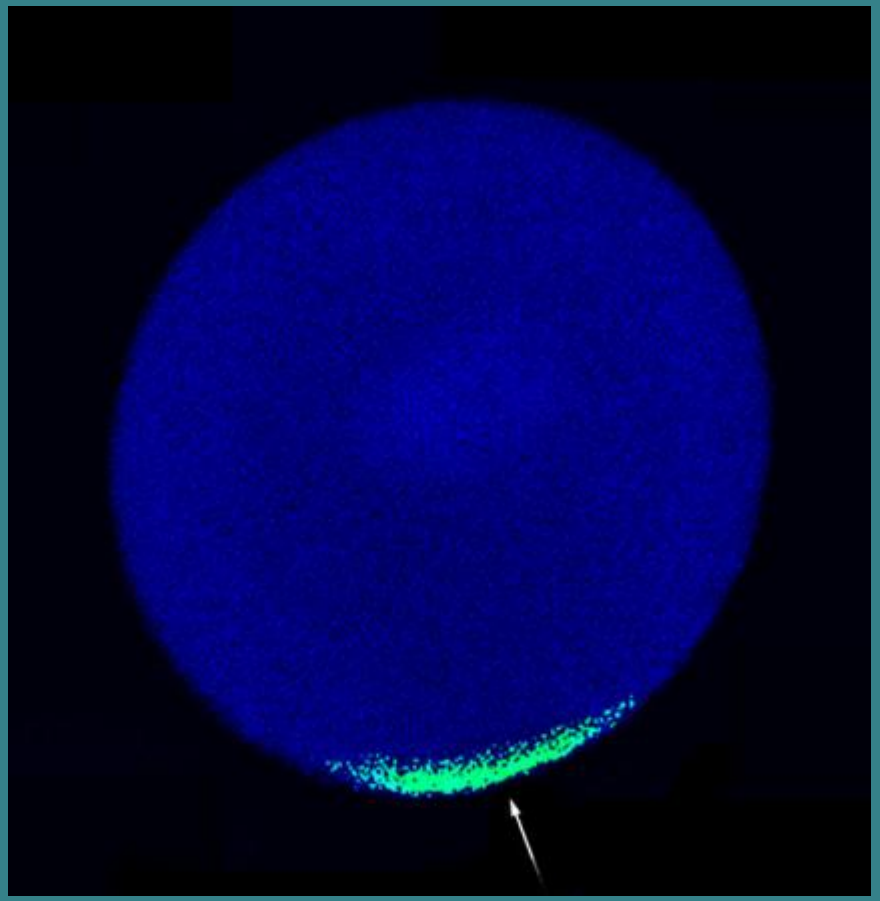
High Sensitivity
(Low Specificity)

Increasing
Specificity

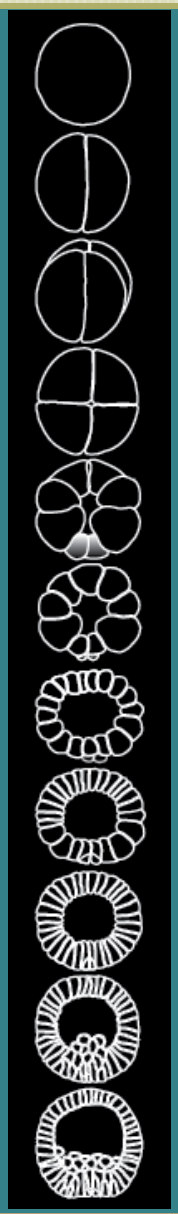
Cis-regulatory analysis

- BAC reporters capture genomic context
- *Cis*-reengineering tests subcircuit function within endogenous regulatory context

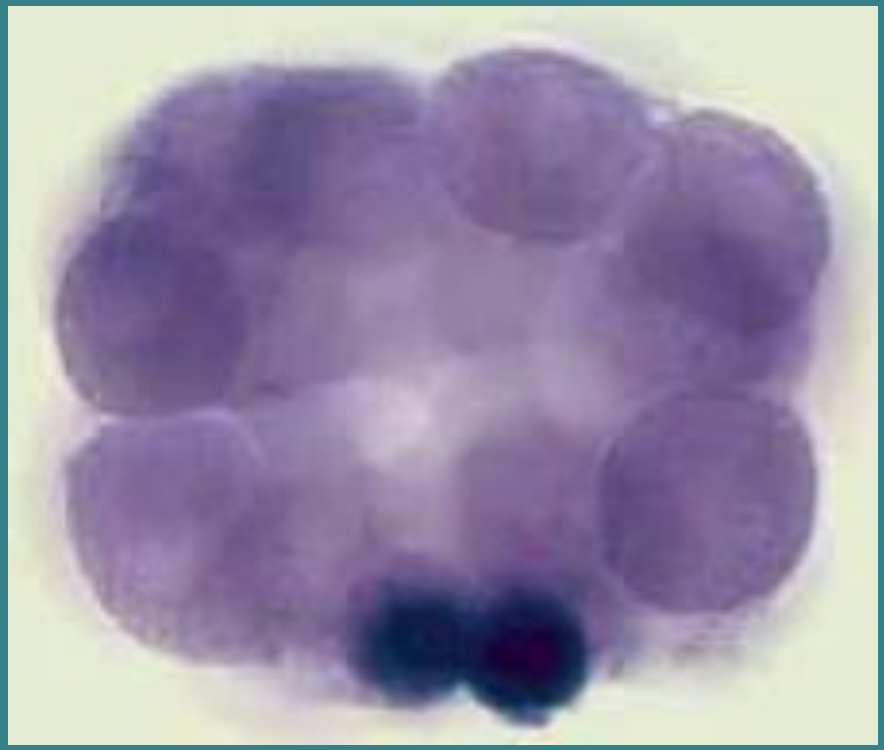
Vegetal Dsh



Ettensohn, et al.

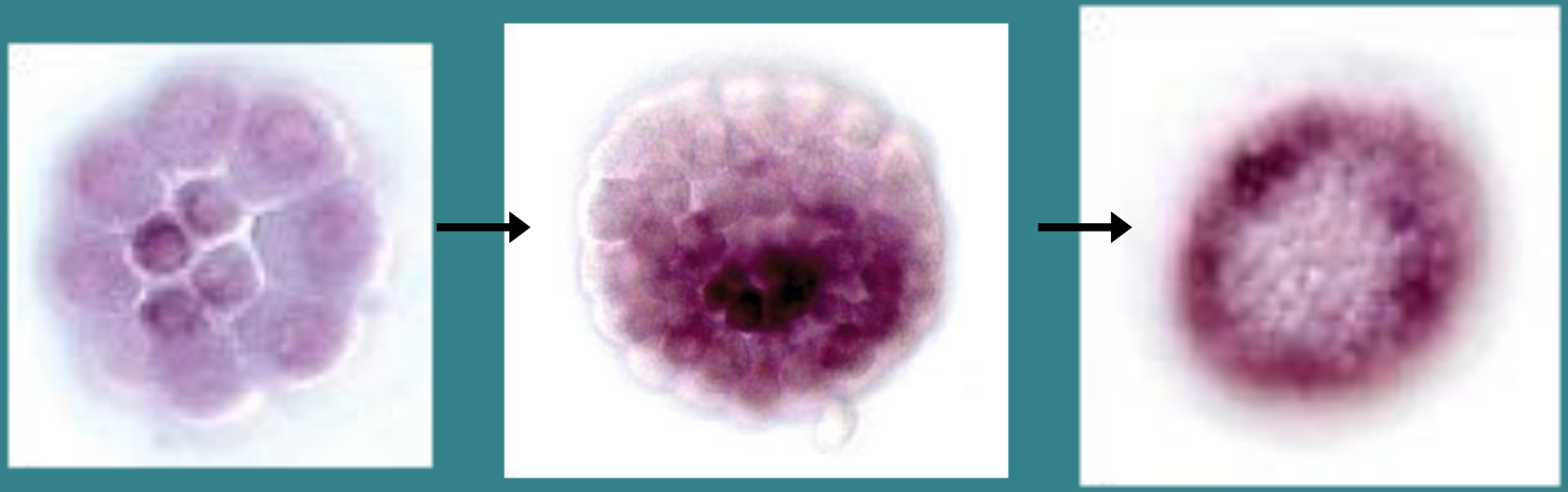


Vegetal *wnt8*

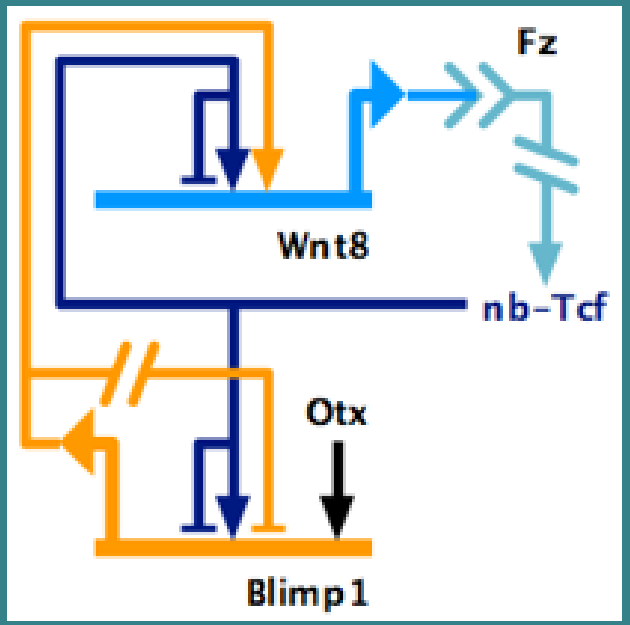


Wikramanayake, et al., Minokawa, et al.

The "Torus Subcircuit"



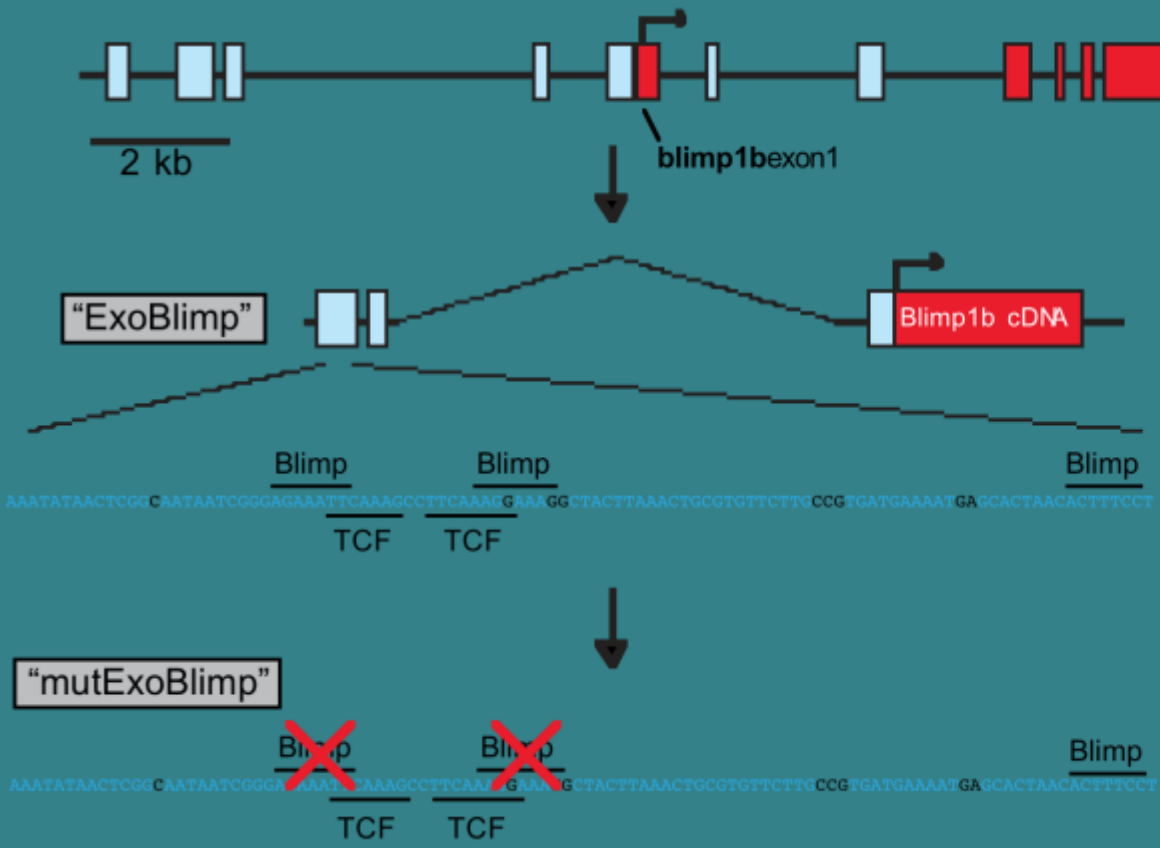
The "Torus Subcircuit"

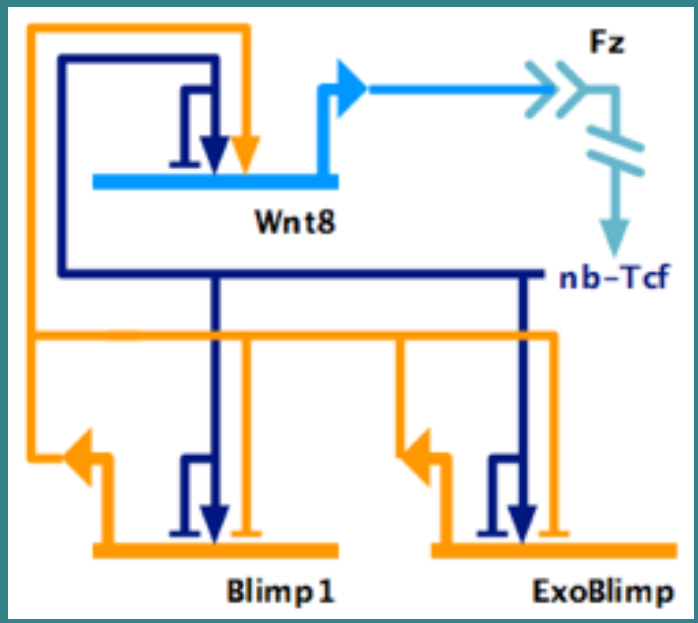


Expansion by ligand diffusion

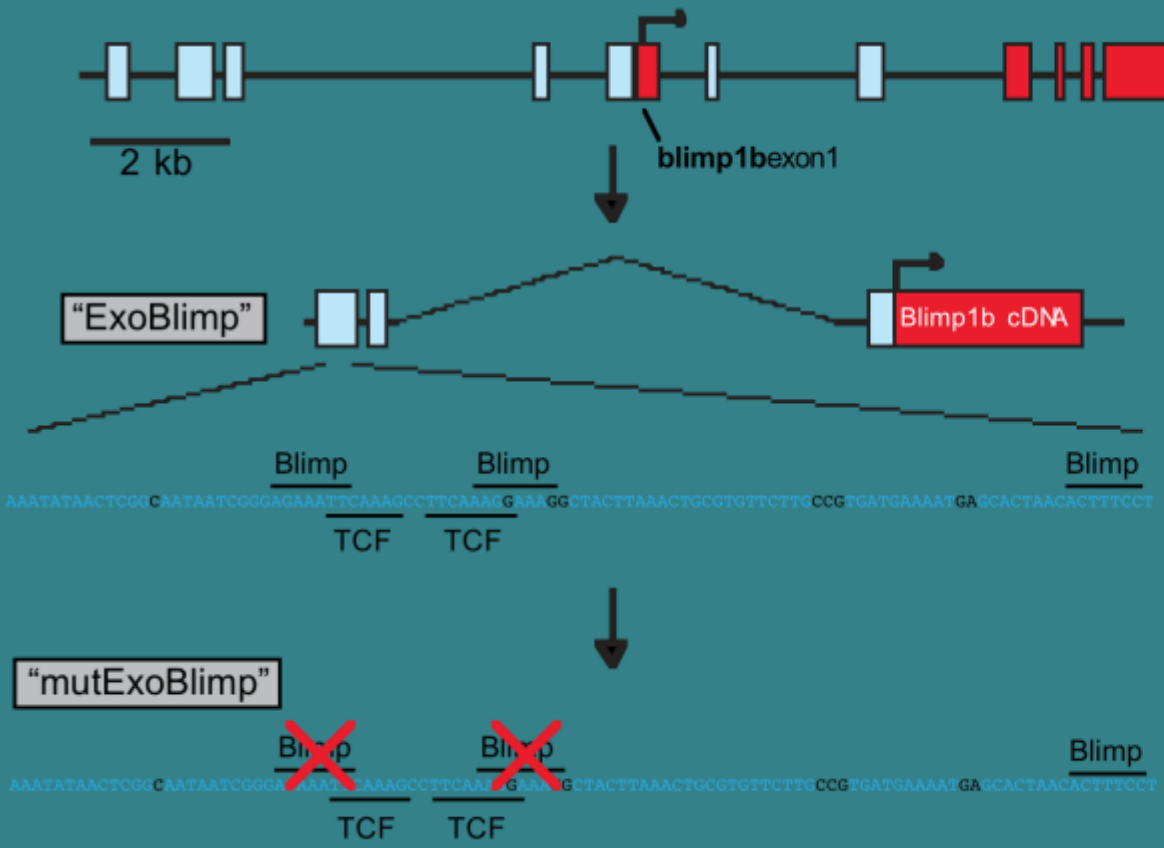
Shutd

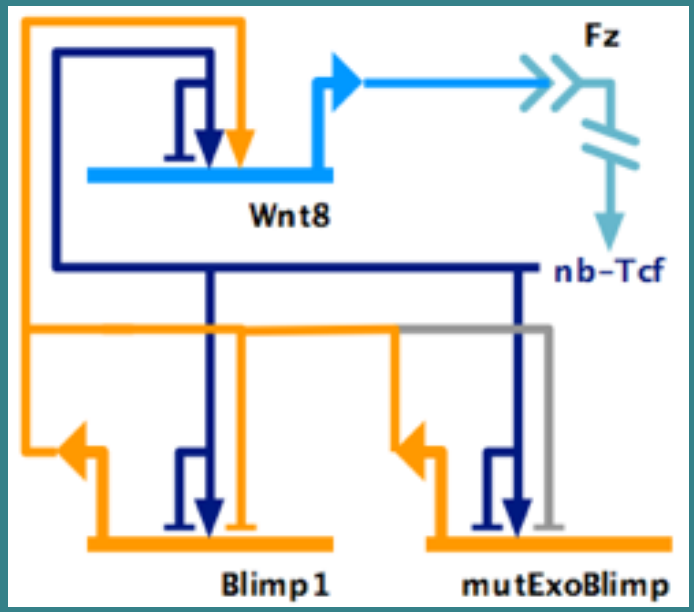
Testing the Torus Subcircuit: ExoBlimp



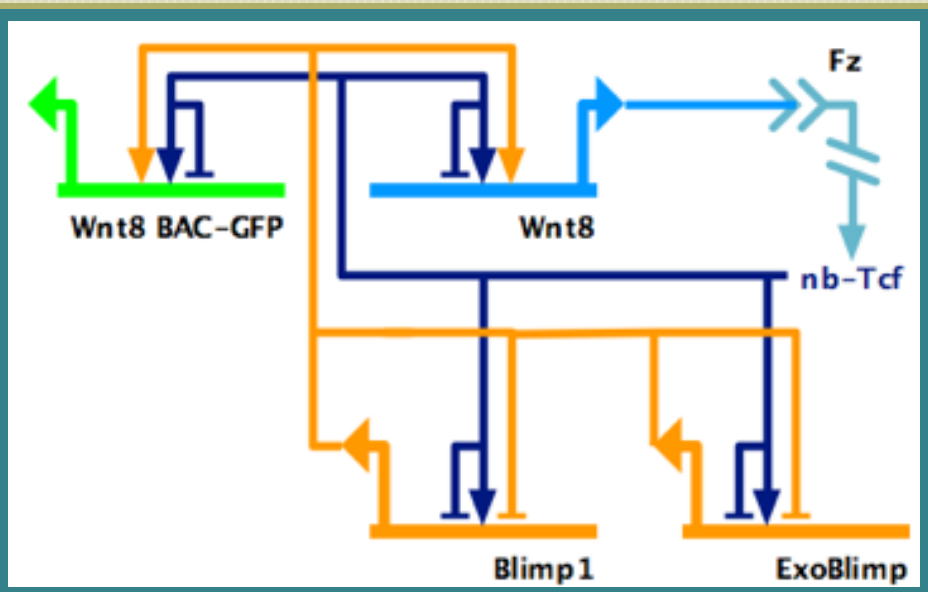


Testing the Torus Subcircuit: mutExoBlimp

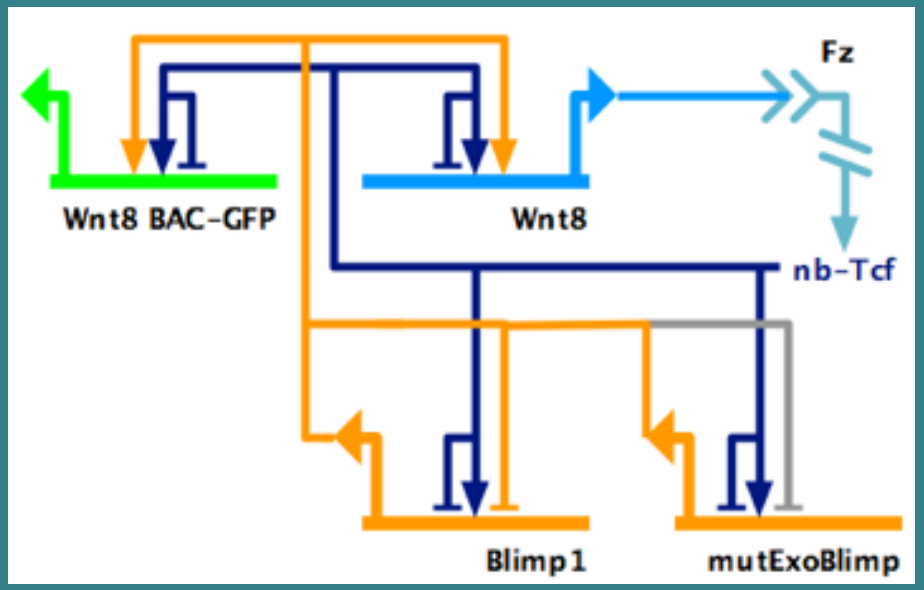




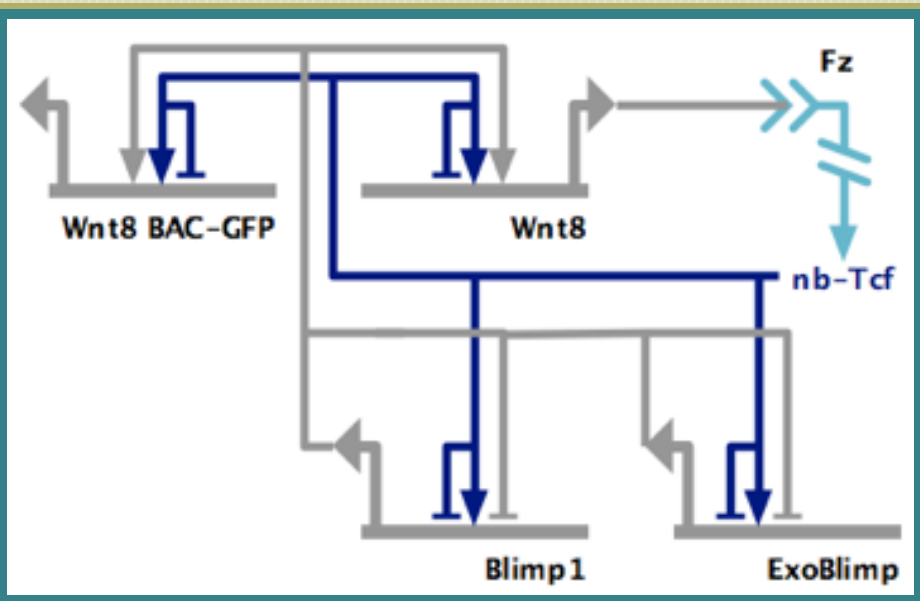
Control



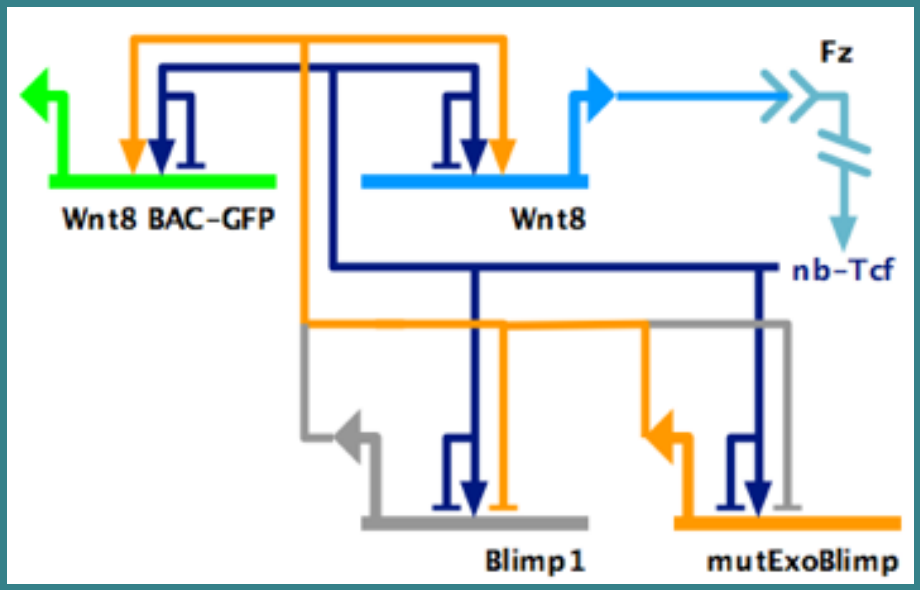
Test



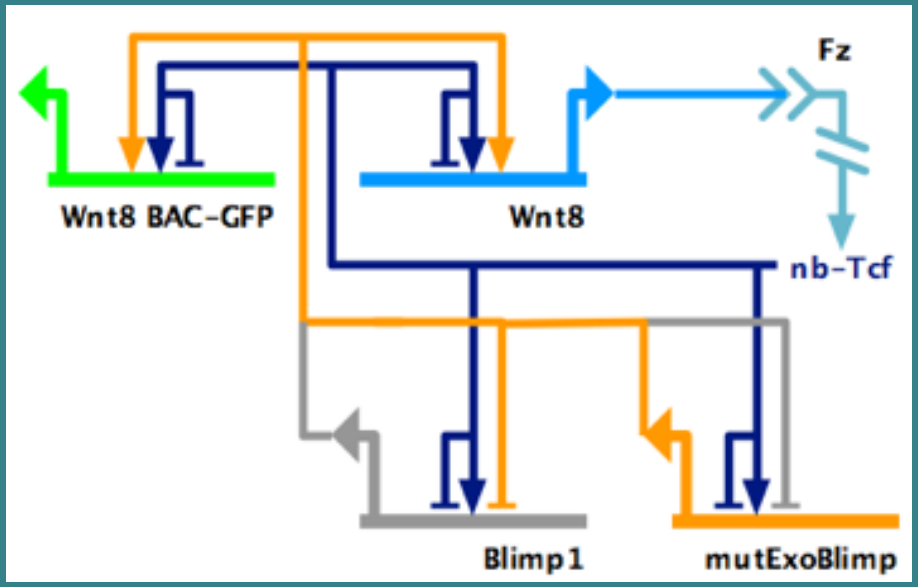
Control

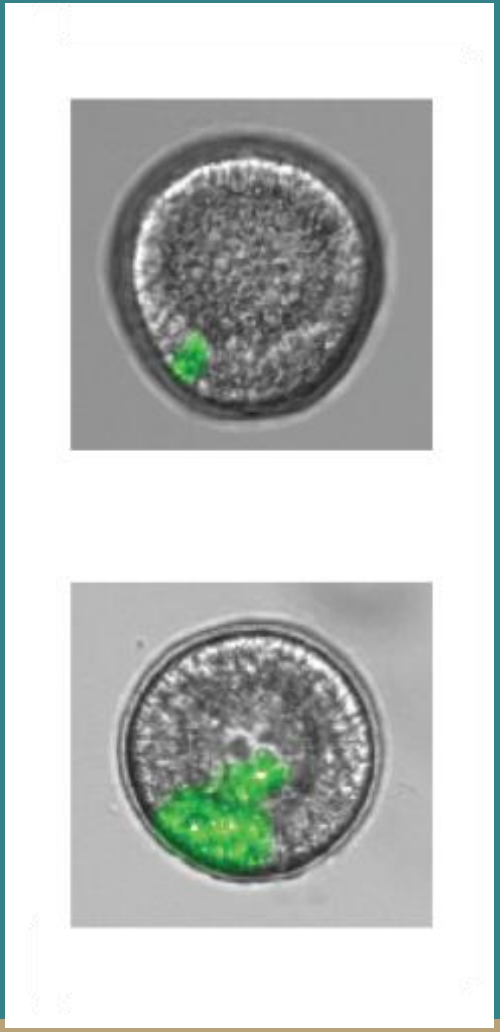


Test

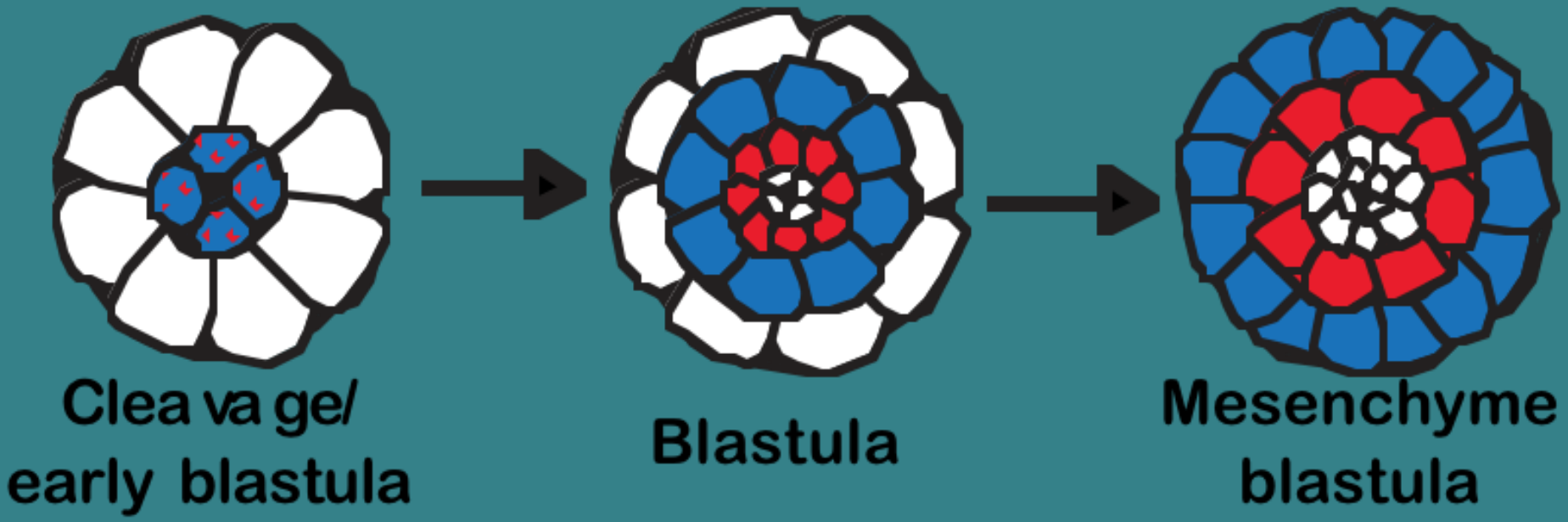


mutExoBlimp assay (I): Wnt8 BAC-GFP



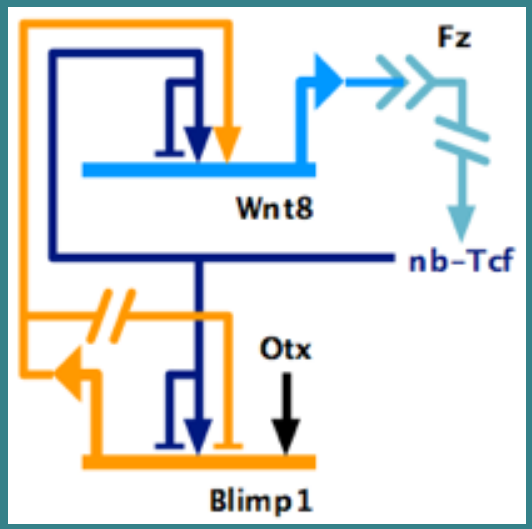


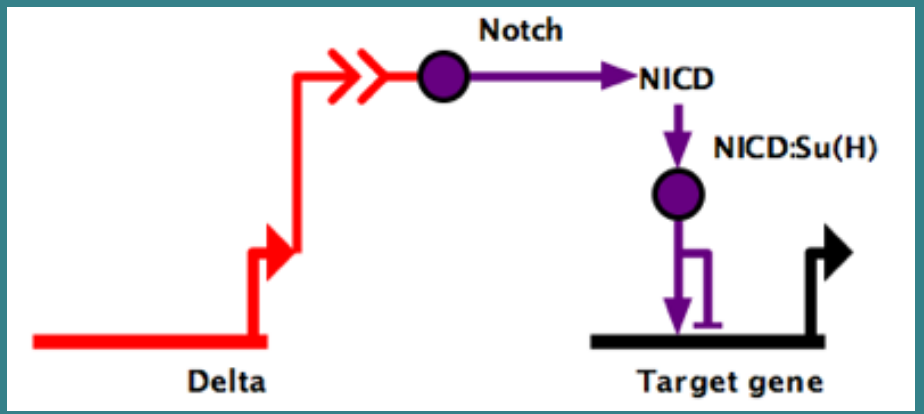
Wnt8 and Delta domains remain spatially aligned



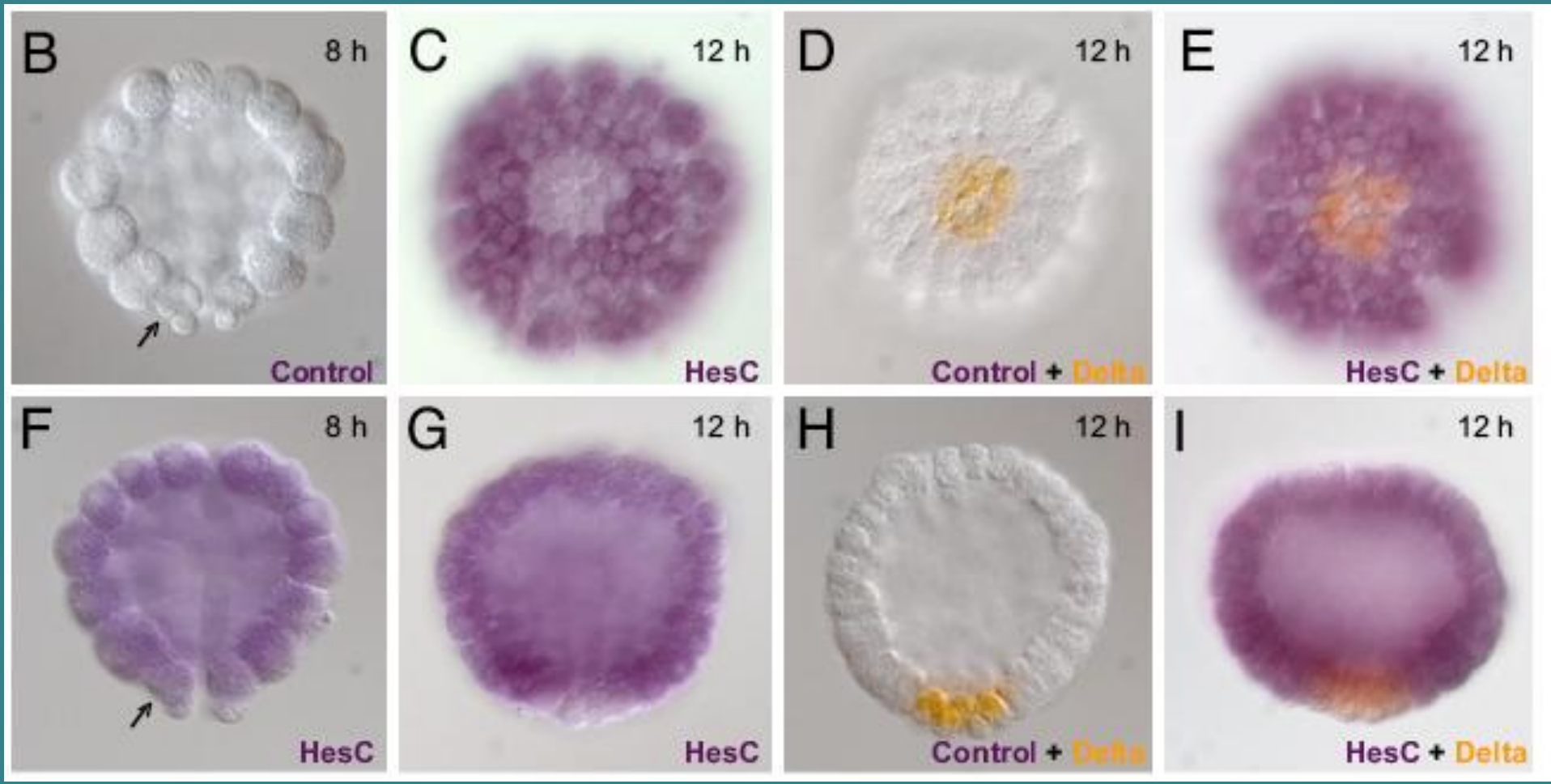
Blue = Wnt8/Blimp1
Red = Delta

Gene regulatory network subcircuit controlling a dynamic spatial pattern of signaling in the sea urchin embryo
Smith and Davidson, PNAS, **105**, 20089 (2008)

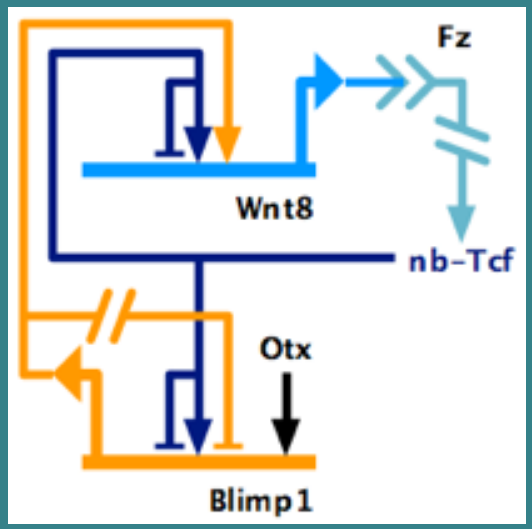


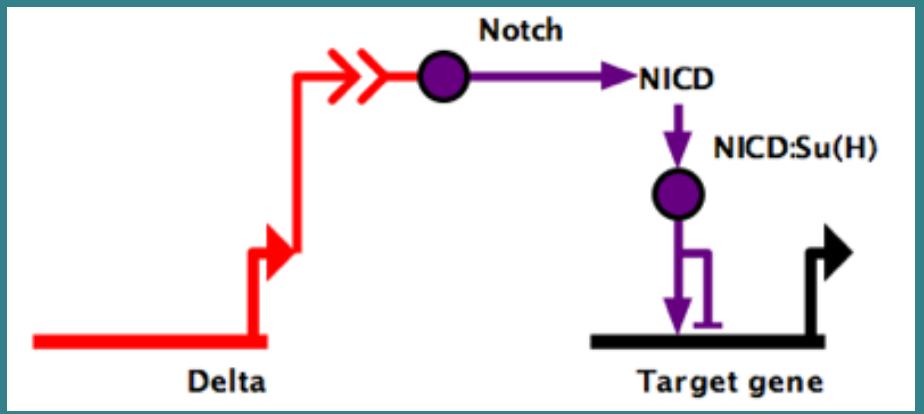


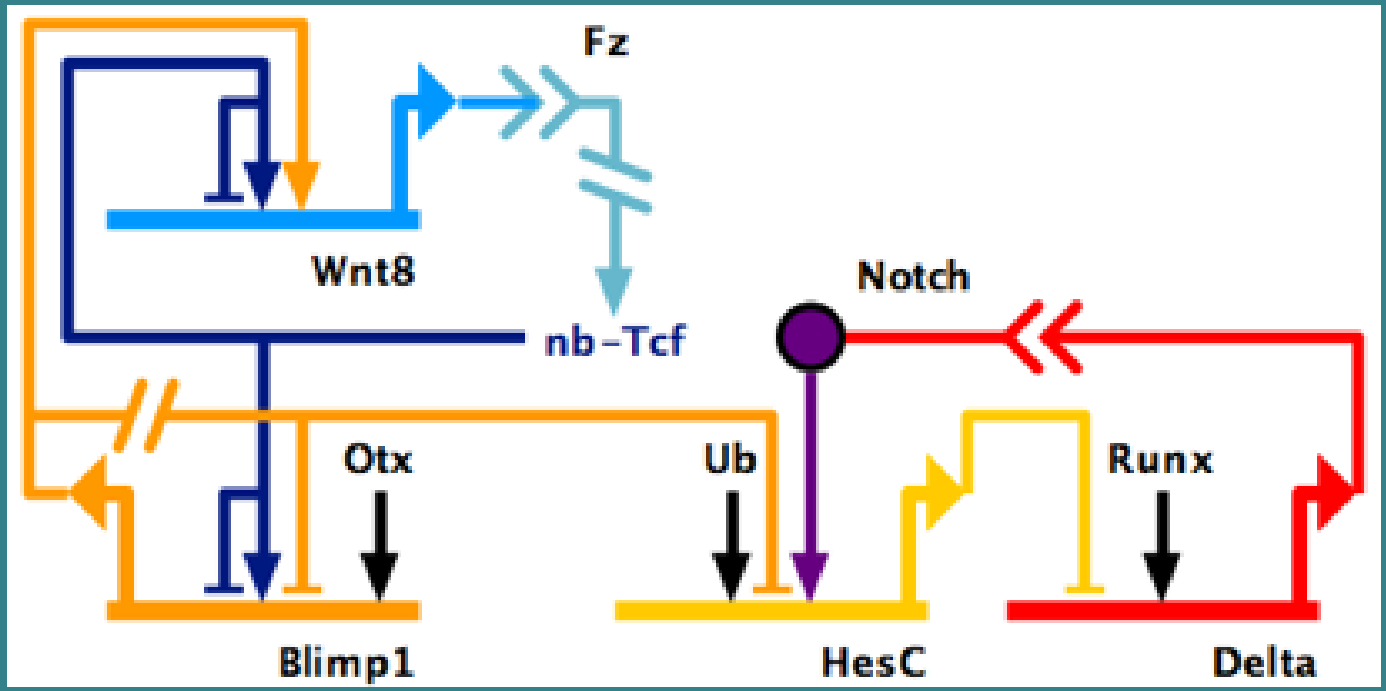
hesC and *delta* expression domains are exclusive



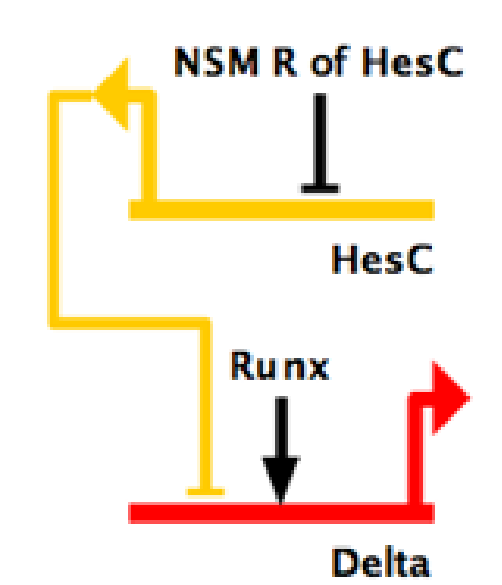
Revilla-i-Domingo *et al.*, 2007







cis-Regulatory analysis of *hesC*



ACGCTCGAGCGCCACGGGCACGTGCTGTCTATATATAGCGGTCGTCTGCCGGGTAGTAAGTCATAACA

ACCGCTCTCACGCTCGTCACAAGATCCATACTCCTCTTCGTGGATACCAGTGACTATCTCTCTGCCTC

TAAGGAATCTAACTCCTGTGATTATACTTCTTGCTTGTGCATTGAGACCATCTATATCAGCAATCATG

end exon 1 | begin intron 1

CTTACTTCATCTGGATACCAACAAATGGACATGTGCTCTAACAGACCTAGGGTAAGTTCTTCTCAACC

ATTTCTTCAATCACACATTTTCTCTTTTTGAATTCAACAAGTTTAAGTCGAAACCGTGTCATCAGTAA

CCATGTTTACTTAGTAAGTTAACAGTAATCACGTTAACTTGTTCATCATGATCATTGATCTAATCATT

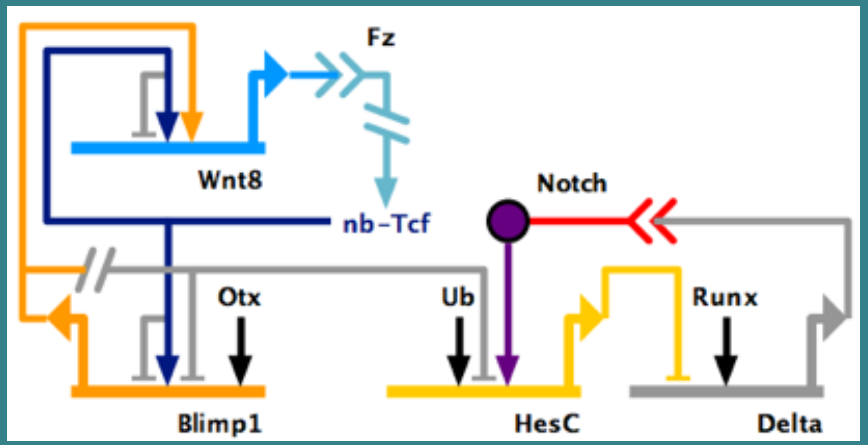
GATCATTGATTAATAGTTGATTTCAGCAAAGGCCAAATACCTTCAATCTACTTTCAACTTAACTGACA

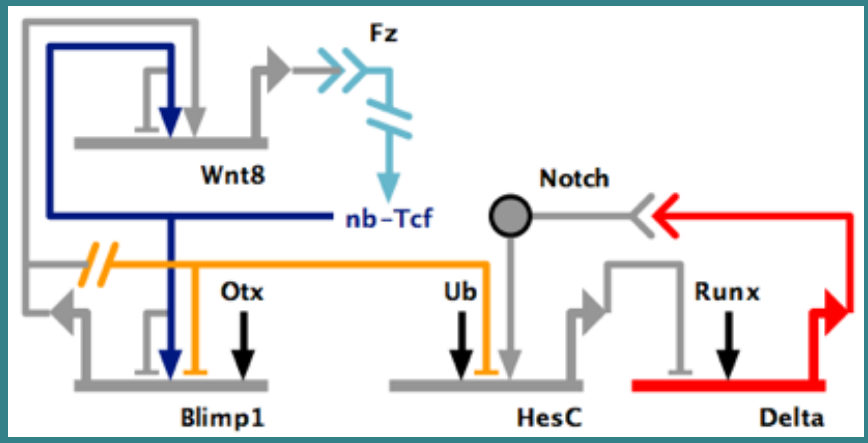
ATAACATCAAGGAACGAACATCAGACTCGTATGTTGATAAATTGATTTCATATTTGTATGATTTCCAT

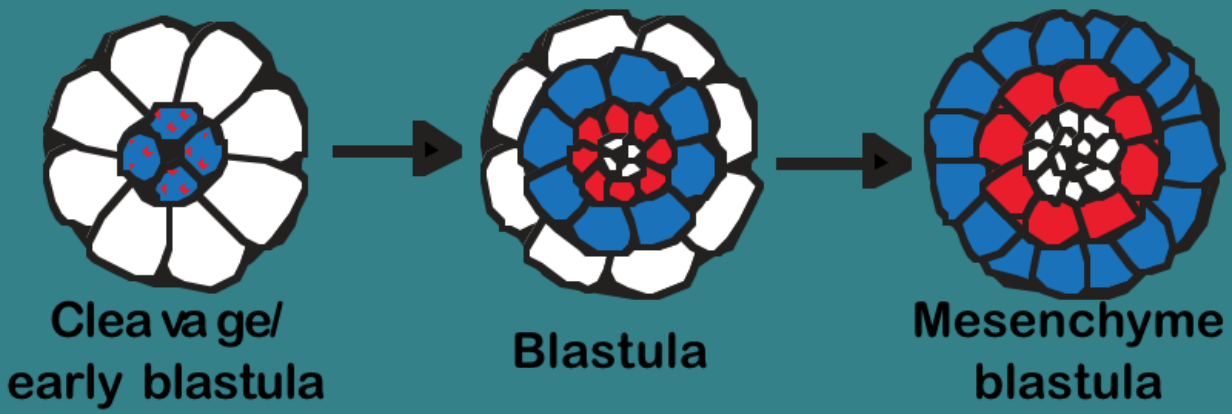
end intron 1

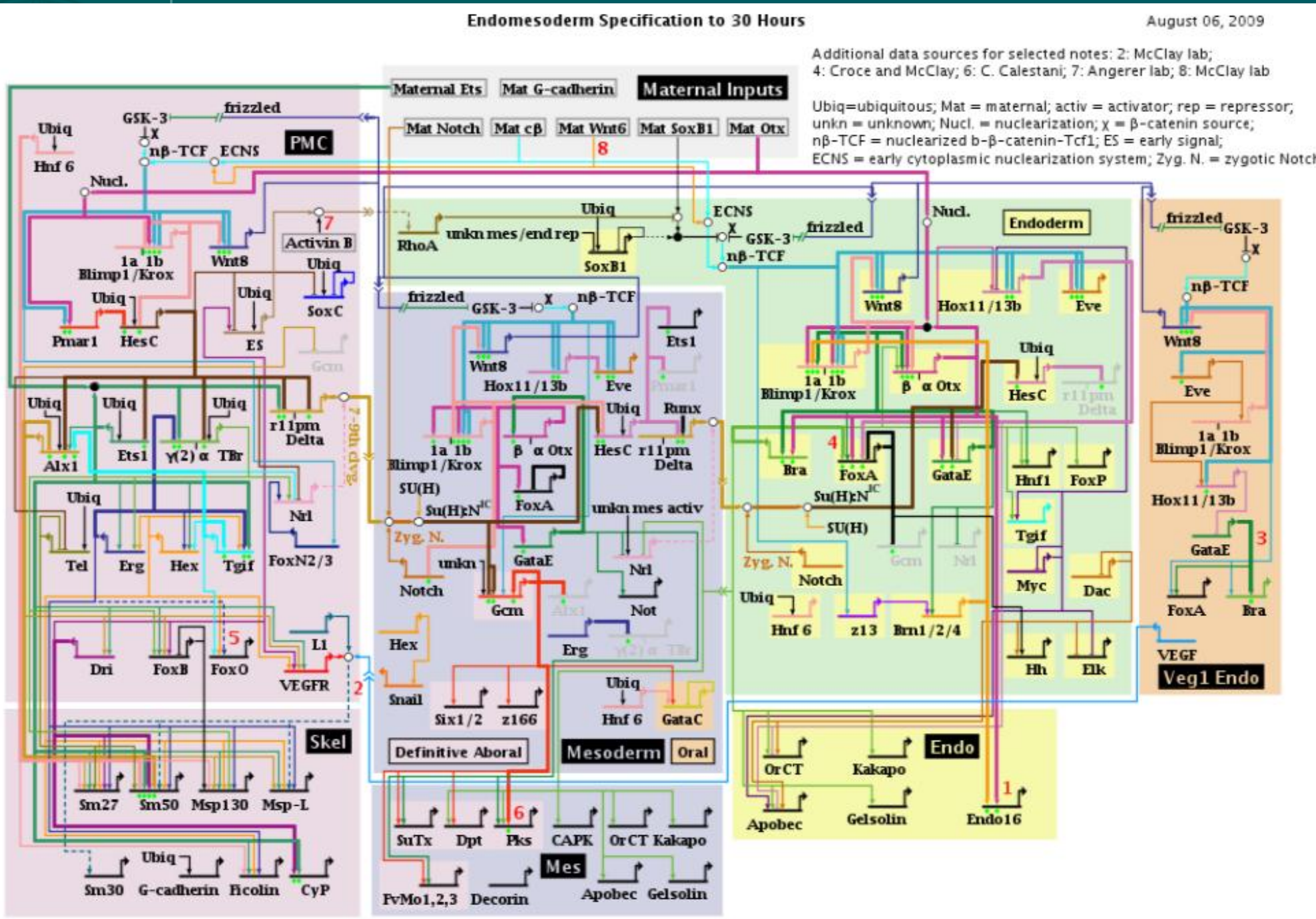
GCACTTGTATTTTCAATTTGAGGACTGTTTGACTTATTCTTGTACCCTTTCTTGTATTCATCTACA

5´(A/C/T)(A/G)(G/T)N<u>GAAAG</u>(G/T)(A/G/T)-3´

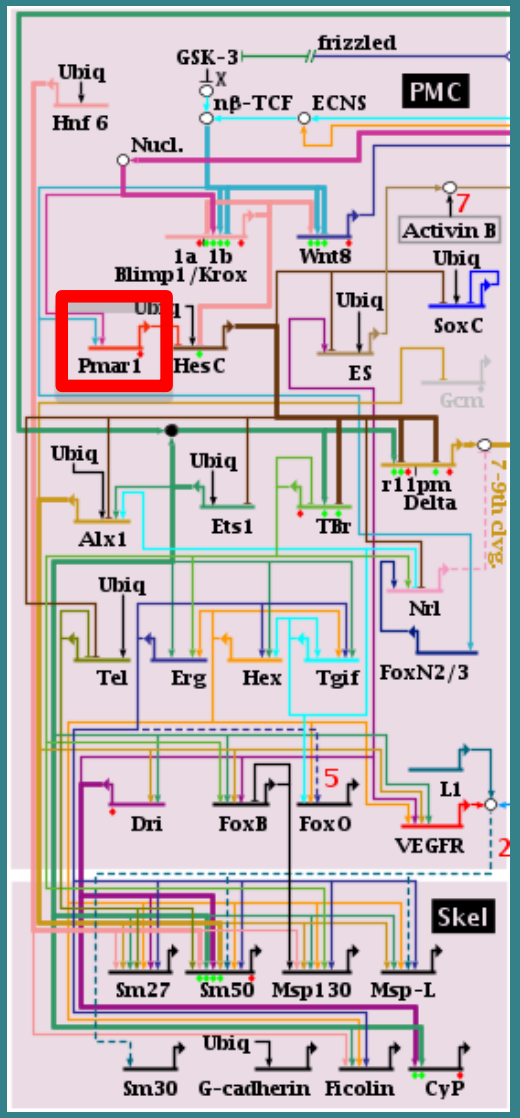


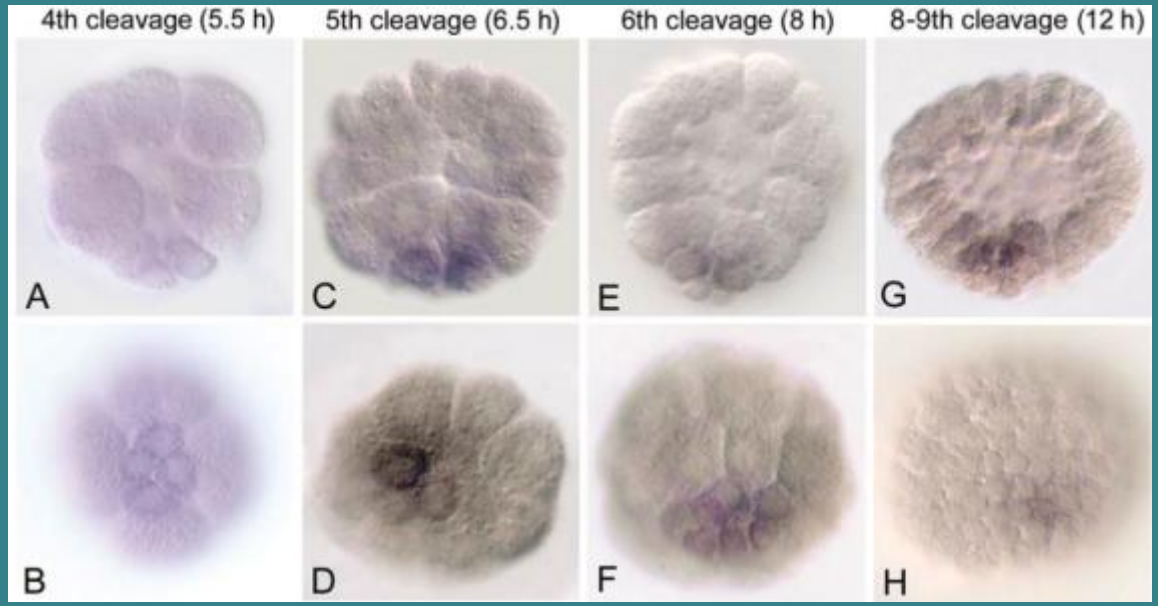


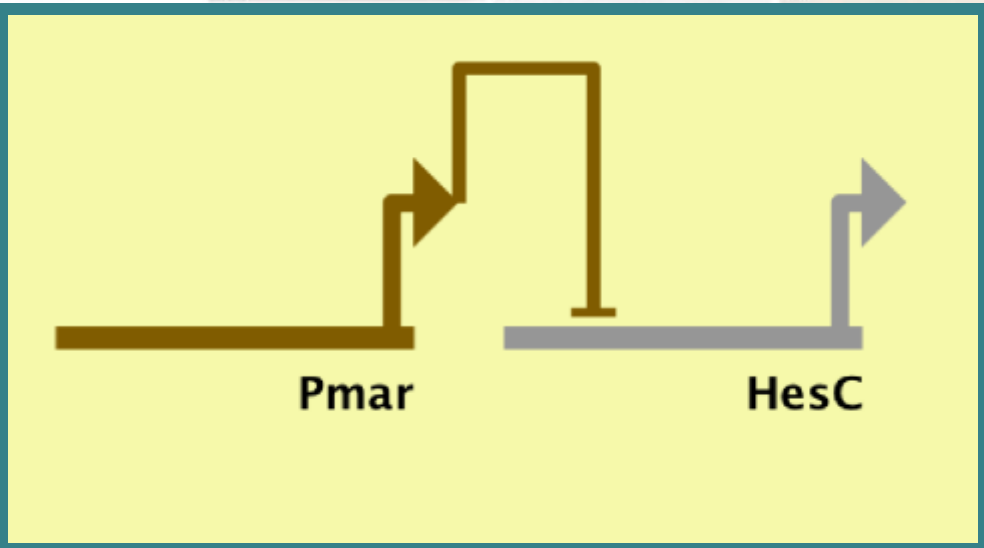




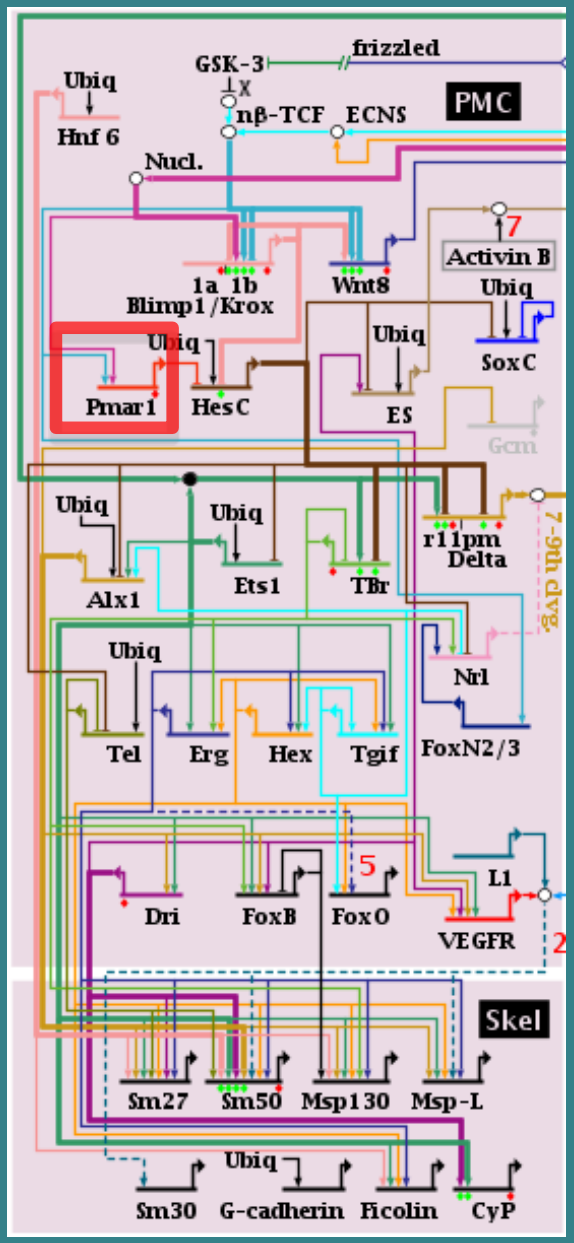
This model is frequently revised. It is based on the latest laboratory data, some of which is not yet published.

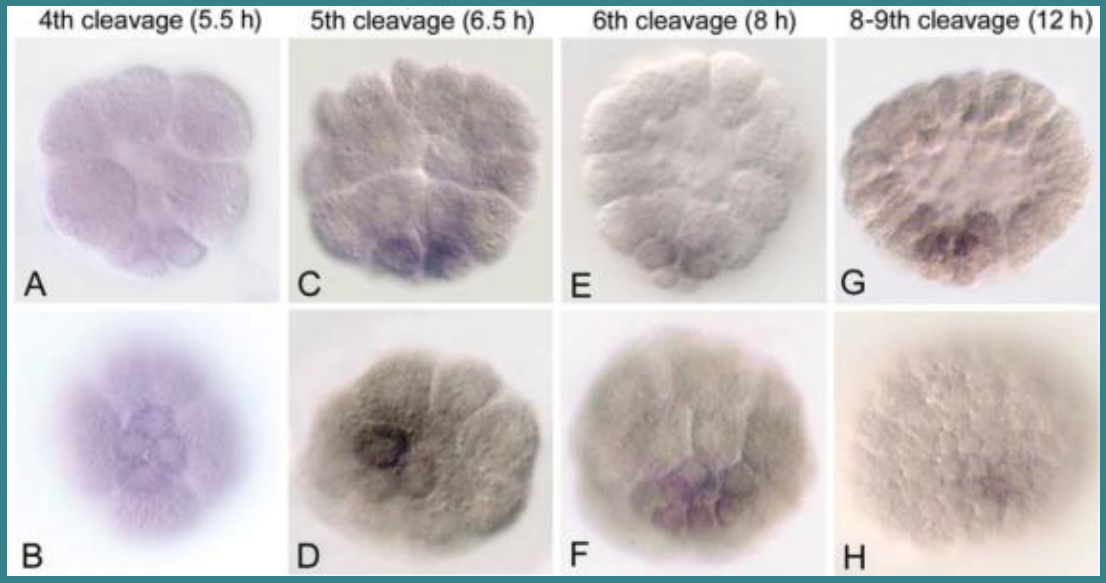


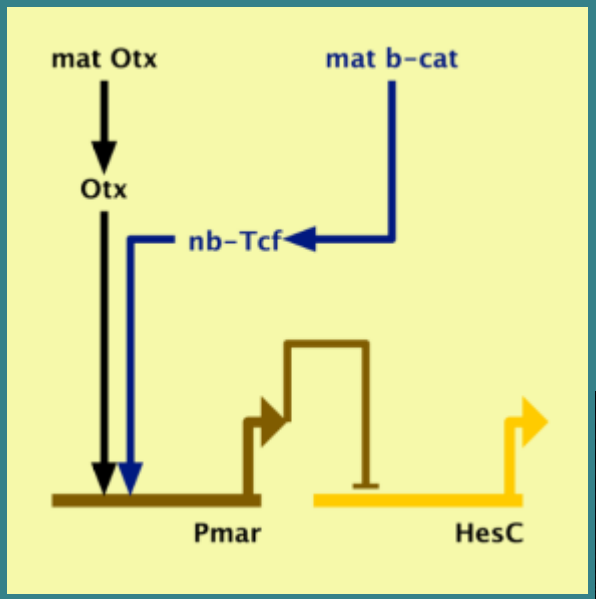


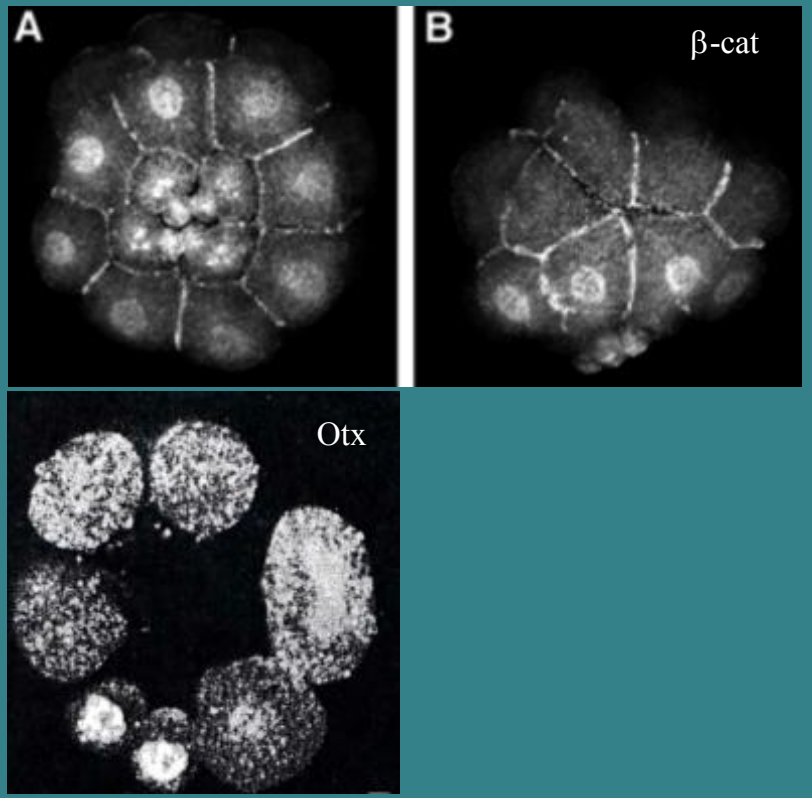


Revilla-i-Domingo *et al.*, 2007; Oliveri *et al.*, 2002, 2003

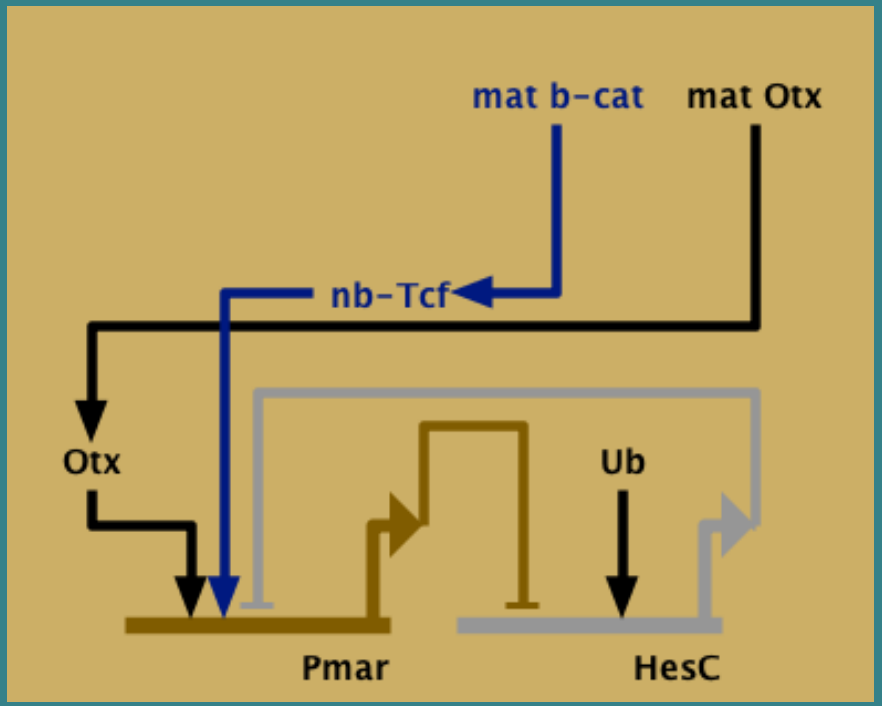


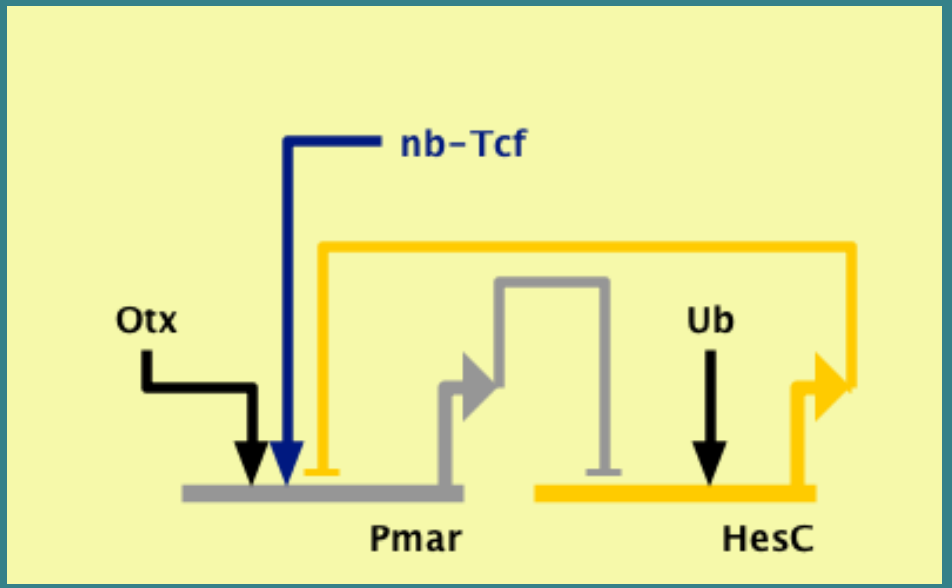




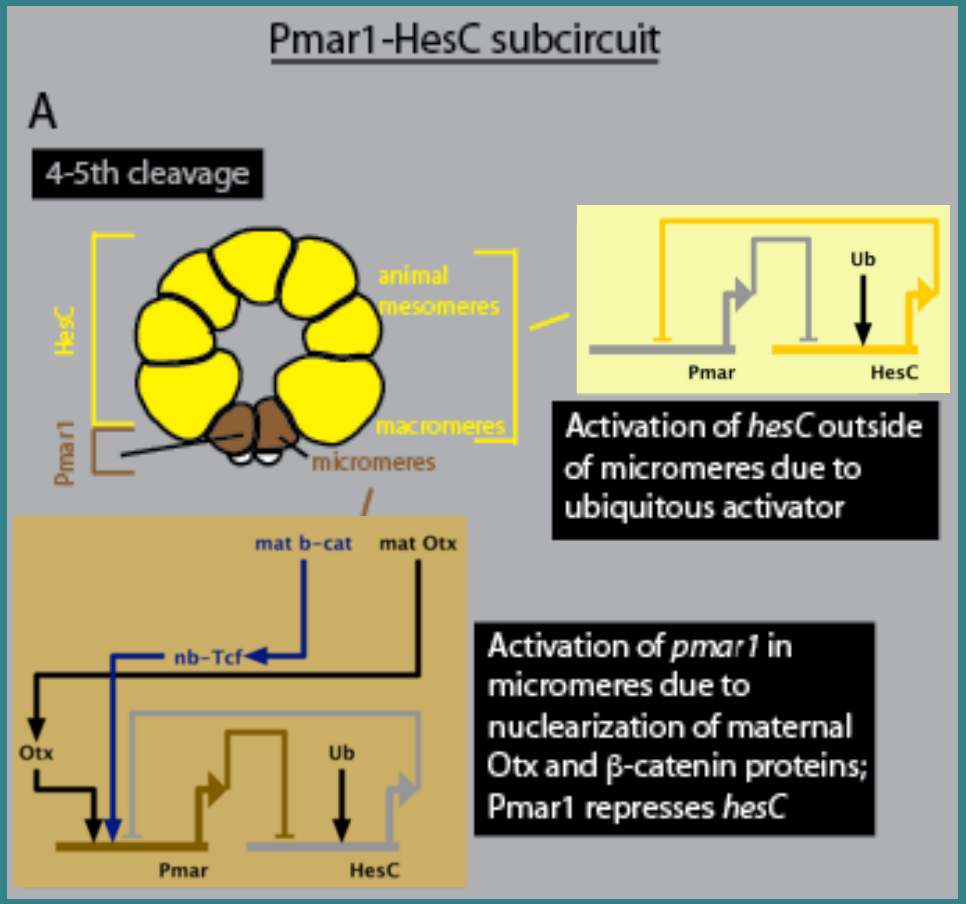


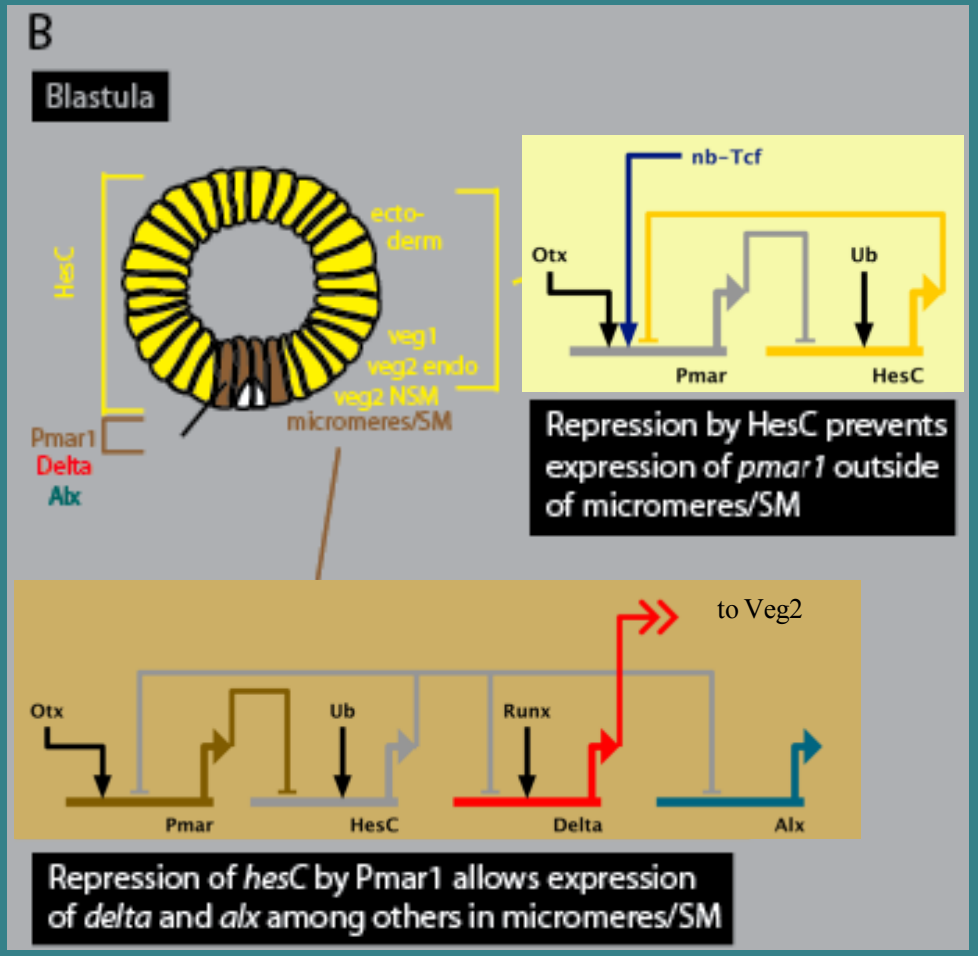
Revilla-i-Domingo *et al.*, 2007; Oliveri *et al.*, 2002, 2003; Logan *et al.*, 1998; Chuang *et al.*, 1996

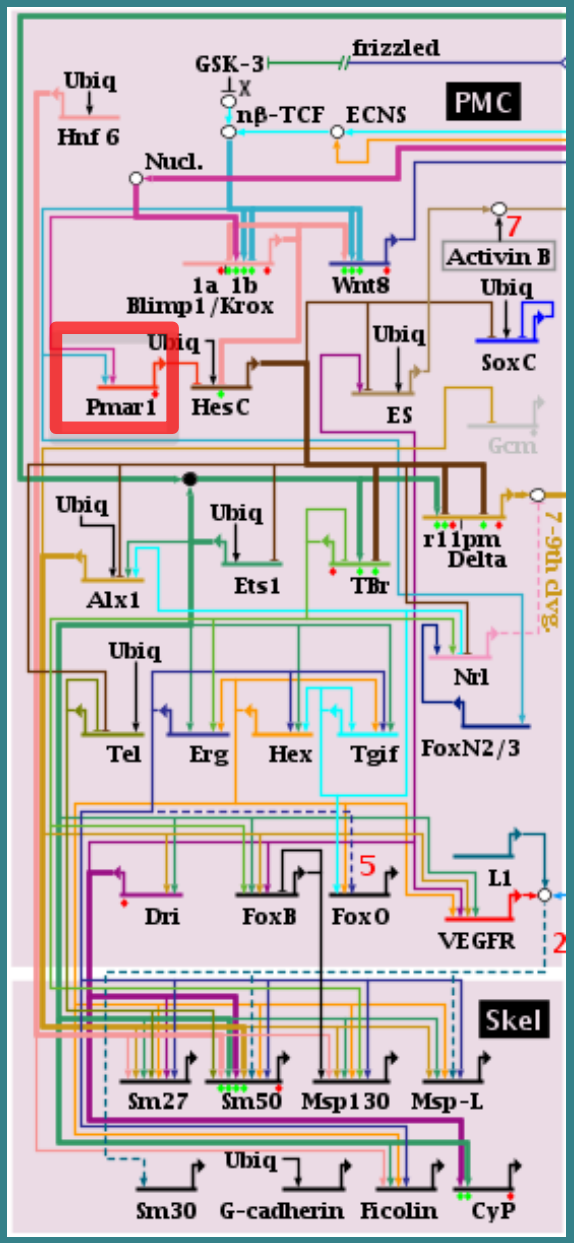


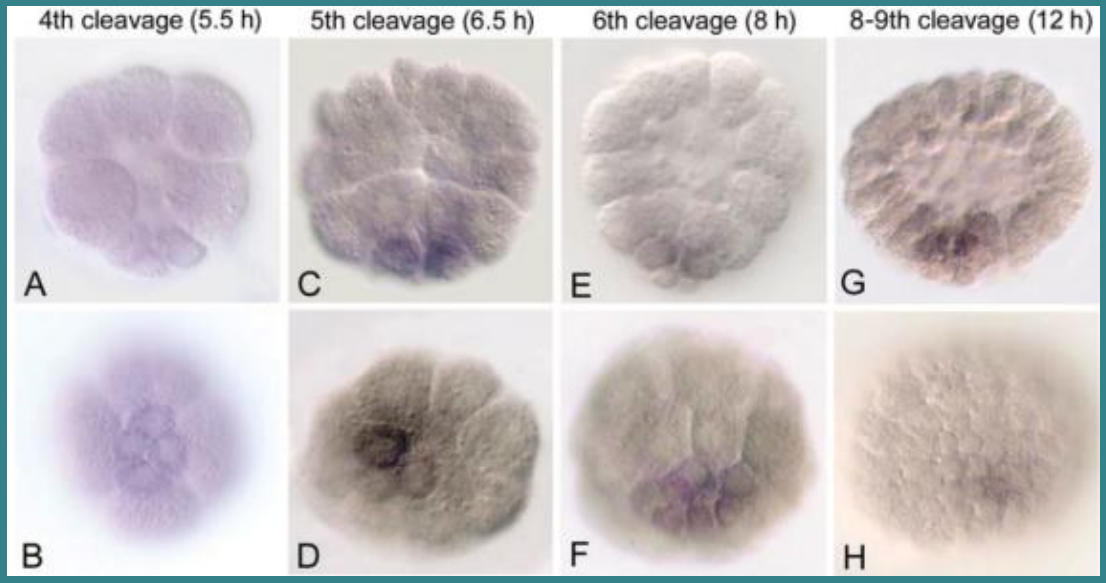


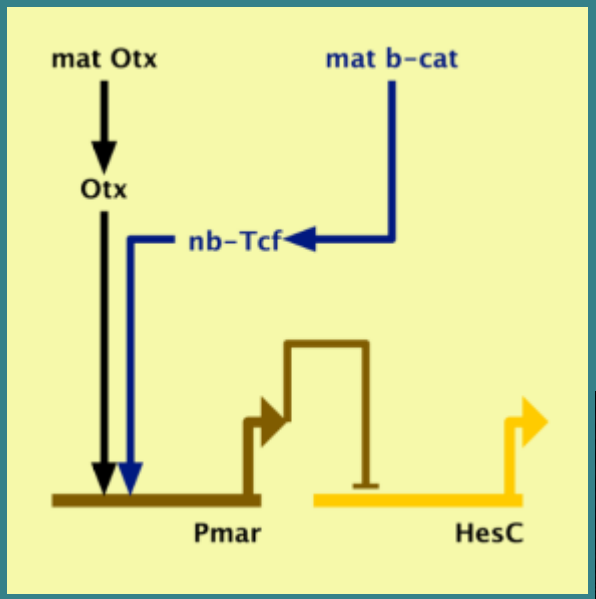
Reciprocal repression

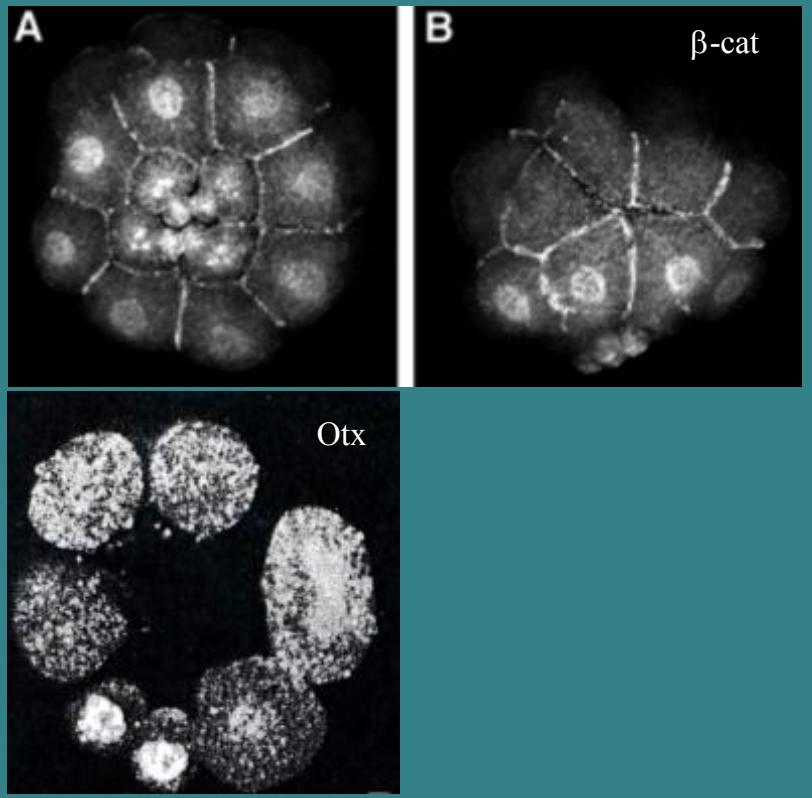




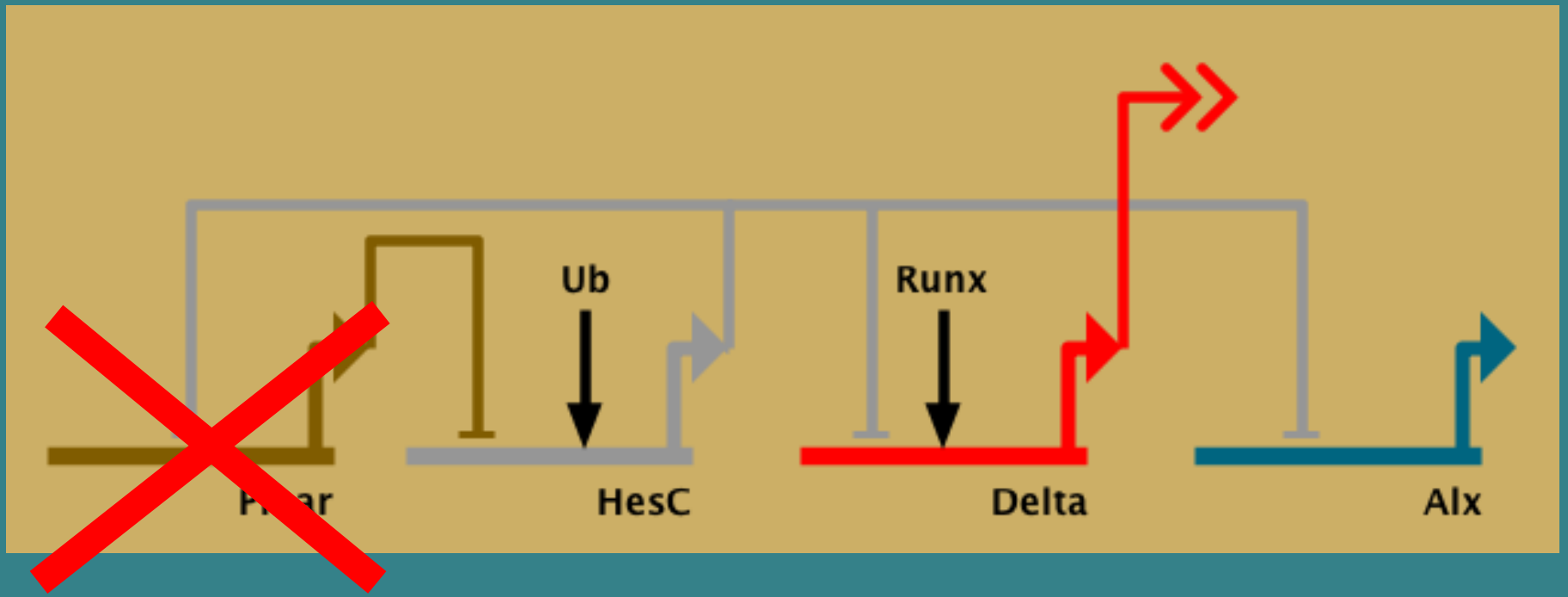


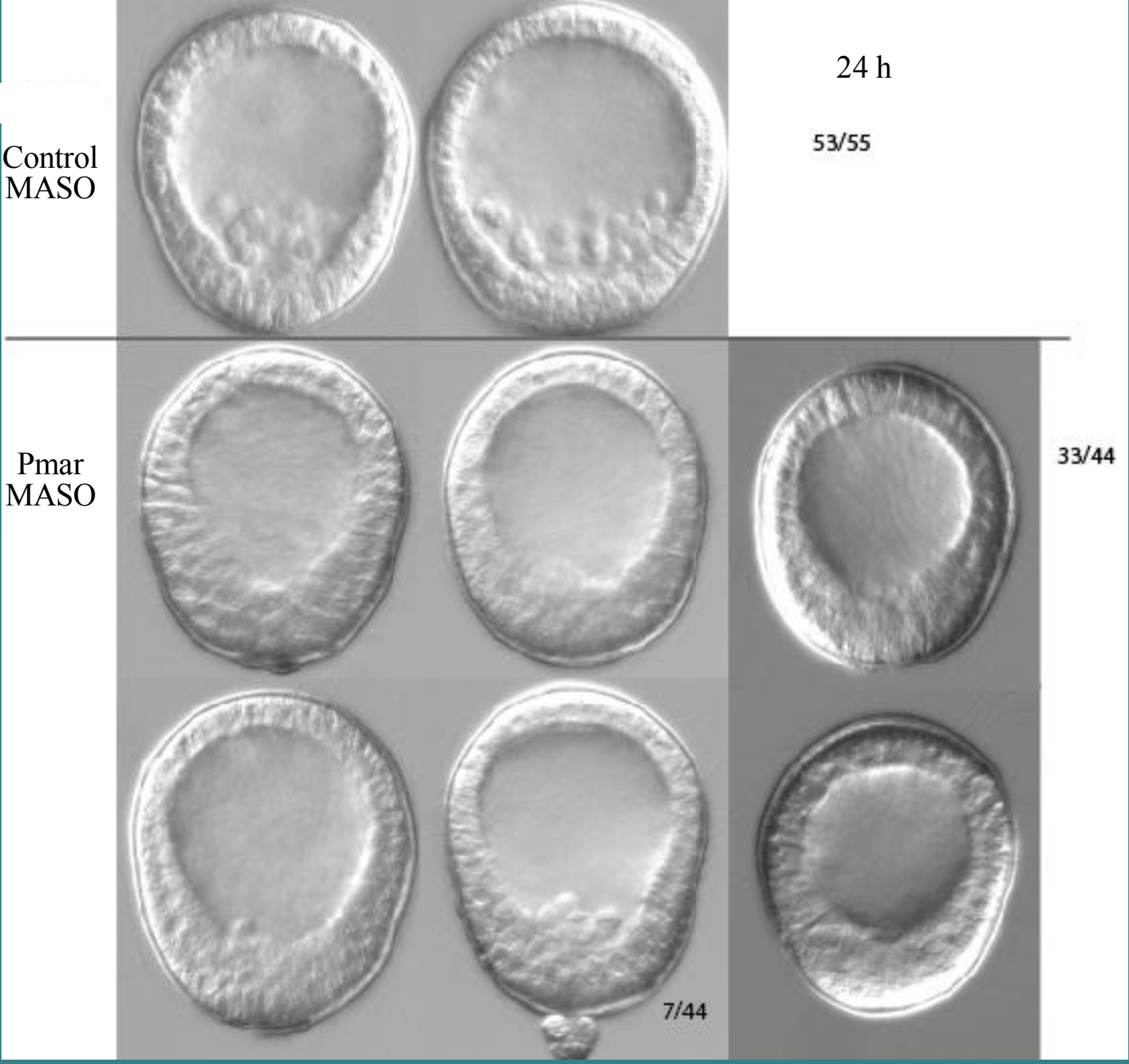






Revilla-i-Domingo *et al.*, 2007; Oliveri *et al.*, 2002, 2003; Logan *et al.*, 1998; Chuang *et al.*, 1996





Pmar MASO: shuts down skeletogenic mesoderm specification program

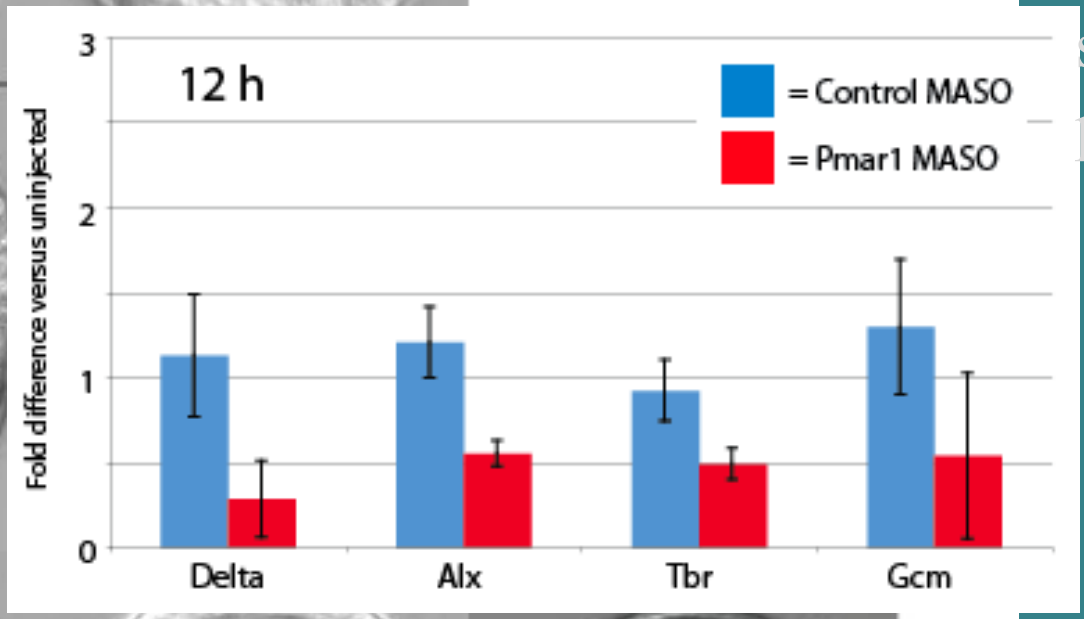
Control MASO

Pmar MASO

24 h

53/55

7/44

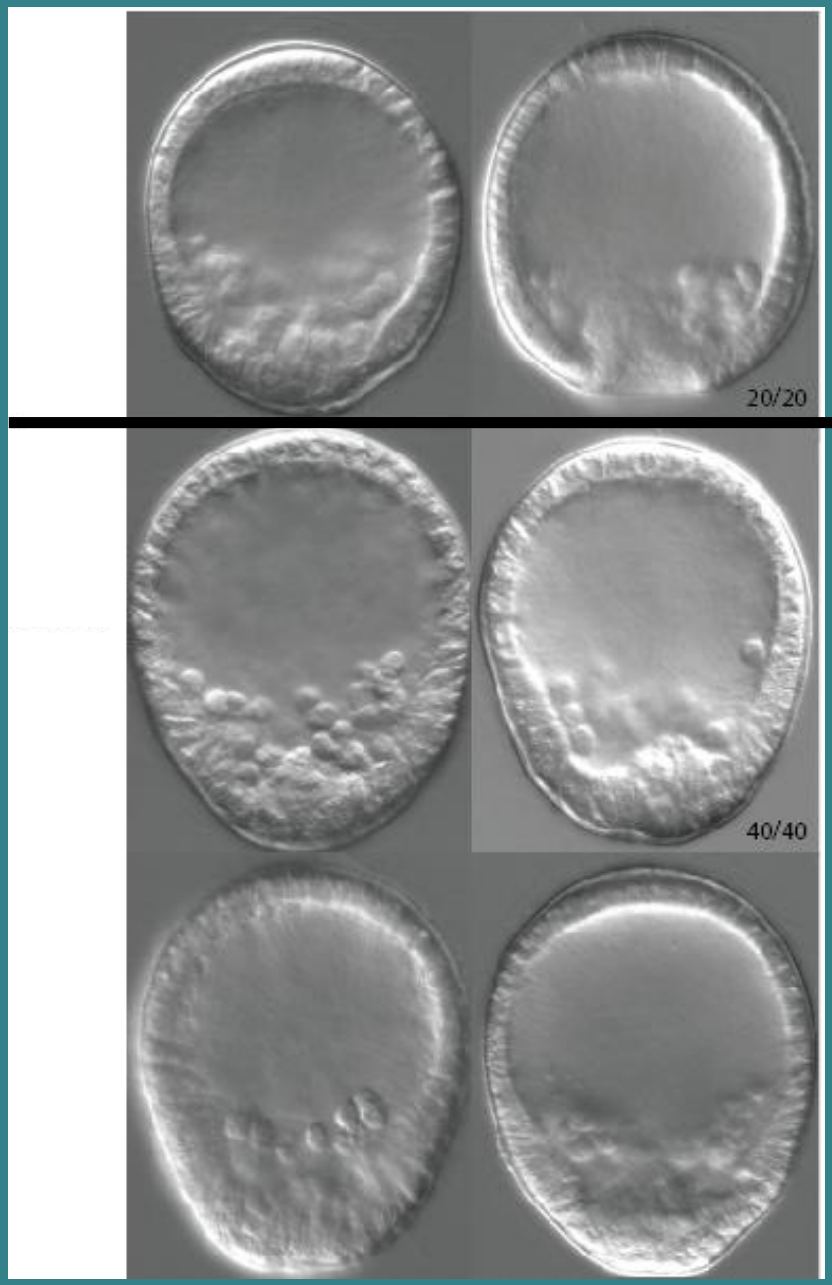


Pmar MASO: shuts down skeletogenic mesoderm specification program

However: embryos recover from Pmar MASO by 30h!

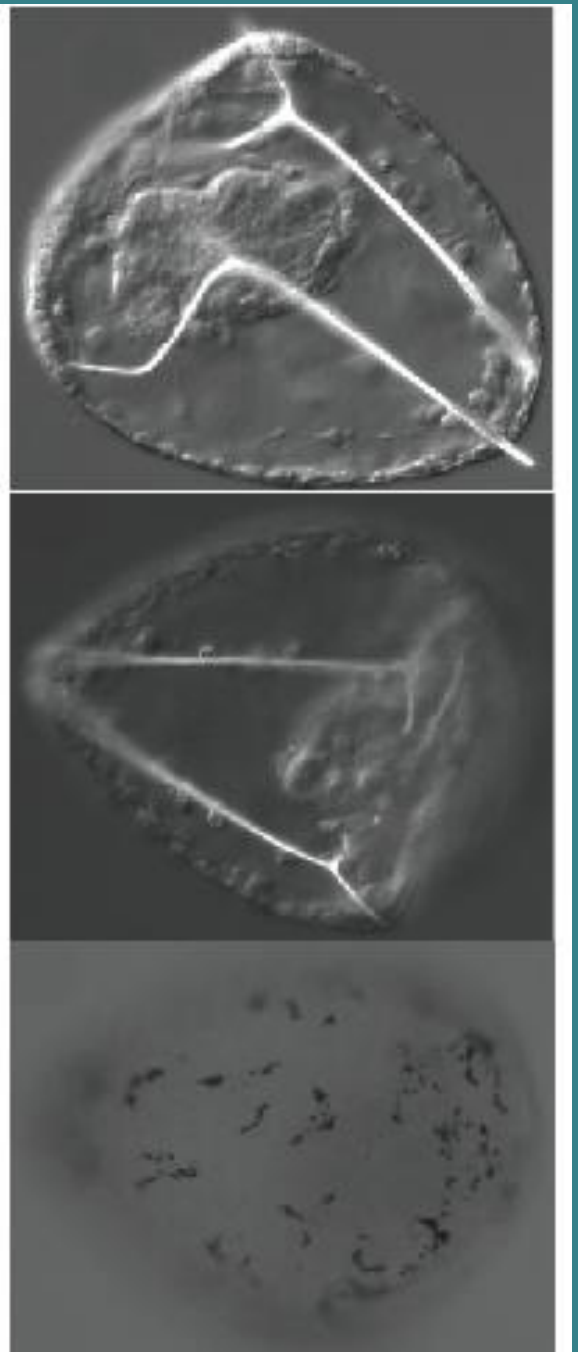
30 h

72 h



Control MASO

Pmar MASO

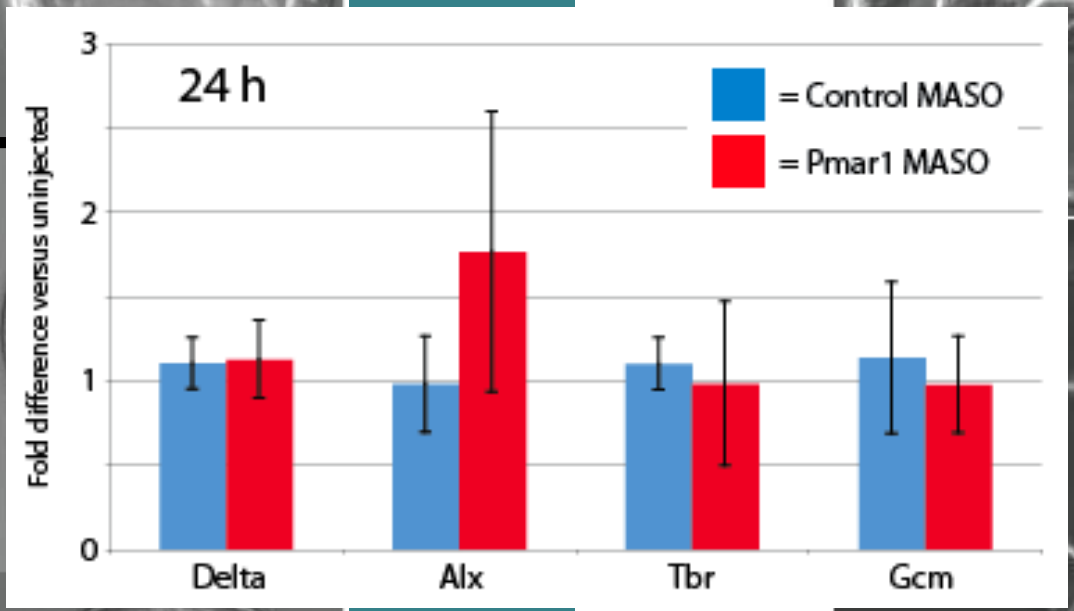


However: embryos recover from Pmar MASO by 30h!

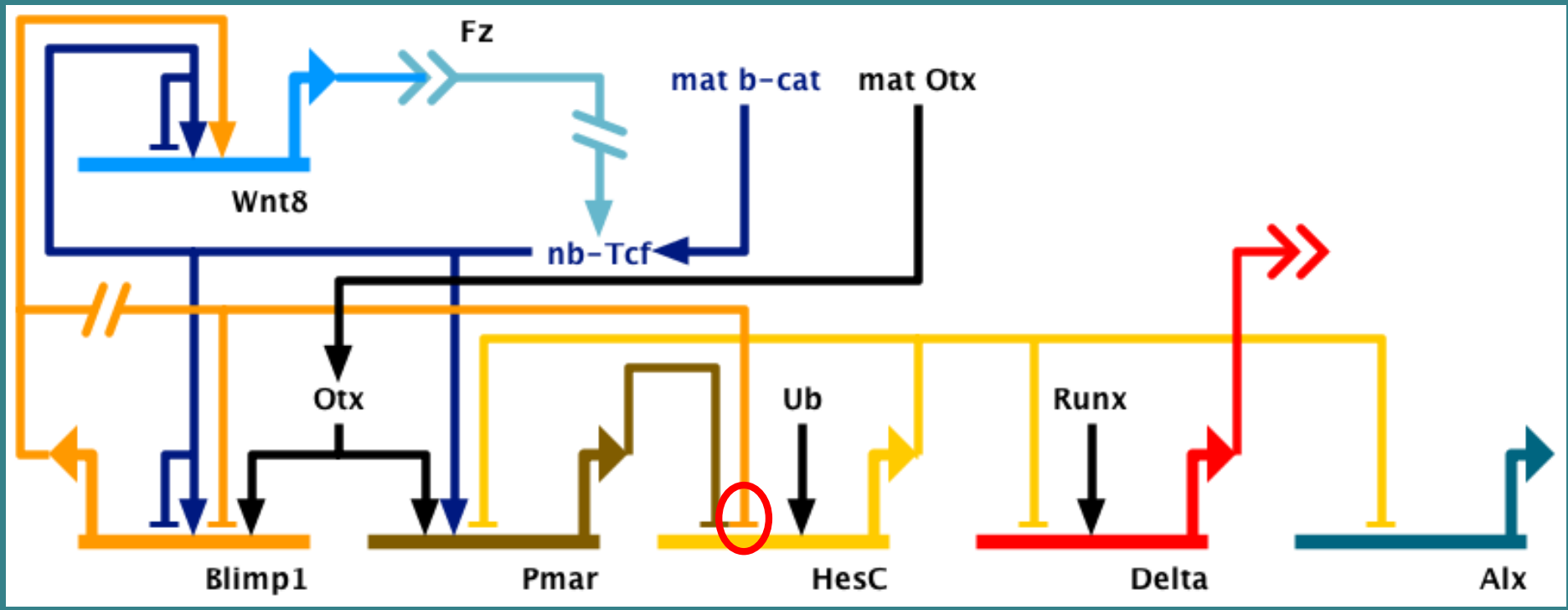
30 h

72 h

Control
MASO

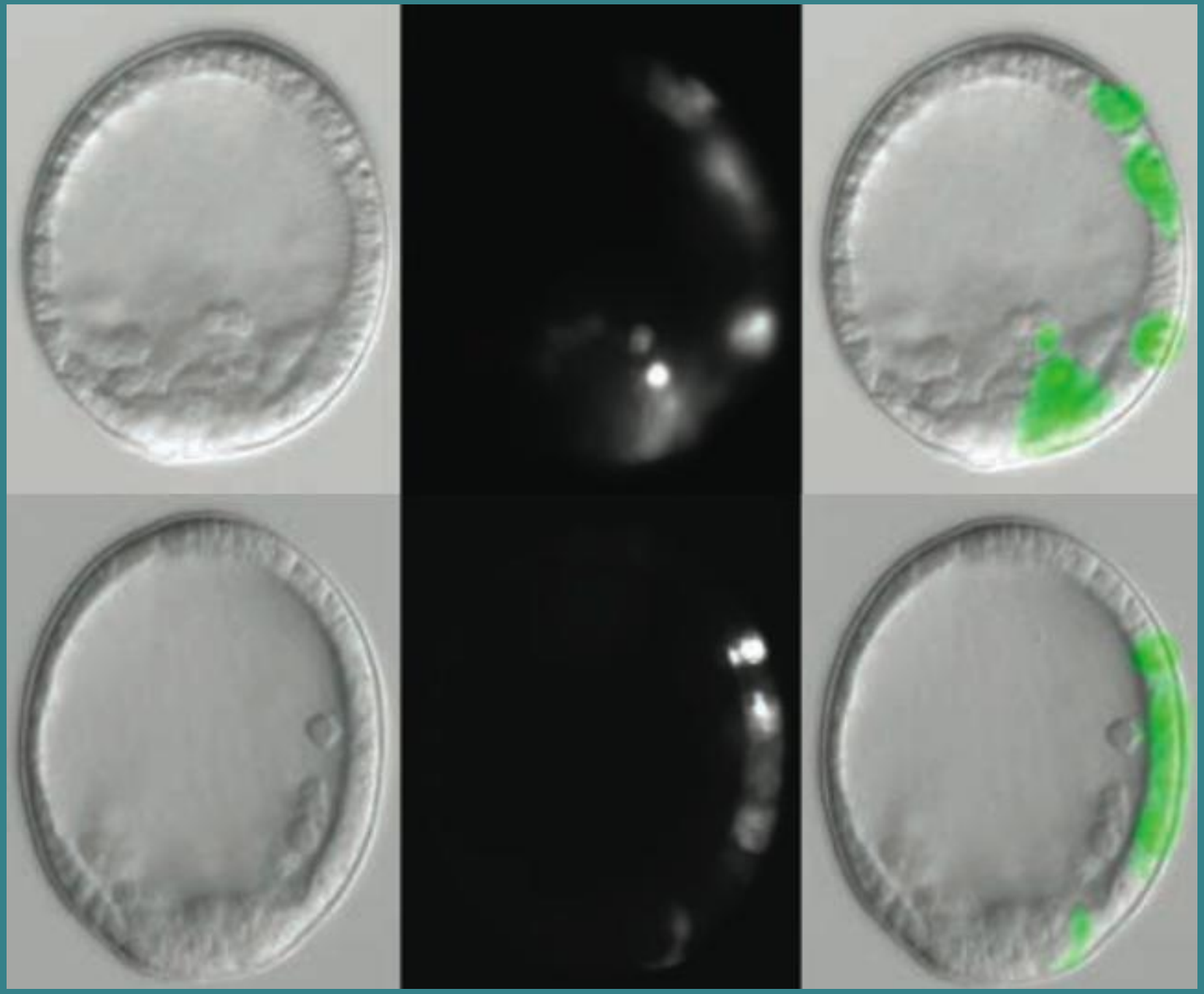


Does regulative recovery depend on repression of *hesC* by Blimp?



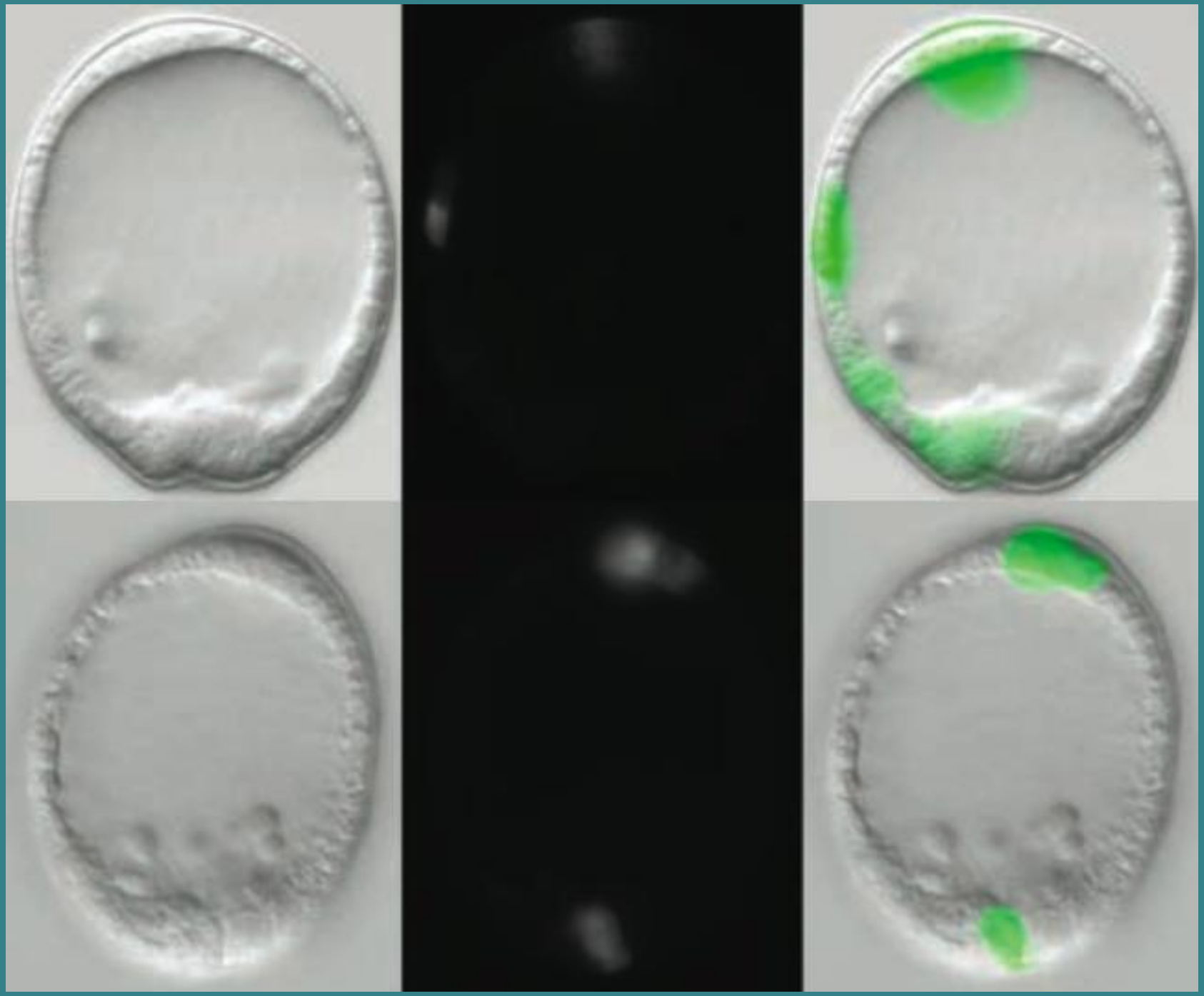
Control

ΔBlimp
HesC



HesC
ΔBlimp

ΔBlimp
HesC



Two systems:
same maternal/early inputs

The older one is able to compensate for disruption of the newer one