Network Based Models in Bioinformatics and Biocomputing

Instructors: Dr. Jeff Solka and Dr. Jennifer Weller

Lecture 3

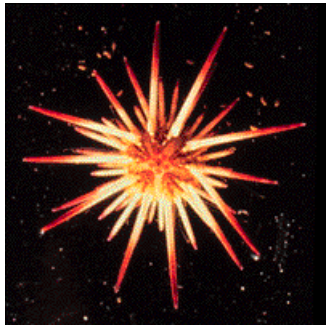
Assignments

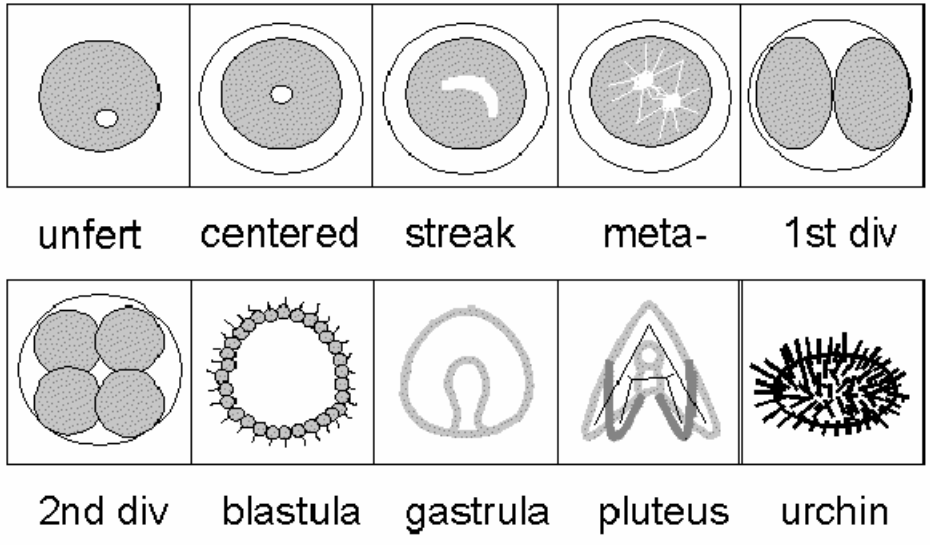
- Read Chapter 2 in The Regulatory Genome
- Homework 2 is posted.

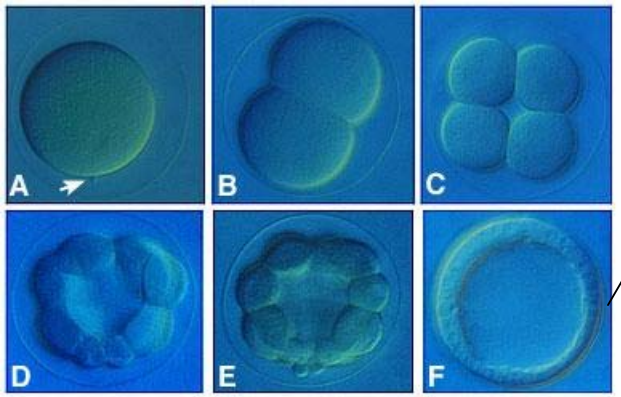
Agenda

- Chapter 2 from Davidson
- We will dissect the way in which CRE architecture allows specific distinct functions to occur:
  - A single set of components permits a wide range of activities to occur.
  - Both up- and down-regulation are mediated
  - The combinatorial use of effectors allows gating (or switching) of signals from the signal transduction pathways
  - Similar mechanisms are used for all CREs
    - The diversity of outcomes depends on the local concentration of effector proteins
  - The final output is indirect: the biochemical alteration that occur downstream of the switch are the conditional outputs of the state
    - The CRE processes the input but does not perform any of the actual functions (adminstrators)

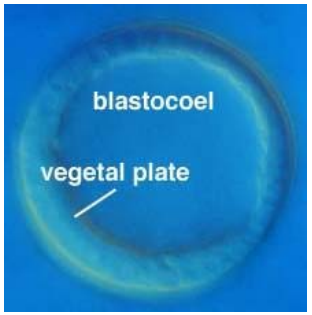
Urchin Development

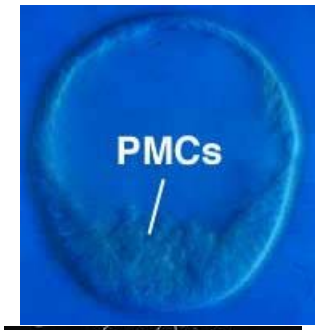


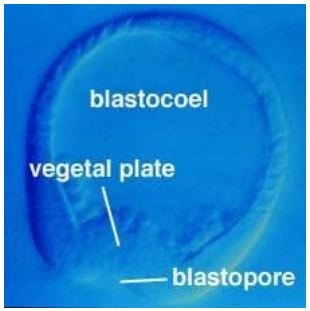


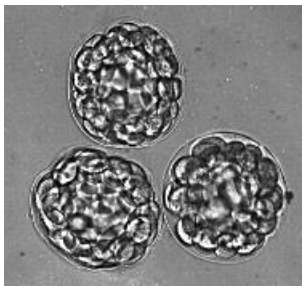


Cleavage stage

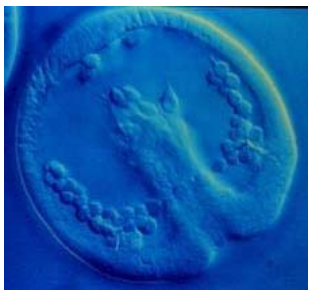








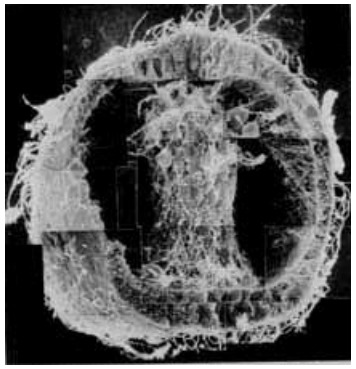
Blastula stage

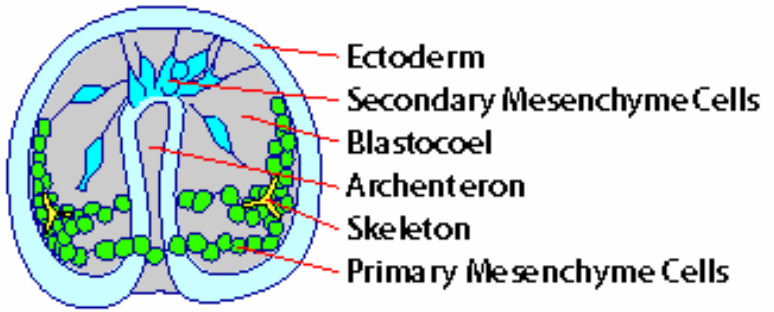


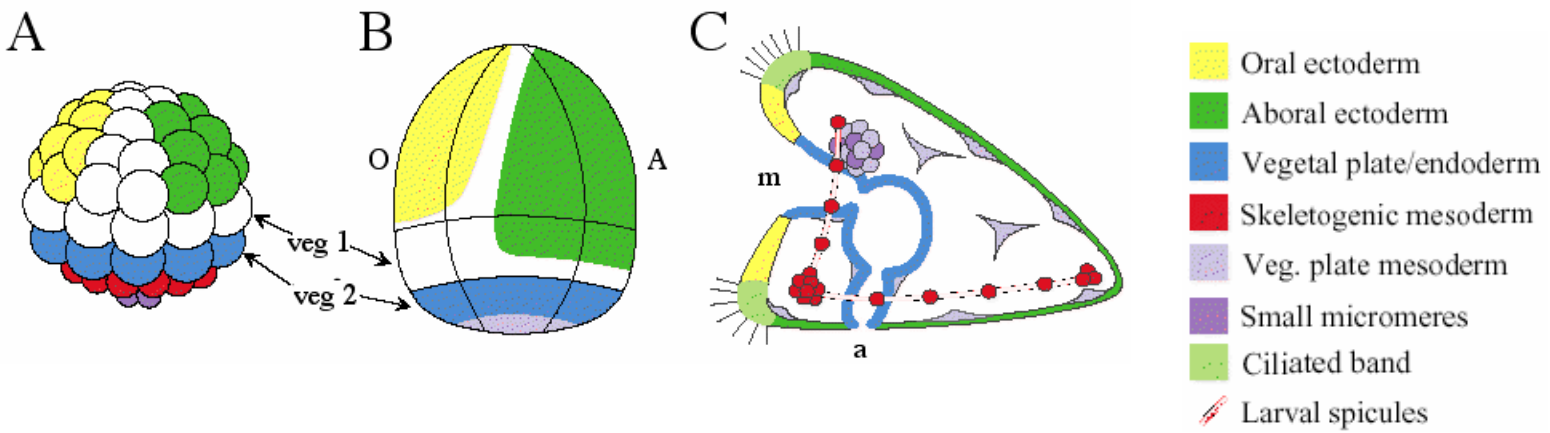
Gastrula stage

http://www.bio.davidson.edu/courses/GENOMICS/method/UrchDev.htm I

Urchin Development, cont







Images: ibid

Identifying and validating CREs

- Required for all methods is a DNA molecule
  - Sequence it
  - Bind proteins to it
- If you sequence it you can look for recurring patterns
  - Based on already known patterns
  - Based on comparative genomics, searching for conserved sequences
- If you use the physical polymer and see if proteins bind to it with high affinity
  - You can identify the protein
  - You can perform a footprinting assay to see exactly what part of the sequence the protein is protecting
  - You can then modify the protected sequence and see if this has an effect on the specificity or stability of the protein binding

Dissecting CRE function

- CREs are non-random clusters of target sites, comprising several hundred basepairs in actual length, distributed over a much larger, but finite length of genomic DNA.
  - Computationally it is possible to pose a pattern question, locate all instances of that pattern that meet your boundary conditions
    - Does the downstream gene respond in the way you would predict for the elements you have found?
      - In Drosophila, for a number of early developments genes specifying spatial expression along the anterior/posterior axis the answer was ‘yes’ for about 50% of the modules identified.
      - A second element the specifies lateral stripes of expression was also identified, and genes having this cluster of CREs showed the predicted spatial expression.

Comparative Genomics and CREs

- One of the characteristics of a CRE mentioned last time is the conservation of sequence, to a level equivalent to that of protein-coding sequences.
  - If you have genomic sequence from species at the necessary evolutionary distance it is possible to identify likely CREs even in the absence of gene expression patterns.
    - One marker of the presence of these elements is the absence of very large insertions between shorter conserved sequences
      - Small differences in spacing don't usually matter, and those that are far apart can be even farther apart, but there is an intermediate distance that cannot be handled by looping and that may not allow protein complexes to form.

Showing CRE function

- The CRE directs the time and place of expression of a downstream protein, but the actual identity of that protein is irrelevant.
  - Fuse the stretch of DNA that is proposed to be the CRE to a gene whose protein is readily detectable
  - Insert the construct into a cell at the desired stage
  - Follow the subsequent events in the cell to see when and where the new protein product is expressed
- Follow-up questions include whether the components are essential and unique
    - For this a knock-out or knock-down experiment must be performed

Handling Redundancy

- An isolated system may produce the predicted pattern so that the expectation is you will be able to produce a particular phenotype in a knock-out experiment
  - If this does not occur then this implies the system is redundant – this violates the uniqueness assumption
    - However, to date whenever such a situation has been completely investigated the modules have varied by at least one component, which was simply not active under the experimental conditions tested
    - That is, for each spatially separated but contemporaneous expression it has so far been shown that a different CRE was used to drive the expression
    - The more ubiquitous the gene product, the more complex the set of CREs that direct it.

Making a reporter Construct

- This methodology will be outlined for *C. elegans*, but the general methods are similar for all model systems. This information has been condensed from a very complete presentation in the Worm Book, from which the following pictures have been used.
- The three general types of reporter gene constructs are:
  - transcriptional reporters
  - translational reporters
  - “smg-1-based" transcriptional reporters

Transcriptional Reporters

- Transcriptional reporters incorporate a promoter fragment (piece of DNA) from a gene of interest (GOI), but use it to drive the expression of a protein such as the Green Fluorescent Protein (GFP) or beta-galactosidase (lacZ).
  - GFP can be visualized in live animals while lacZ requires fixing and staining.
  - Usually several thousand bases from the 5' end of the GOI are tested, as most likely to contain CREs, but of course there are exceptions.

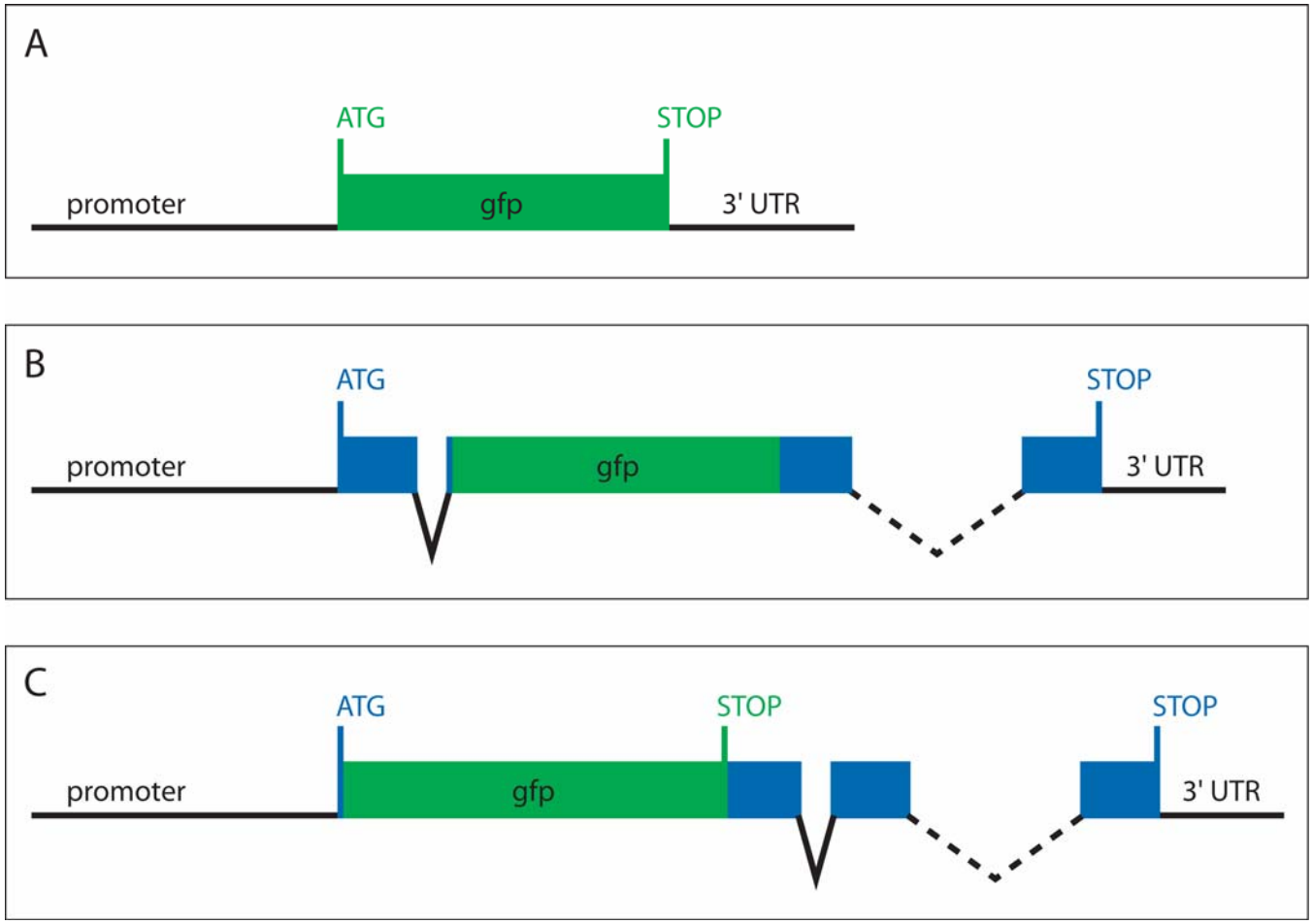
Translational Reporters

- Translational reporters use an in-frame gene fusions between GFP and the coding sequence of the GOI.
  - In-frame means you retain the correct 3-codon per amino acid window for both the GOI and the GFP.
  - Ideally, you will incorporate the entire genomic sequence of the GOI with the elements of the 5′ upstream region, exons, introns, and 3′ UTR.
    - This lets you retain CRE information inserted in the middle of the gene
  - The choice about where to insert the GFP is somewhat arbitrary, although modeling is performed to try to avoid disrupting the protein function or topology of the GOI.
    - This may allow the gene product to be translocated normally to its proper location (giving spatial as well as temporal data), and carry out its downstream function.
    - Obviously, disruption of the GOI is a possibility, and too much disruption may cause the chimeric protein to be degraded very rapidly.
- If mutants are available, translational reporters can be used in rescue experiments, which provides support for the proposed functional importance of the GOI.

Smg-1-based reporters

- *Smg-1*-based reporters are another type of transcriptional reporter, but the aim is to include all the CRE components for the GOI.
  - GFP, including its stop codon, is inserted *between* the promoter and the first exon of the GOI.
  - The transcription of GFP is susceptible to the complete set of CREs found upstream, intronically or downstream of the GOI, to the limits of the size of the insert.
    - GFP is made, the GOI is not
  - Technically it has been noted that the GFP stop codon results in this mRNA being targeted for a particular rapid turnover process, so these experiments are done in a particular genetic background in which that process has been elliminated (the smg-1 gene).

Cartoon schematic of the types of reporter constructs



(A) Transcriptional reporter, (B) translational reporter, (C) "*smg-1*-based" reporter. Exons are in blue, GFP in green.

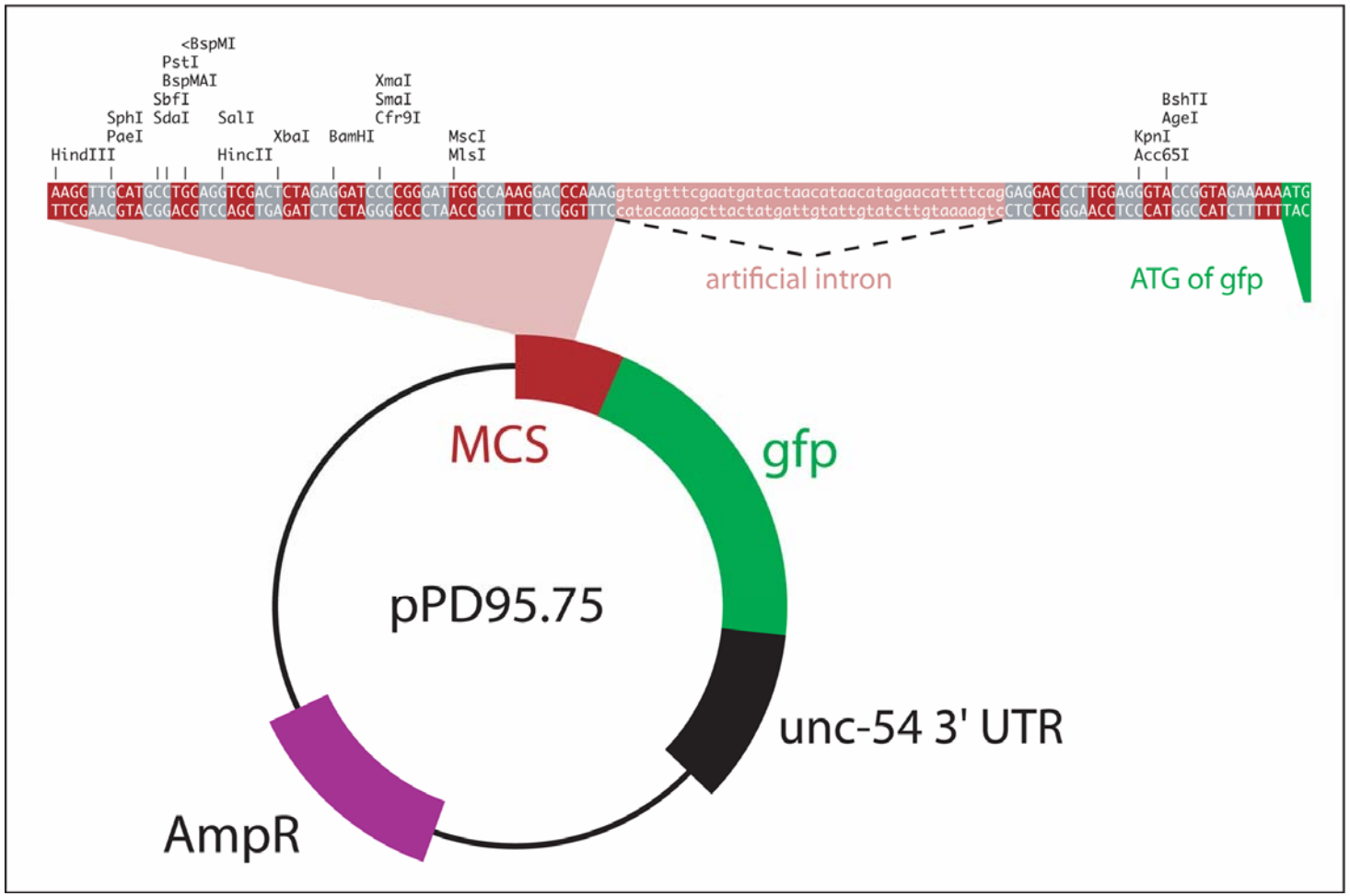
Elements incorporated in Reporter Constructs

- Size of the insert: the C. *elegans* genome is very compact so usually the CRE is encompassed within ~2kbp of the TATA box.
- There are many example of CREs in introns and in 3' UTRs.
  - Longer that usual introns have become a signal that such elements are likely to be present.
  - More rarely they are 3' downstream or very far 5' upstream.
- The initial experiments generally begin with a 1-2 kbp length of DNA that is immediately 5' to the GOI.
  - A computational assessment of known elements and/or conserved elements is performed to guide the selection of this sequence.

Tools to study transcriptional regulation

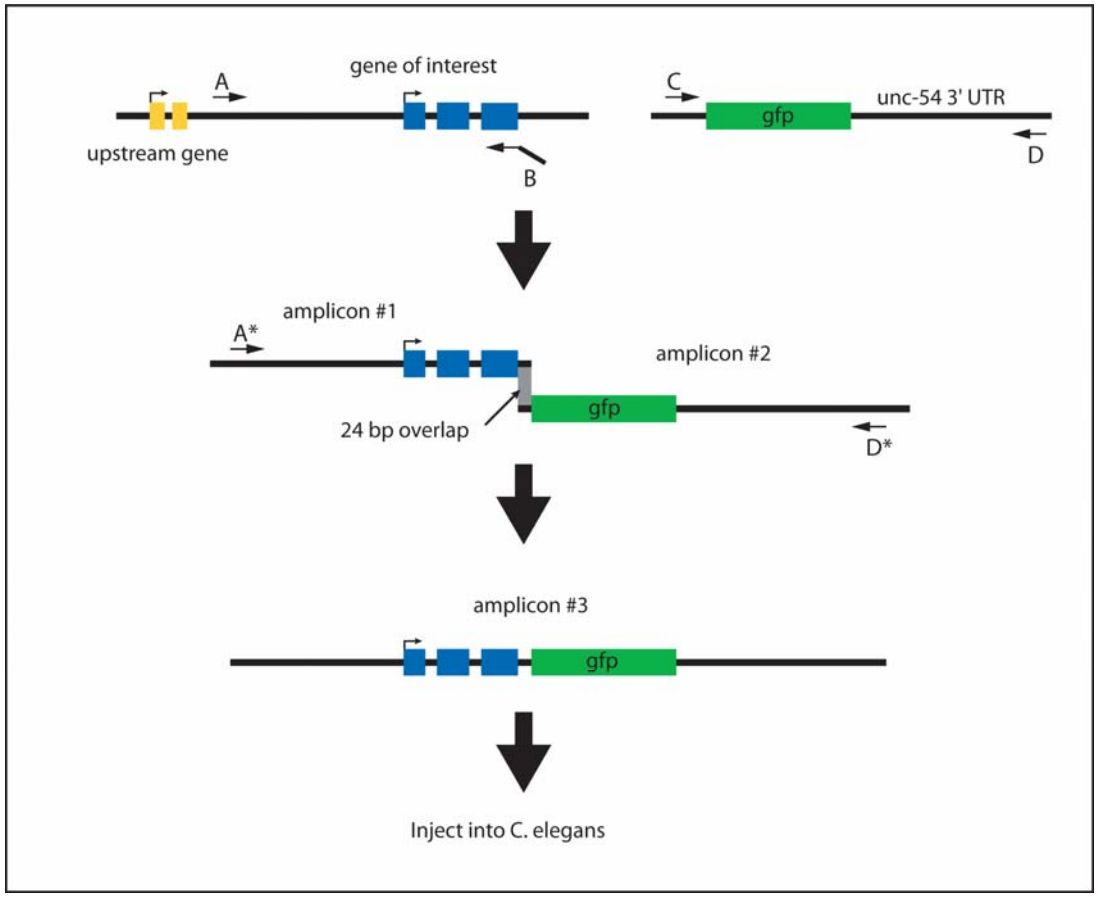
- A GFP ‘coding cassette’ is inserted in different locations within a large genomic clone (tens of kilobases) surrounding the GOI, to generate transcriptional and/or translational fusions.
- The region may be purified by restriction digestion and gel purification of a sub-fragment from a genomic DNA clone, but it is usually obtained by using the polymerase chain reaction (PCR) to make targeted copies.
- The fragment is then inserted into a plasmid (or related cloning vector) that allows modules to be recombined in highly specific ways.

Vector backbone



pPD95.75. The multiple cloning site (MCS) of pPD95.75 spans from HindIII to MlsI. The reading frame upstream of GFP is indicated by alternating red and gray boxes. The artificial intron lies between the MCS and the GFP coding region (the ATG of GFP is indicated in green).

PCR fusions



- Shown: C-terminal translational PCR fusion.
- Primers A and B amplify the genomic region of the GOI).
  - Primer B adds a 24 bp overlap in frame to the GFP coding region.
  - Primers C and D amplify the reporter gene and 3' UTR
  - Primers A* and D* are used to fuse the two amplicons
    - gray box indicates 24 bp sequence overlap. without purification

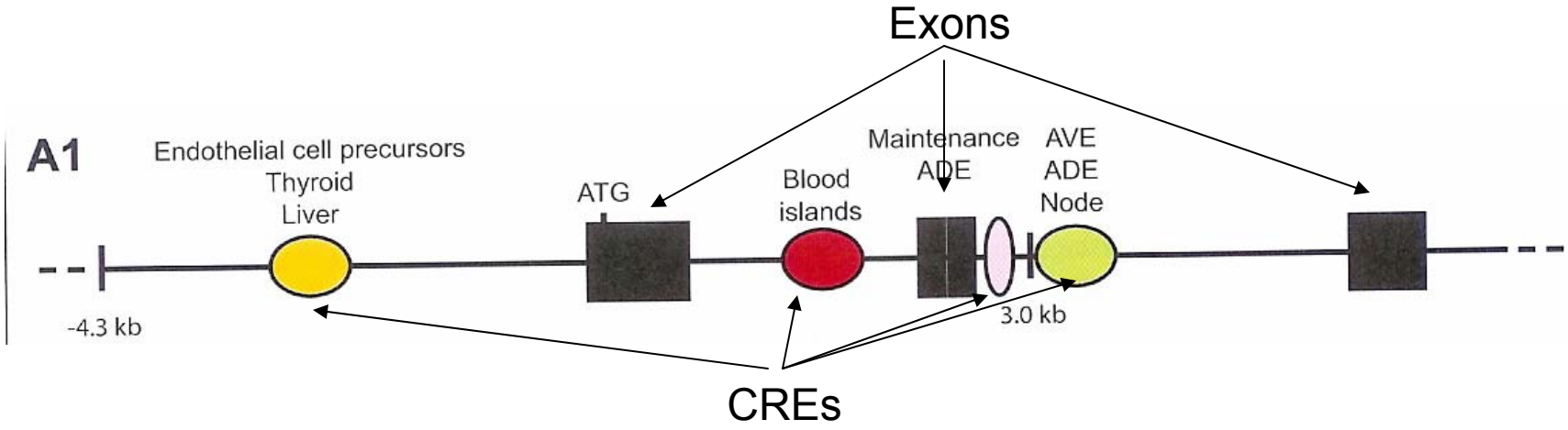
Promoter Analysis

- Once a pattern of expression is obtained, the goal is to determine the important regulatory elements.
  - Sequential deletions of putative promoter regions linked to a GFP reporter gene are easily made by a number of highly-controlled techniques.
  - Once the control elements have been localized to small genomic regions (several hundred base pairs or less), they can be mobilized to the upstream of "basal" promoters to assay for enhancer or represser activity.
- Reporter gene constructs are artificial and can easily misrepresent the pattern of gene expression.
  - Important positive- and negative-acting control elements that have inadvertently been excluded lead to a number of outcomes, including mosaicism, loss of expression, or ectopic expression.
  - A common outcome when CREs for fidelity have been lost is expression in the anterior and posterior intestinal cells or a small set of head neurons.
  - Small changes in promoter regions can dramatically alter expression patterns
    - It is critical to confirm reporter gene expression patterns with an independent technique such as *in situ* hybridization, antibody staining, or mutant phenotype.

New CRE combinations

- The constructs used for engineering expression vectors, have their own elements and thus may produce a new combination that has an activity that does not reflect an *in vivo* process.
  - The vector diagrammed above has a 3′ untranslated region (UTR) derived from the unc-54 gene (muscle myosin heavy chain). These sequences have been shown to not always be neutral when combined with promoters from other cell types.
  - Since there may be dramatic effects of co-transformation markers on expression levels, at least two should be used.
  - “basal” promoters may show bias for (or against) being effective with particular CRE modules.

Fig 1: the mouse hex gene



- In the early embryos this gene is expressed in two different locations – it is the earliest gene known to be essential for liver formation, B cell development, and is needed for the thyroid and forebrain. It was first isolated by sequence similarity to a human gene used in hematopoeisis.

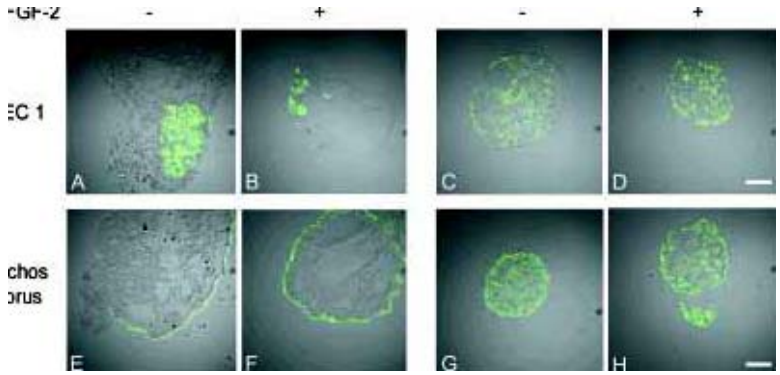
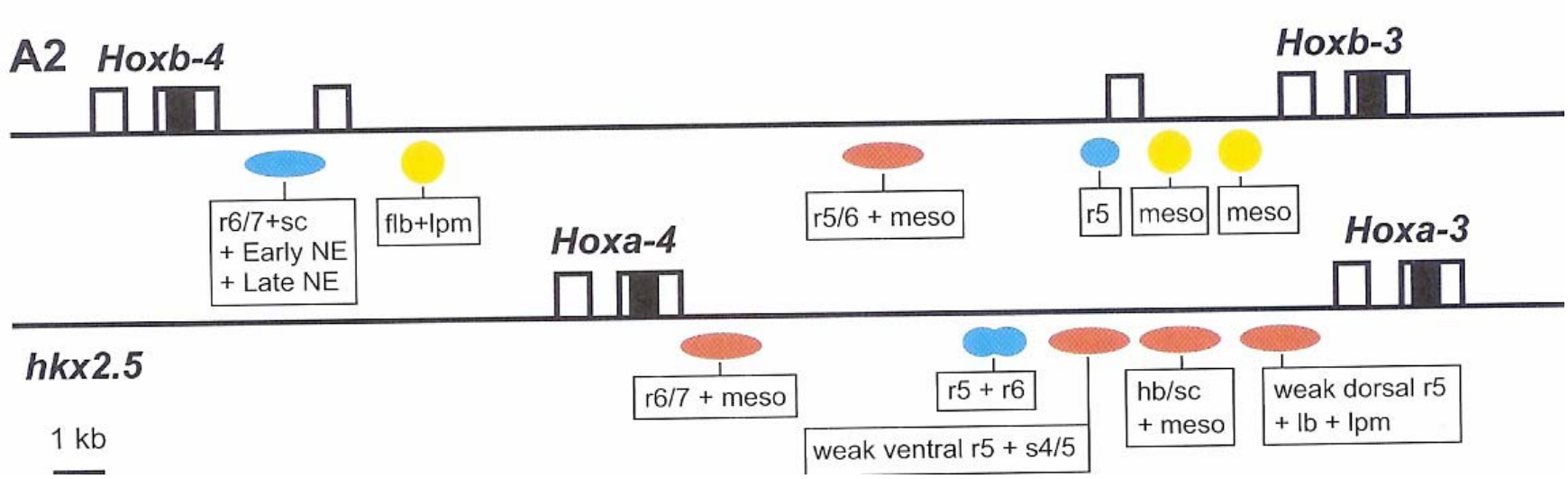


Fig 1-2: hox gene example



- The ovals are CREs and they are labeled according to the spatial expression they direct.
  - Letters indicate body parts -- r: rhombomere, s: somite, lpm: lateral plate mesoderm, lb: limb bud, flb: forelimb bud, hb: hindbrain, sc: spinal cord.
  - Colors indicate tissue expression: yellow=mesodermal, blue=neural, red = neural+mesodermal

Hox genes in brief

- The homeotic (genes that affect embryo development by specifying the character of a body segment) Hox proteins are sequence-specific transcription factors
  - Hox genes provide cells with specific positional identities on the anterior-posterior (A-P) axis.
  - The hox proteins contain an `homeobox' domain.
  - In Drosophila and other insects there are eight different Hox genes that are encoded in two gene complexes:
    - ANT-C and BX-C
  - In vertebrates there are 38 genes organized in four complexes.
  - A new nomenclature system would assign the four gene complexes the letters A to D.
    - Based on sequence similarities the genes can be sorted into 13 `paralog' groups.
    - The order of the paralogs along the chromosomes are conserved in the four complexes.
    - The gene name is obtained by concatenating the gene complex letter designation with the group number; examples: HOXA1, HOXB4, HOXC8, HOXD13, etc.
- Information has been restated from the Hox gene database overview, which can be found here: http://www.iephb.nw.ru/labs/lab38/spirov/hox_pro/hoxdb.html

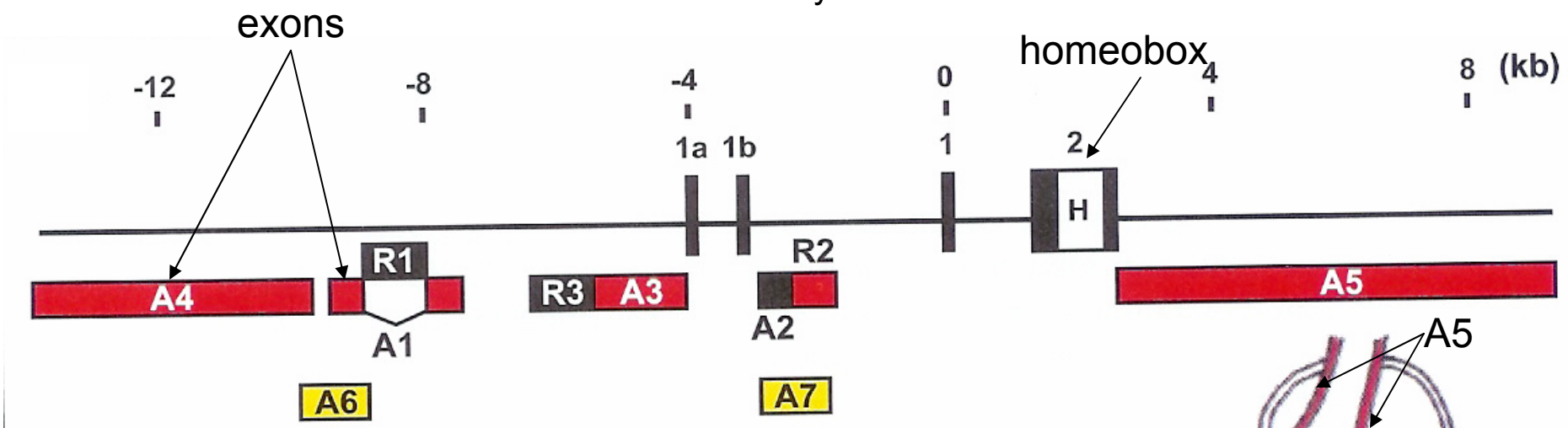
Homeobox domains

- Homeobox genes are defined by the presence of a characteristic 183 base pair DNA sequence (the homeobox)
  - It codes forr a relatively conserved 61amino acid section of protein (the homeodomain).
  - In mammals the four Homeobox containing gene clusters (HOX-clusters) are a highly conserved group of genes evolutionary related to the Drosophila Antennapedia- and Bithorax-complexes.
- Gene expression patterns from the HOX clusters show that domains of gene expression are spatially restricted in different embryonic sites and axes.
- There is a linear correlation between the position of a gene in a HOX cluster and its relative Antero-Posterior or axial domain of expression in many embryonic tissues.
  - This collinearity is conserved in arthropods and vertebrates
    - the regulatory mechanism for controlling the spatially restricted domains of HOX expression is an important feature in maintaining the organization of these gene clusters.

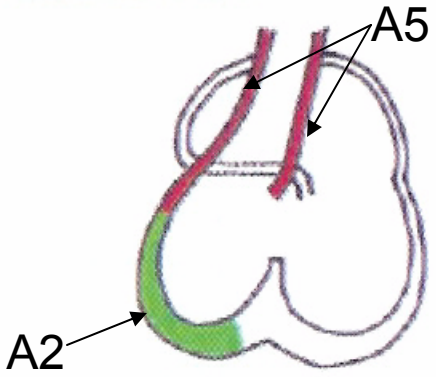
NK2 tf

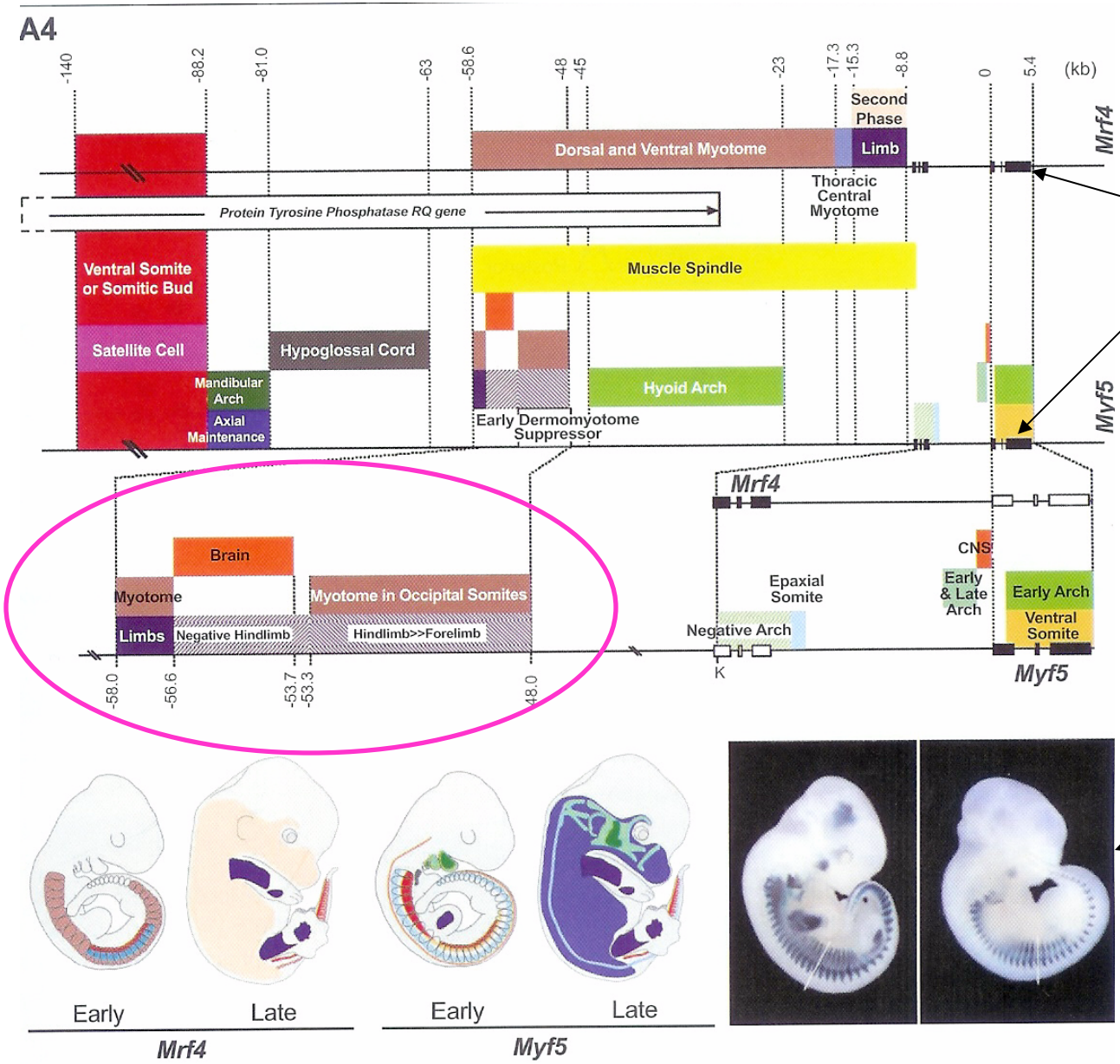
A Drosophila homeobox-containing gene called 'tinman' is expressed in the developing dorsal vessel and in the equivalent of the vertebrate heart. Mutations in tinman result in loss of heart formation in the embryo, suggesting that tinman is essential for Drosophila heart formation. Aabundant expression of Csx (mouse homolog), is observed only in the heart from the time of cardiac differentiation.

- A=enhancer
- R=repressor
- red=dev. heart
- yellow= other tissue



A1: cardiac crescent, entire linear heart tube, outflow tract, right ventricle
A2: anterior cardiac crescent, anterior linear heart tube, outflow tract, right ventricle
A3: outflow tract, right ventricle
A4: cardiac crescent, RV medial wall and inner trabeculae, LV medial wall and inner trabeculae, interventricular septum
A5: RV lateral wall
A6: thyroid
A7: thyroid, spleen, pharynx, stomach





exons

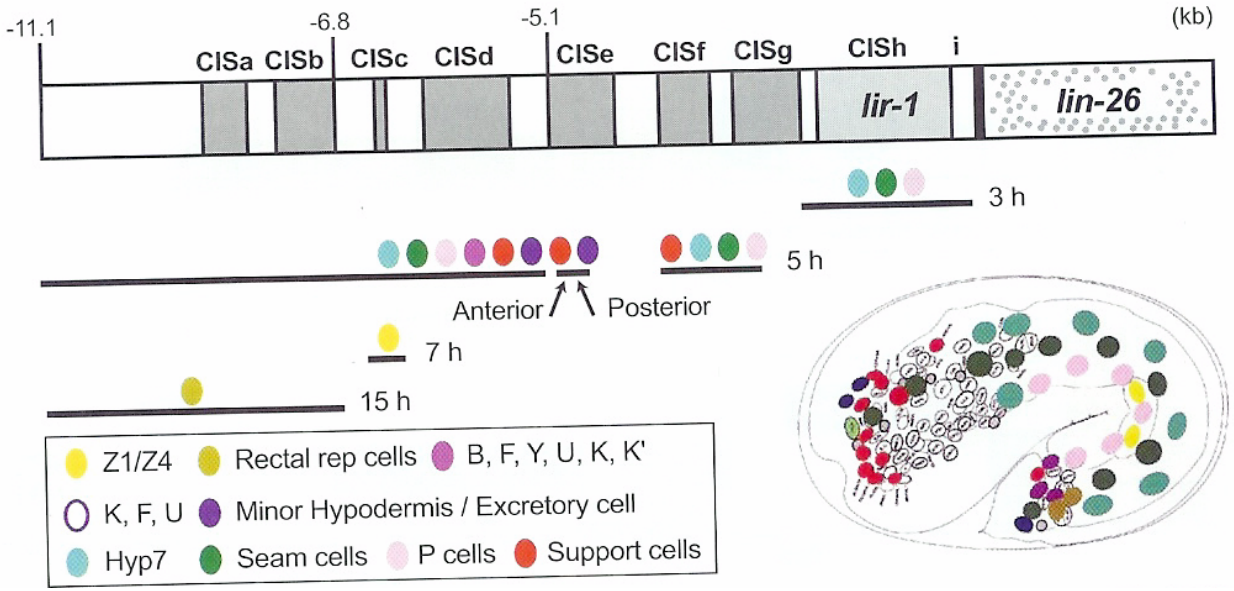
If you delete the circled region expression is altered to the embryo on the right (left is normal)

Terminology

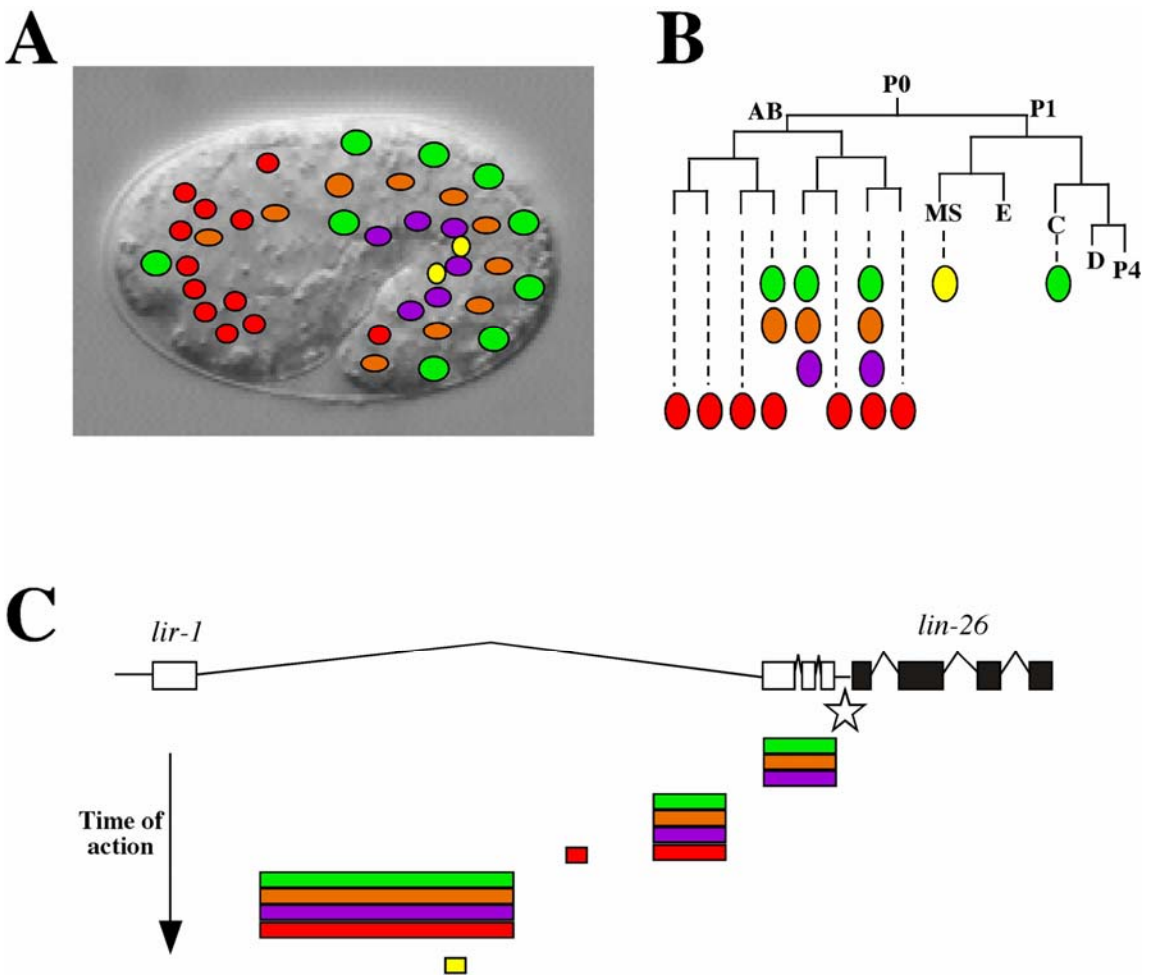
- Myotome:
  - The segment of a somite in a vertebrate embryo that differentiates into skeletal muscle.
  - A muscle or group of muscles derived from one somite and innervated by a single segment of a spinal nerve.
- Somitomeres: In the developing vertebrate embryo, somitomeres are loose masses of paraxial mesoderm-derived cells
  - They form alond each side of the neural tube (~3 wk)
  - Approximately 50 pairs of somitomeres begin developing in the cranial (head) region, continuing in a caudal (tail) direction until the end of week four.
  - The first seven somitomeres give rise to the striated muscles of the face, jaw and throat
  - The remaining somitomeres, driven by periodic expression of the *hairy* gene, begin expressing adhesion proteins (ex. N-cadherin and fibronectin) and then compact and bud off, forming somites.
    - The somites give rise to the vertebral column (sclerotome), the associated muscles (myotome) and the overlying dermis (dermatome)
- Definitions have been restated from: Answers.com and ISCID Encyclopedia of Science and Philosophy

C elegans lin-26 gene

- This is a tf expressed in epithelial cells, germ line cells and the uterus of the worm
- Tissue-specific CREs are in CISa-g.
  - The proximal element i interacts with them



Regulation of lin-26



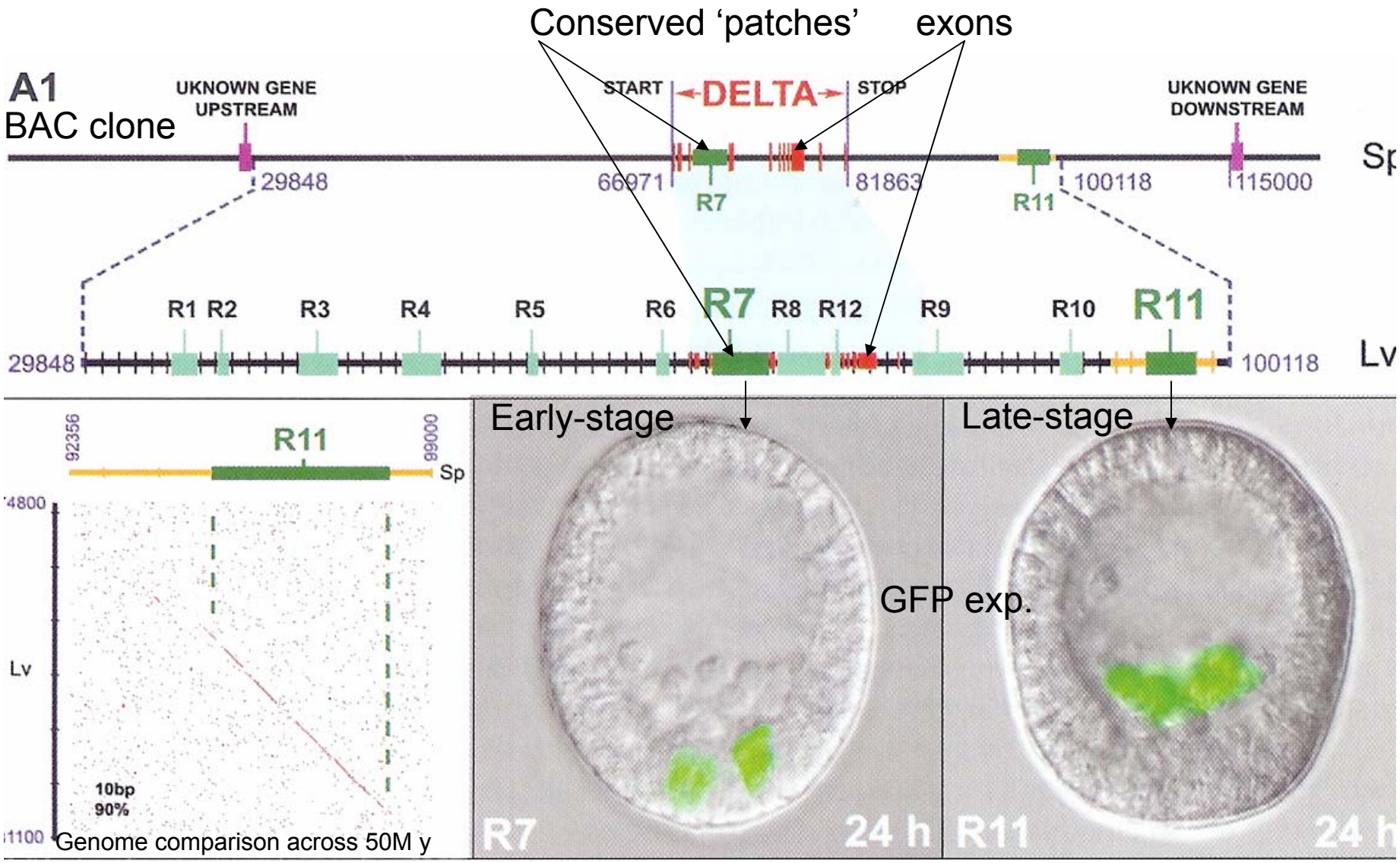
- Regulation by tissue-specific elements.
- A) Shown image is a comma stage embryo.
- B) A partial embryonic lineage showing the origin of lin-26 expressing cells with color coding matching the cell types shown in (A
- major hypodermal cells include
  - hyp 7 (green),
  - seam cells (orange)
  - P cells (purple)
  - support cells (red)
  - somatic gonad precursors (yellow).
- C) The promoter elements for lin-26.
  - All expression is dependent on a “core” element (star) located in the intergenic region between lin-26 and lir-1 Tissue-specific control elements are located within an lir-1 intronwith color-coding as in (A) and (B).
  - Most control elements function in cells related by tissue-type but not by lineage.

http://www.wormbook.org/chapters/www_transcriptionalregulation/transcriptionalregulation.html

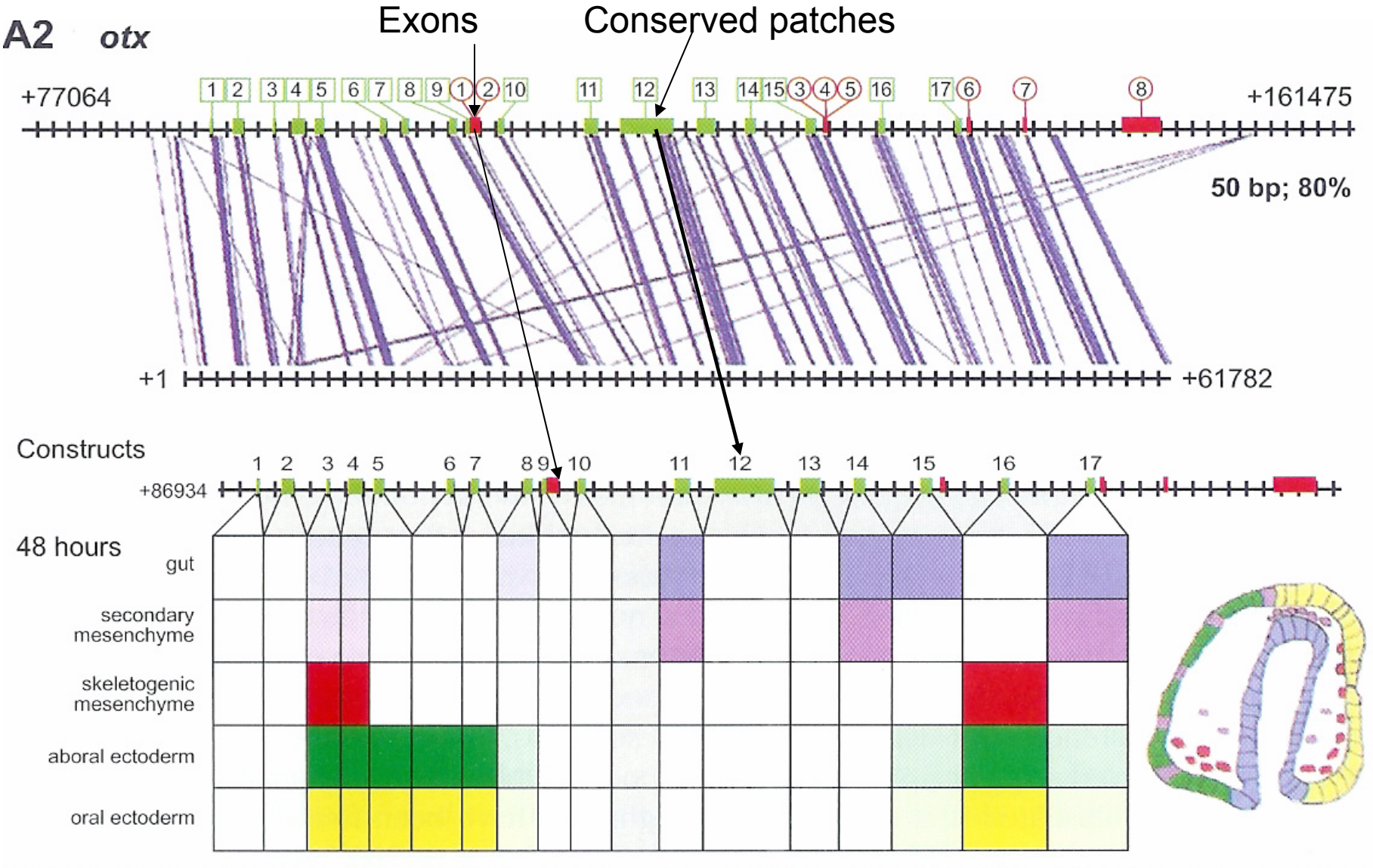
Using Sequence Conservation as a Predictive Method

- In sea urchins the CREs for ten genes were identified in the Davidson lab using just comparative genomics
  - the delta gene
  - the otx gene 2.2 A1, A2
- The sequence 'patches' were then pursued, and ~2/3 of the patches identified had downstream genes expressed in the embryo in the first two days.
- Comparative genomics has allowed modular boundaries to be identified for some comparisons
  - Mouse/human for pax6 and otx2 but a wider outgorup (chicken) was needed for sox2.

Sea Urchin delta gene



Sea urchin otx gene

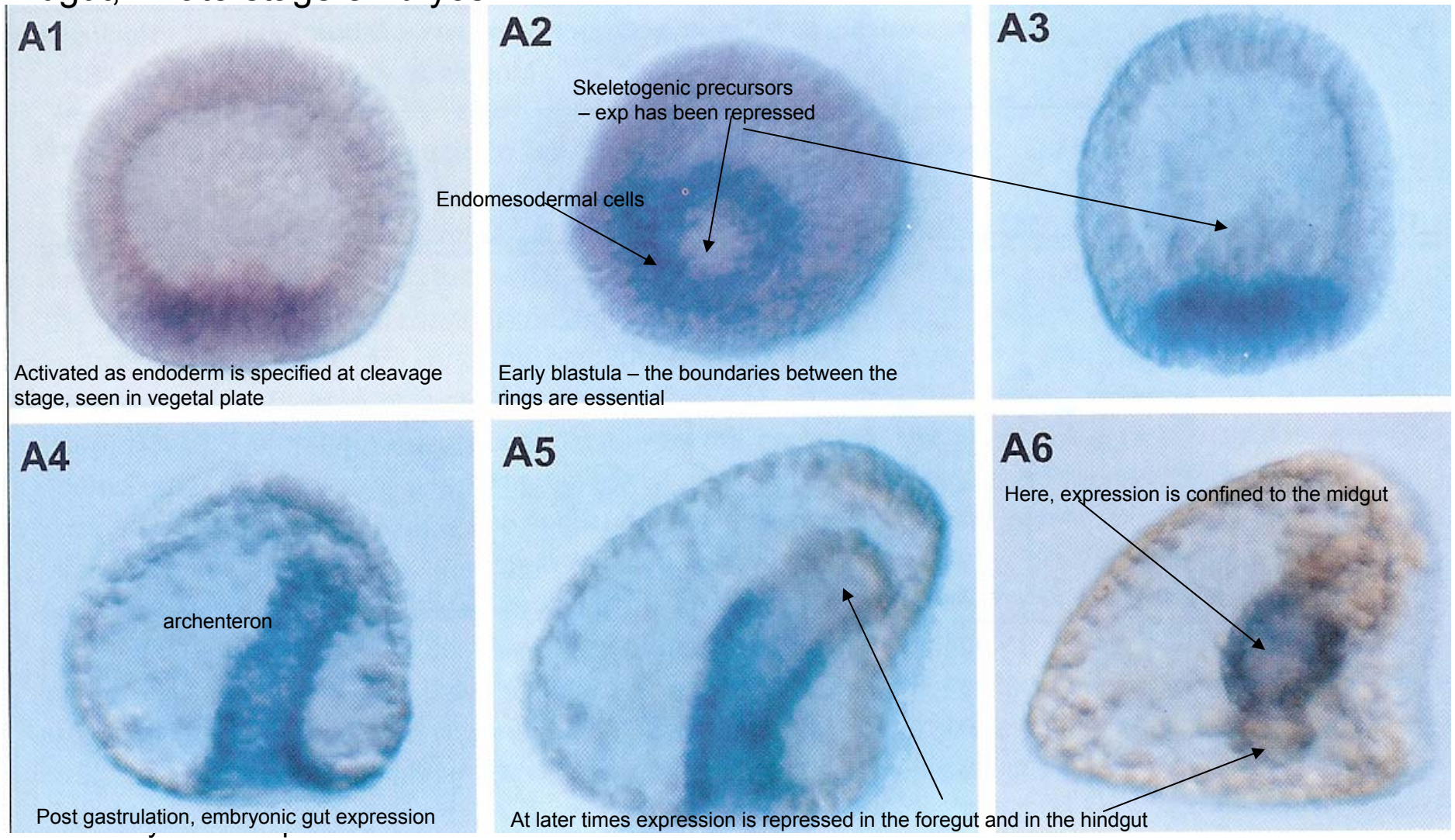


CRE logic deconstruction

- There are two classes of inputs in CREs
  - tfs generate the spatio-temporal variation
  - Other factors mediate between the tfs and transcription apparatus and do not show this type of variation
  - It is the tf binding sites that represent the nodes in the gene regulatory networks that will be diagrammed
- A systematic mutational analysis of tf binding sites shows that
  - At any given state not all of the sites are occupied
  - Every site uniquely influences the output
  - Combinations give output that is not an ‘additive’ effect of the two individual effects

Sea urchin endo-16 gene expression

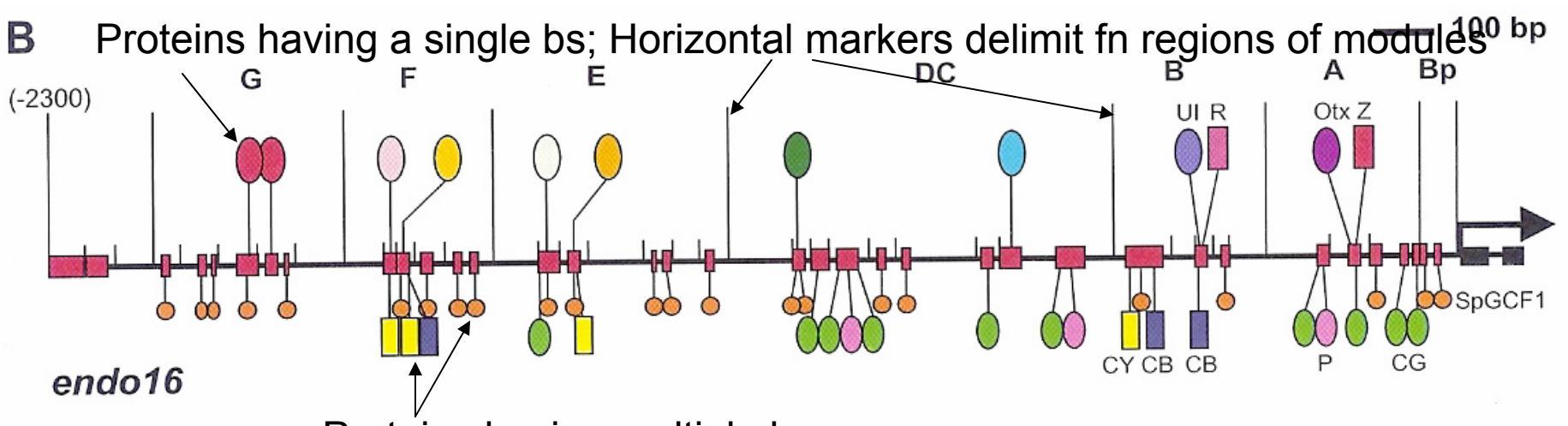
This is a secreted protein expressed in the endoderm, on the inner wall of the midgut, in late-stage embryos.



Background on endo-16

- There are 13 known proteins that bind to sites distributed in 6 CREs (G-A).
- There are repressive interactions mediated by the domains DC, E, F
- There are enhancer regions, G, B, A
  - A is mediated by the tf Otx, which acts as its driver
    - A causes earlier activation and is the core processing unit for the other CREs
  - The spatio-temporal activation of B is driven by Brn1/2/4

Endo-16: Necessary and sufficient segment of DNA to govern correct expression in an egg



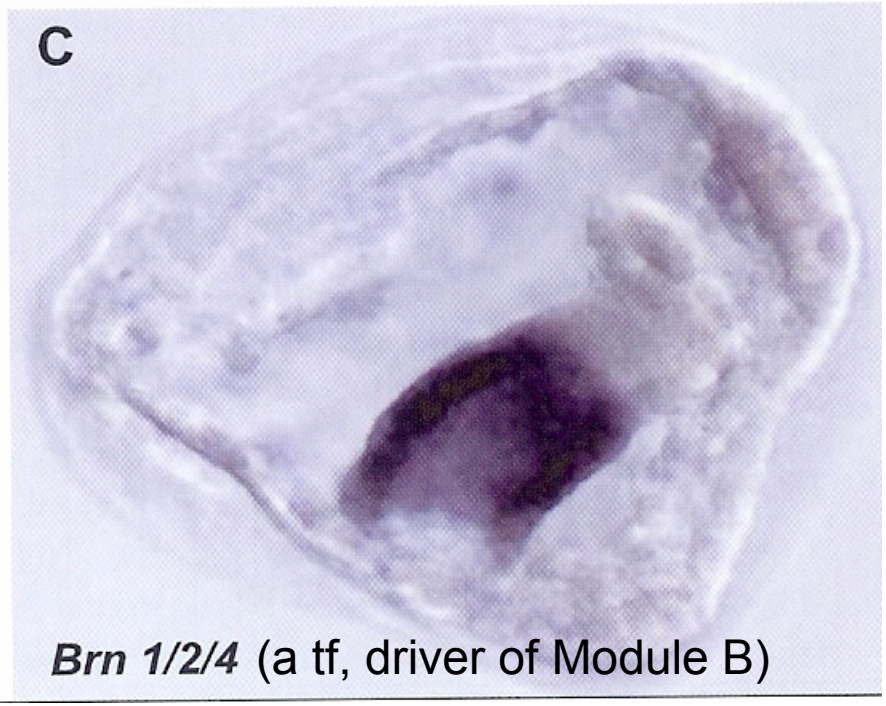
G Positive booster

F
E } Repression in adjacent ectoderm

DC Repression in skeletogenic mesenchyme

B Expression in midgut of late embryo
- Controls late rise in expression
- Activates switch resulting in exclusive use of its own input

A Expression in vegetal plate in early embryo
- Sole communication to BTA for whole system
- Synergistic amplification of B input
- Transduction of FE, DC repression



Brn 1/2/4 (a tf, driver of Module B)

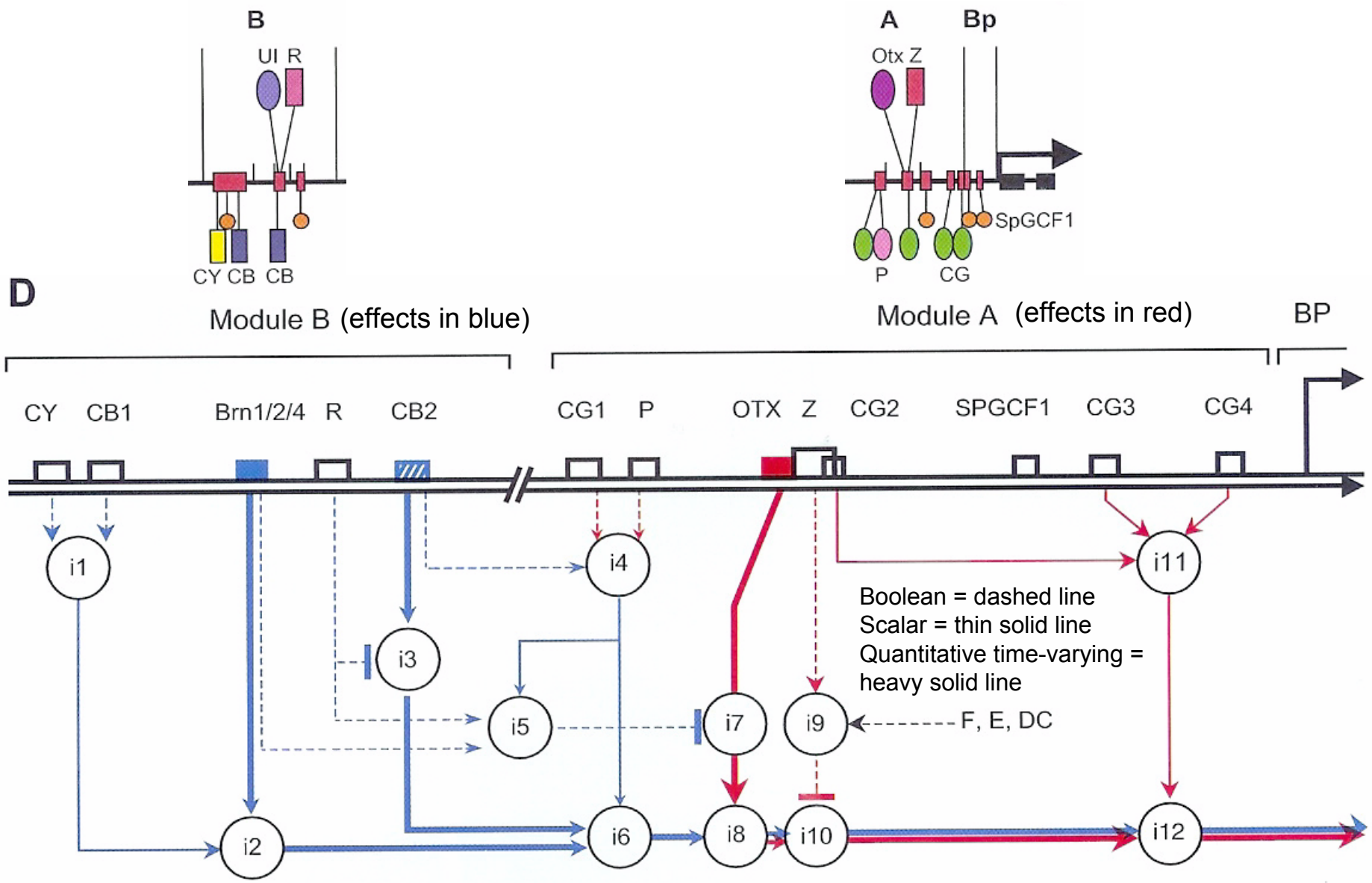
Describing the endo-16 Logic Circuit

- The basal promoter (Bp) is where the RNA polymerase and associated factors assemble in order to make the transcript.
  - All inputs to the Bp are channeled through Module A, which determines whether or not transcription will occur and the rate, when it does.
    - Mutations at specific sites in Module A have been shown to block all upstream effects from reaching the Bp.
- Repressive Modules: these are DC, E and F
  - They are known to be essential since certain regions must have a lack of the product – these constitute negative inputs to Module A
    - The tf Otx is present and active in most cells in early development, and is a postivie driver of Module A. Since it is needed for many functions it cannot be repressed (ie the concentration modulated) – another type of preventative action must be used.

Endo-16 Effects

- Positive inputs to Module A
  - Later in development it is not necessary to modulate Otx, the repressor modules are not needed and all effects are due to modules A+B.
    - A new tf, Brn1/2/4 is the main driver at the later stage, which binds to a site in Module B.
  - Two new interactions must occur between modules A and B:
    - A switch sensitive to the concentration of Brn1/2/4 must be present, which is the R regulator in B
      - Brn1/2/4 is activated after the gut forms (pos gastrulation). The switch turns off the input to the Otx regulator of A → only inputs to B will be effective
    - Amplification in A of the B input, a set-up function is needed.
      - Several proteins (having specific binding sites) are needed: CB2 in B and CG1 and P in A.

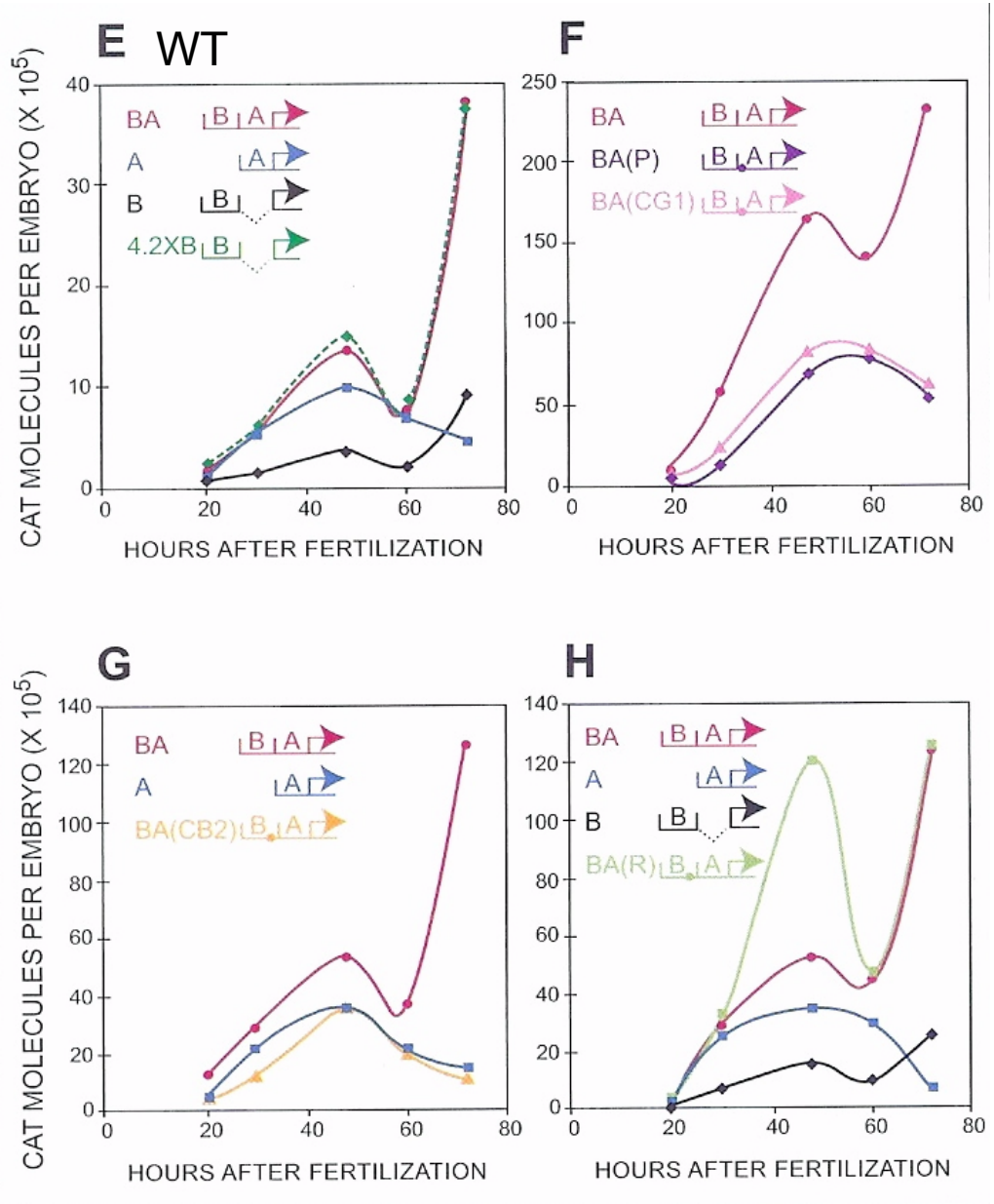
The Logic Circuit Diagram



Time-varying driver intx (colored box), time-and-space invariant intx (open boxes)

if CY & CB1	i1 = 1
else	i1 = 0.5
	i2 = i1 ï Brn124(t)
if R	i3 = CB2(t)
else	i3 = k ï CB2(t) (1<k<2)
if P & CG1 & CB2	i4 = 2
else	i4 = 0
if Brn124(t)>threshold & R & i4 á 0	i5 = 1
else	i5 = 0
	i6 = i4 ï (i2+i3)
if i5 = 0	i7 = OTX(t)
else	i7 = 0
	i8 = i6 + i7
if (F or E or DC) & Z	i9 = 1
else	i9 = 0
if i9 = 1	i10 = 0
else	i10 = i8
if (CG2 & CG3 & CG4)	i11 = 2
else	i11 = 1
	i12 = i11 ï i10

- If CY (target site CY), and i≠0 (here it is 1) means the site is present and occupied.
- i=0 means that either the site was mutated, or the factor was inactivated so productive occupation of the site is not possible – this is 'else'
- CB1 and CY1 interactions together are the i1 circuit, and cause an increase in the output of the spatio-temporal regulator of Module B, which binds at the Brn1/2/4 site
- The output of the Brn1/2/4 module is at i2.
- There is another time-varying positive input generated by the interaction at CB2 (peaking at 40h).
- An interaction at R is required for the B-A intermodule switch, which shuts off the otx input (i5,i7); this operates only if there is input from Brn1/2/4 AND CB2 is present and occupied.



- How the logic is tested: kinetic experiments fr induction and repression of the basal reporter gene chloramphenicol acyl transferase (CAT)
- The construct in its normal architecture, BA, is red
- B alone is black
- A alone is blue
- Targeted deletions/mutations are given in violet, with parentheses and dots

Factors vs Drivers

- In the endo-16 regulatory modules, only 2 of the 9 proteins having specific binding sites function as ‘drivers’ of gene transcription, having spatio-temporal regulatory functions: Otx and Brn1/2/4.
  - Absolute concentration or activation/inactivation is used to fill the need for these drivers
- The other proteins process and modulate the effects of the two drivers, having repressor, switching and amplification effects.
    - Their action is still directly encoded by DNA binding sites.
    - These proteins can be ubiquitous and at generally high concentrations all the time: they can only act when the drivers are present anyway.

Computing and Genomic Regulatory Logic

- What would be needed to compute rather than experimentally test the logic circuits described?
  - First, truly be able to identify all target binding sites
    - These are not yet fully characterized for any species
  - Have available for interpretation the set of functions that occur given a particular pattern of site occupation and factors
    - There are closely related families of tfs that have considerable binding site overlap and functions that are not identical but the distinctions are not well understood
  - Have available the rules for combining the functions in order to infer the output – e.g. when amplification will occur and by what factor
    - Almost none of the combinatorial rules are yet understood
    - Some factors can have opposite effects given a new partner or co-factor
- So computational predictions are a goal, not a reality. In the meantime the approach of using characterized modules with saturated tf characterization allows dissection of complete functional effects and modulators
  - The differences between closely related CREs may start to reveal the rules we seek.

Saturated Characterization of a CRE

- What is needed to completely characterize a CRE?
  - Mutational analysis of the binding sites for sequence and distance
  - Mutational analysis of transcription factor DNA- and protein- and co-factor-binding regions
  - Analysis of functional results
    - Transcriptional activation or repression
    - Intermodule linking
    - Amplification of input
    - Concentraion dependence
    - Types of quantitation transformations
      - Scalar
      - Multiplier
      - Power
      - Boolean

Common Circuit Responses

- For linking between modules is is clear that Boolean logic most commonly applies:
  - AND meaning that a positive output only occurs in domains where two different regulatory factors are coincidentally bound.
- Switch-like behavior is the most frequent response to signal transduction events
  - Intercellular ligands are used to modify a tf, for example if there is no signal ligand and the tf is present then it will act as a dominant repressor
  - If a ligand and a tf are present (and usually a co-activator as well) then the tf cannot repress the Bp and transcription will occur
  - The repression of transcription is usually Boolean and usually dominant.

Target site occupancy

- Transcription factors have a binding constant in the formation of the complex with the DNA binding site.
- This is a bimolecular chemical reaction and thus depends on the concentration of the two reactants
  - The more of a reactant is present the more you push the reaction toward the product
  - Thus occupancy is determined by the binding constant (an intrinsic property) and the nuclear concentration of the DNA sites (invariant) and the transcription factor (which can be modulated).
  - The binding equilibrium of the DNA-tf duplex may be modulated by additional interactions with adjacent proteins, expressed as cooperativity constants

Measuring Binding Constants

- It is possible to label both the protein and DNA and measure the on- and off-rates of duplex formation
  - You can do competition experiments with labeled and unlabeled protein.
- It turns out that on-rates (the protein finding the recognition site and binding, which is diffusion driven) do not vary much over a wide range of tf classes (for a given concentration) but the off-rates vary by orders of magnitude.
- In terms of relative concentrations, in animal cells, there are hundreds of sites and thousands to tens of thousands of a given tf at a given time.
  - The tfs bind transiently to DNA for which no strong interaction occurs (very transient) mostly.
  - For positively acting factors the site must be occupied if you see initiation of transcription (which has an invariant rate once it has begun).
    - It is possible to measure transcript production rates, and processing and cytoplasmic transport are not rate-limiting, it is possible to measure protein production rates; if you know transcript and protein production and turnover rates you will know relative concentrations of proteins at any given time.

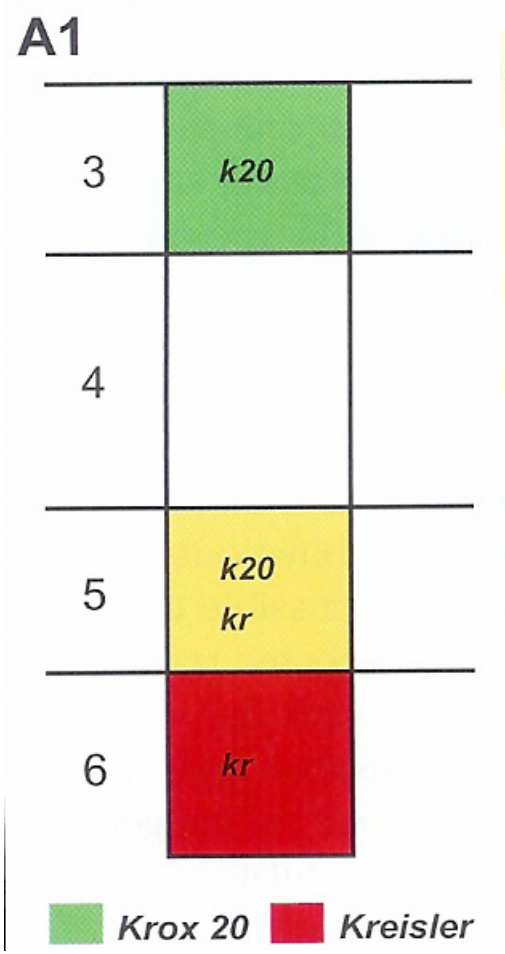
Cis-Regulatory Design

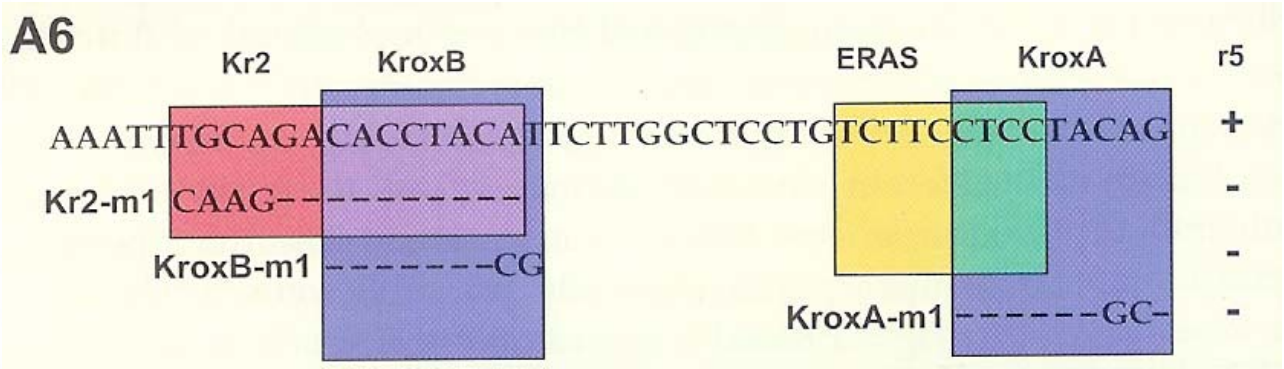
- In development a primary function of cis-regulatory information processing is to integrate diverse driver inputs so that the output is always unique – AND logic again.
  - Two factors must be present together – why?
    - The interaction may give cooperative binding (greater stability)
    - The interaction may give the complex the ability to bind a co-factor
    - Each factor may be correctly oriented to give needed input to the basal transcriptional apparatus (BTA)

Rhombomere development

- The next example shown is rhombomere specification in the amniote hindbrain.
- Rhombomeres form short-term metameric units along the anterior/posterior axis of the hindbrain
  - each unit is distinct by virtue of a particular regulatory state
  - each leads to development of a specific set of ganglia in the face/lower head and population in the neural crest
- The correct expression of a specific hox gene (hoxb3 in mice)specifies the state leading to the required and subsequent outcome for each rhombomere.

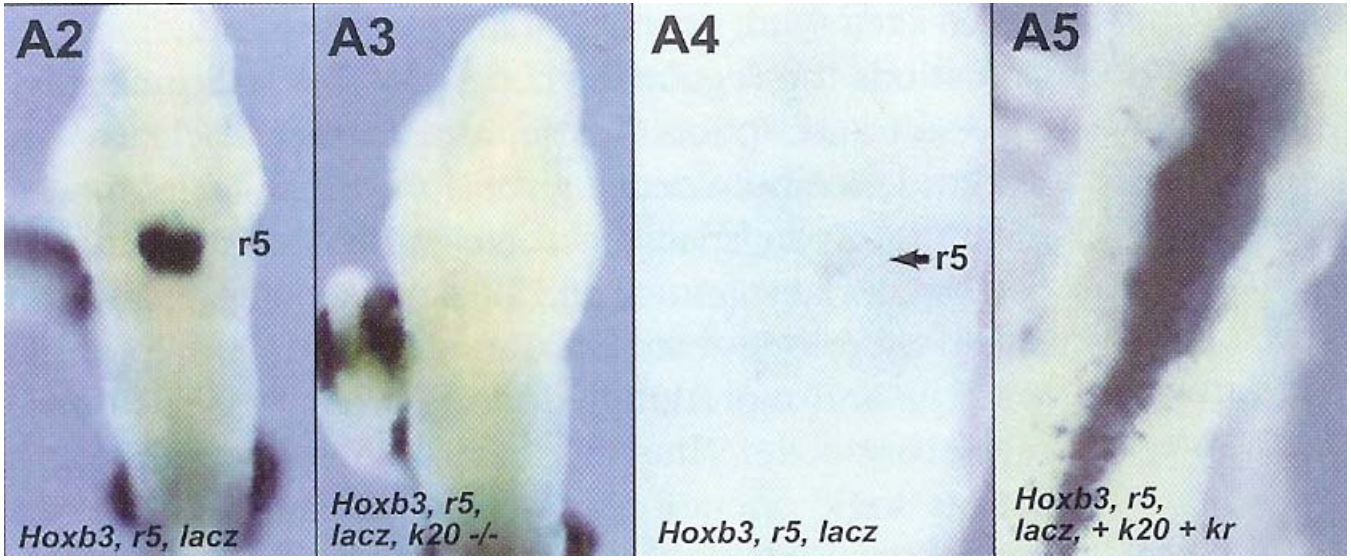
Mouse hoxb3 gene example





- An enhancer element controls hoxb3 in rhombomere 5 (r5)
- There are two spatially distinct drivers having input, the tfs Krox20 (k20) and Kreisler (kr).
- As shown, k20 is expressed in r3, r5 and kr is expressed in r5, r6 so they overlap in r5.
- The resulting function only happens when both are present (AND), the integration executed by the hoxb3 enhancer gives r5-specific expression
  - Show that if you knock out either k20 or k3 binding the activity is destroyed

Experimental Verification

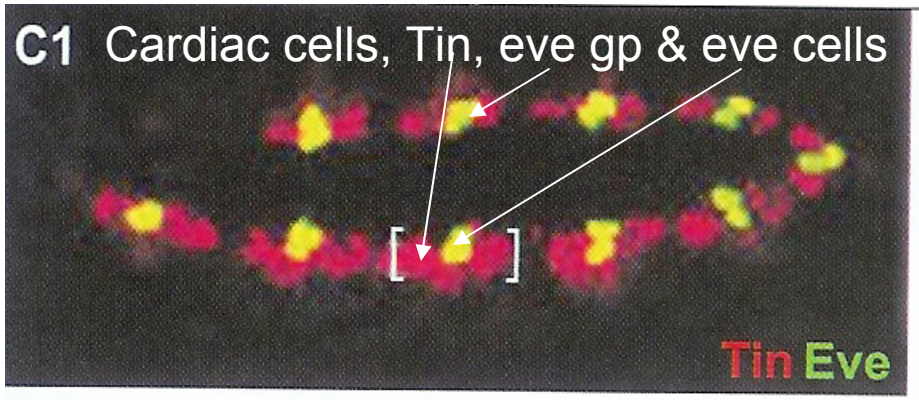


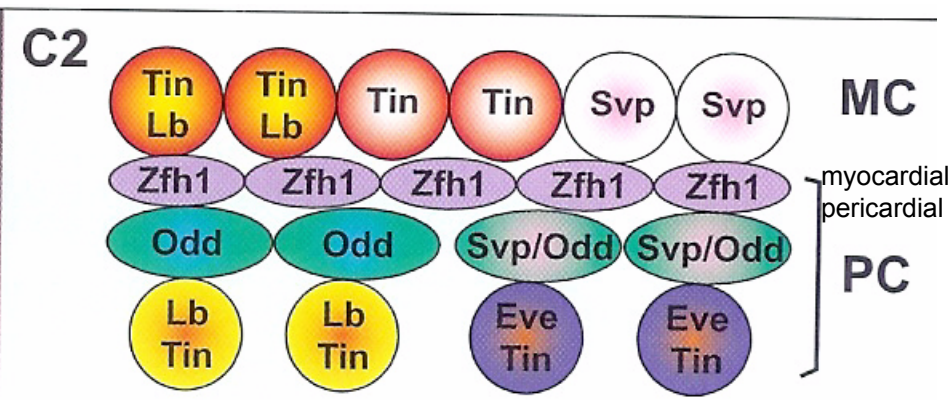
- A2 shows where the r5 unit is in a mouse embryo, a lacZ construct behind the hoxb3 module is being expressed
- A3 has mutated both k20 sites – now no activity can be induced
- A4 puts in the hox3b element but without significant amounts of endogenous k20, kr no expression occurs
- A5 adds expression vectors that make k20 and kr constitutively and now the lacZ fusion is everywhere, ectopic (other controls showed that the k20 and kr alone did not give such a result).

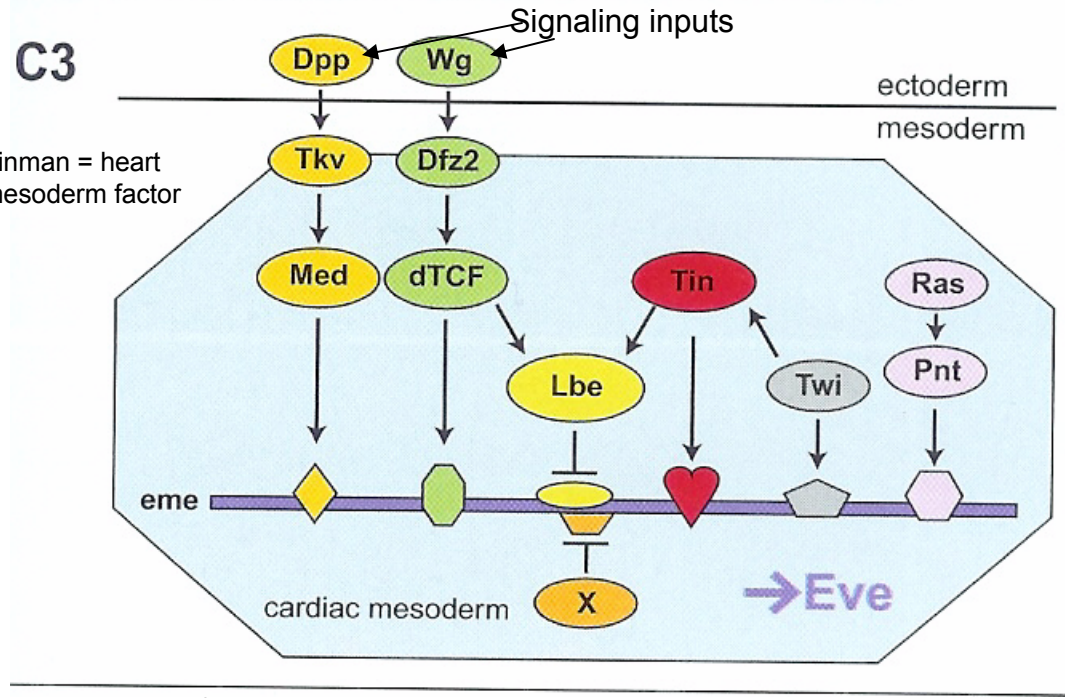
Conditional Processors

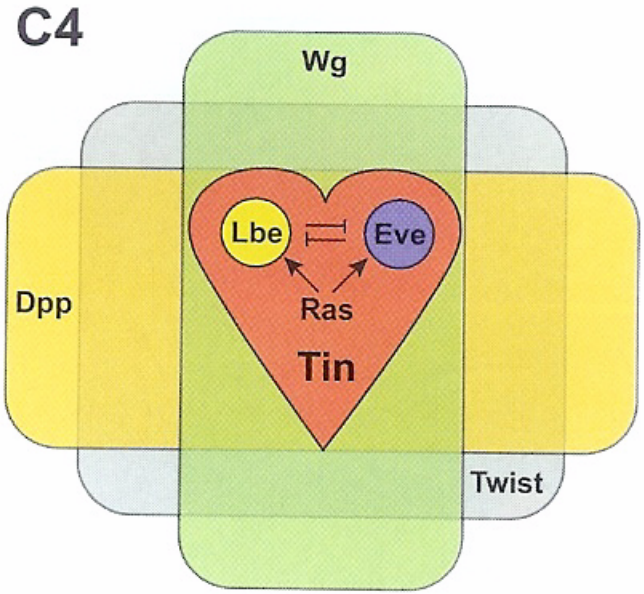
- Conditional processors have both positive and negative inputs. One example is the *eve* gene in D.m., which is controlled by a CRE that has binding sites for 5 proteins: 4 positive and one negative regulator.
- Positive include
  - *Twist* (expressed in all mesodermal cells and some other types of ventral cells
  - *Tinman* (expressed in dorsal mesodermal cells, some of which become the heart progenitor cells)
  - *Wg* (a signaling ligand expressed in the ectoderm covering this region, in stripes along the anterior/posterior axis→ activates the tf Tcf in heart cells)
  - *Dpp* (another signaling ligand expressed in the dorsal ectoderm)
  - None of these coincide completely spatially, but there is a small region where they all overlap
- Negative is
  - *Ladybird* (expressed in anterior pericardial cells → thus *eve* is specifically not expressed in those cells)
  - Repression is dominant, so transcription is turned off when *Ladybird* is present, and turned on when Ladybird is absent AND all four of the others are present on the CRE sites for eve.

The *eve* enhancer in D.m.









Repression and cis-Reg Design

- If you have several sets of CREs controlling the expression of one gene then you can combine repressors with activators to get quite complex spatial patters.
- Ex is eve (evenskipped) in D.m.
- qIn stripe 2 one sees expression of even that is a few cells thick at a particular location on the A/P axis
  - The stripe 2 module gives transcriptional activation in response to the Bicoid (Bcd) and Hunchback (Hb) regulators
    - Bcd also acticates the hb gene.
    - The Bcd protein is distributed on a concentration gradient that decreases homogeneously from A → P
    - The Hb protein is present in the Anterior 40% of the embryo
    - The stripe 2 module has several Bcd binding sites, having different k_bs.
    - The Bcd activators are cooperative with one another and with Hb

Setting Boundaries

- The stripe boundaries depend on specific repression of transcription, mediated by the Kruppel (Kr) and Giant (Gt) factors
  - For the posterior boundary is set through a response to the presence of Kr
  - The closest anterior boundary is set by the response fo Gt (+?) in specific cells
  - Further anterior: set by the repressor Cloppypaired1 (Slp1)
  - Etx to the farthest anterior cells
  - Note the Gt and Slp1 are activated by Bcd themselves
- Other stripes are controlled with similar mechanisms, but different specific repressor combintaions – this lets boundaries be set autonomously, which makes the system much more robust.

Evenskipped and stripe 2

Hb = dk green, ant Bcd= blue→ magenta, even =red, eve+Hb = yellow