

### L'ascidie *Ciona intestinalis*: Du génome aux propriétés cellulaires via les réseaux transcriptionnels.

Patrick Lemaire and colleagues  
IBDML, CNRS/université de la Méditerranée  
Marseille, France



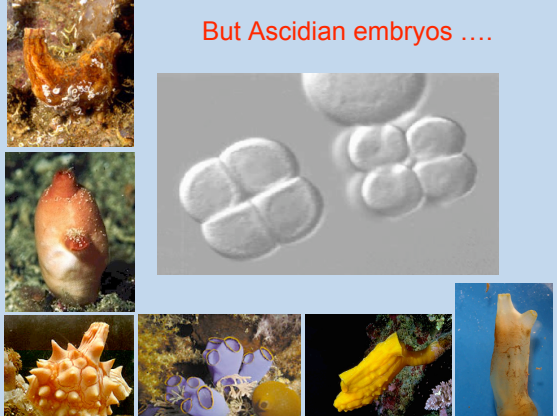
Developmental Biology Institute of Marseille Luminy  
UMR 6216 • CNRS • Université de la Méditerranée





### Adult ascidians don't do much...

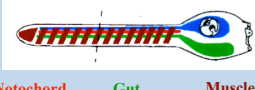




### But Ascidian embryos ....



### Body plan of chordate embryos

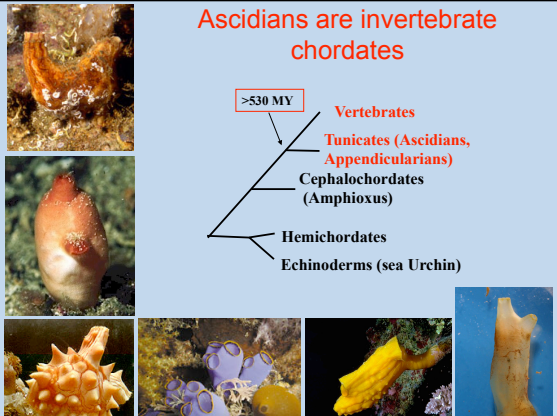
Amphibian		Cell number >10 <sup>6</sup>
Ascidian		2600

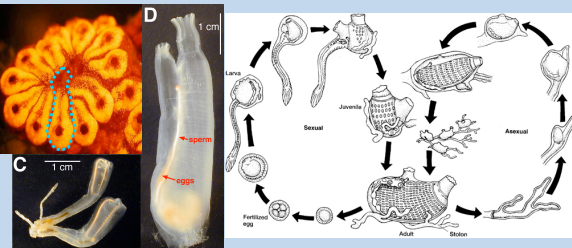
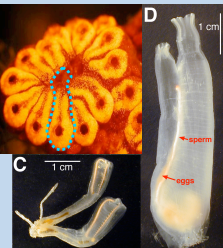
### Ascidians are invertebrate chordates

>530 MY

- Vertebrates
- Tunicates (Ascidians, Appendicularians)
- Cephalochordates (Amphioxus)
- Hemichordates
- Echinoderms (sea Urchin)



### Colonial ascidians: two genetic programmes lead to the same adult form

### A brief introduction to chordate development

**Egg** The embryo is a ball of cells, sometimes with a hollow cavity, the blastocoel.

**Cleavage** Endoderm (gut, liver) and mesoderm (muscle, bones, kidneys, blood) are engulfed into the embryo, which become surrounded by the future skin and nervous system

**Gastrula**

**Neurula** The neural tube closes, the body plan starts to elongate

**Tailbud** Elongation of the tail and trunk

**Larva** Final stage of embryonic development. The organisms usually becomes free living and autonomous

### How do cells know what to do?

- They progressively see their fate restricted until they ultimately adopt a given fate (eg a neuron, a muscle cell)
- They contribute to shaping the body during development by dividing, migrating, or changing shape.
- They talk to one another to coordinate their behaviour

### Ciona embryos have a fixed lineage

A lineage tree diagram showing the developmental path from a single cell to various tissues. Key stages and cell types include:

- A5.1: endoderm, notochord, spinal cord
- A6.3: endoderm, trunk, lateral cells
- A7.7: no outburst
- A7.8: A8.15: spinal cord, A8.16: muscle
- a.2: endoderm, brain, sensory pigment cells, pharynx
- B.1: muscle, mesenchyme, notochord, endoderm, end-dermal strand
- 7.9: b8.17: muscle, spinal cord, endodermal strand
- b8.18: endoderm
- b8.19: spinal cord, fin, stem
- b8.20: endoderm
- b.2: b6.6: endoderm
- b.5: endoderm

Hiroki Nishida

### The Ciona genome(s)

- *Ciona intestinalis* and *Ciona savignyi* were sequenced in parallel. Very similar animals but genomes as different as human and chick.
- 160 Mb
- 16000 to 20000 genes, full ORF Unigene cDNA collection available.
- Compact intergenic regions, no non-coding sequence conservation with vertebrates.
- Predates vertebrate genome duplications (80% single copy genes)
- Contains most gene families present in vertebrates (except adaptive immunity, neurotrophins, olfactory/pheromone receptors)

Dehal et al. *Science* 2002

### A regulatory gene spatio-temporal expression atlas

Over 70% (500) of TFs detected in the *Ciona* genome were analysed by ISH. 109 signal transduction genes in major signalling pathways analysed similarly.

Total number of genes analysed by ISH: ~5000

Imai et al., 2004

### Transient transgenesis by electroporation

Fertilisation + Electroporation → Culture → Expression of pBrachyury-GFP, pGsx-NLSLacZ, and pGsx-TauLacZ


Hudson, Bertrand

### Experimental perturbations of Ciona development

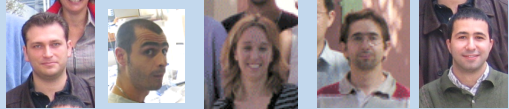
- There are some, but few genetic mutants.
- Efficient loss-of function:
  - Injection of anti-sense morpholino oligonucleotides in unfertilised or fertilised eggs --> blocks translation of target mRNA.
  - Incubation in pharmacological inhibitors
- Gain of function:
  - by electroporation of DNA constructs
  - simple soaking into signalling ligands (eg FGF, Bmp).

### N.I.S.E.E.D. Network for In Situ Expression and Embryological Data

<http://aniseed-ibdm.univ-mrs.fr>

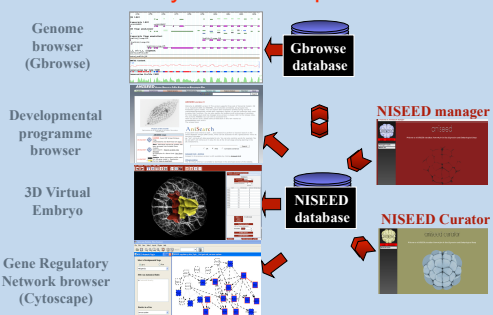


- Cell Identity
- Neighbours
- Origin & fate (lineage)
- Interactions
- Inductions
- Expressed regulatory genes
- Function
- Regulation



O. Tassy   F. Daian   D. Dauga   D. Sobral   P. Khoeiry

### ANISEED a digital framework to represent embryonic development.



Genome browser (Gbrowse) → Gbrowse database

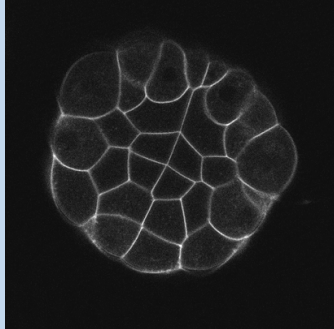
Developmental programme browser → NISEED manager

3D Virtual Embryo → NISEED database


Gene Regulatory Network browser (Cytoscape) → NISEED Curator

Fabrice Daian, Daniel Sobral, Delphine Dauga

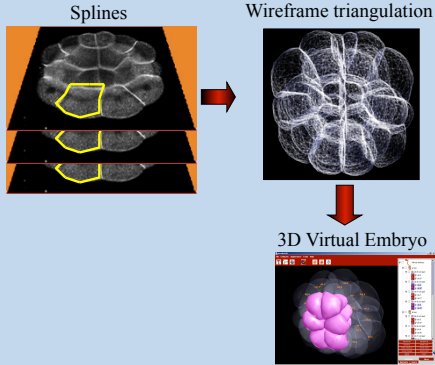
### Imaging ascidian embryos.....



### Filming live embryos in 3D, from the 44-cell to the late gastrula stage.



### 3D Embryo Reconstruction



Splines → Wireframe triangulation → 3D Virtual Embryo

O. Tassy  
F. Daian

### A collection of reconstructed embryos....

Cleavage stages up to 44-cell

Tassy et al., Curr Bio 2006

### Reconstructing gastrulation

Cleavage (64-cell)    Cleavage (76-cell)    Early Gastrula (112-cell)    Gastrula (128-cell)

### Integrating lineage and volume information to identify unequal cleavages

### Gastrulation is driven by individual cell shape changes

### Biometry analysis of embryonic blastomeres

**A**

**B**

Shape descriptors									
Conclude selected cells									
cell	Sphericity	Elongation	Flatness	Squareness	Entropy	SV	Convexity	v/V	
A6.2	178.28	14.82	12.42	49.04	95.53	14.57	86.65	3.43	
B6.3	62.85	28.33	52.95	39.85	76.03	28.86	39.27	0.73	
A6.3	87.83	14.81	7.87	44.85	97.23	12.25	33.86	6.56	
B6.1	70.84	11.47	31.54	31.98	51.29	15.33	77.38	3.74	
B6.8	72.03	47.67	18.57	41.81	66.24	16.94	34.98	0.68	

**C**

Tassy et al., Curr Biol. 2006

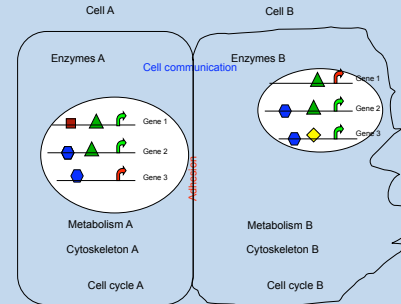
### Evolution of cell shapes between early and late 112-cell embryos.



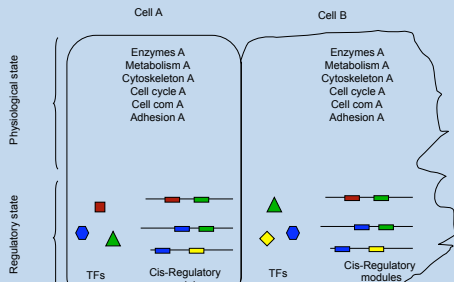
### How is cell behaviour controlled by genomic information?

- Ultimately, the genome encodes the proteins that dictate their behaviour to cells (shape, adhesion, migration, fate)
- Initiation of transcription is one of the major events that specifies in which cell a given protein or protein complex is present.
- Understanding the transcriptional programme may thus help us understand how cell adopt their physiological states.

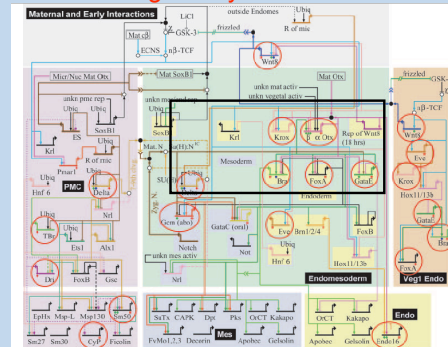
### From physiological state to regulatory state



### From physiological state to regulatory state

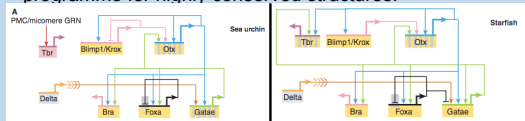


### The Urchin endomesodermal Gene Regulatory Network

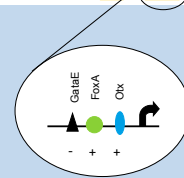
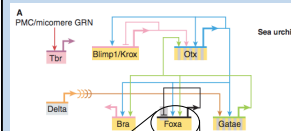


### Why study Gene Regulatory Networks?

- They constitute the regulatory programme for development, and a necessary step towards the dynamic modelling of the regulatory program
- Their architecture reveals general principles that cannot be appreciated at a smaller scale level of analysis: The networks have an intrinsic modular structure, and modules can be co-opted for novel functions.
- GRN modules evolve at a different pace. Some, “the kernels”, are evolutionary very robust and encode the programme for highly conserved structures.



### How to reconstruct Gene regulatory networks?



- 1) Take census of all transcription factors and co-factors
- 2) Determine their expression profiles
- 3) Identify their enhancers
- 4) Identify their upstream regulators

### How to reconstruct Gene regulatory networks?

- 1) Take census of all transcription factors and co-factors encoded by the genome
  - 358 core transcription factors
  - 147 possible transcription factors
  - 462 possible transcriptional co-factors
  - (Human >1400 probable TFs)
- 2) Determine their expression profiles
- 3) Identify their enhancers
- 4) Identify their upstream regulators

### How to reconstruct Gene regulatory networks?

- 1) Take census of all transcription factors and co-factors encoded by the genome
  - 358 core transcription factors
  - 147 possible transcription factors
  - 462 possible transcriptional co-factors
  - (Human >1400 probable TFs)
- 2) Determine their expression profiles
  - >80% of TFs with a full description of expression by ISH
- 3) Identify their enhancers
- 4) Identify their upstream regulators

### How to reconstruct Gene regulatory networks?

- 1) Take census of all transcription factors and co-factors encoded by the genome
  - 358 core transcription factors
  - 147 possible transcription factors
  - 462 possible transcriptional co-factors
  - (Human >1400 probable TFs)
- 2) Determine their expression profiles
  - >80% of TFs with a full description of expression by ISH
- 3) Identify their enhancers
  - In vivo determination of Transcription factor binding sites (Chromatin immunoprecipitation)
- 4) Identify their upstream regulators

### Overview of transcriptional regulation by promoters and enhancers

Small scale studies revealed that functional enhancers often contain multiple binding sites for a small number of transcription factors, and are located in regions of non-coding sequences conserved between related animals.

Figure 1 Transcriptional regulation by promoters and enhancers. General

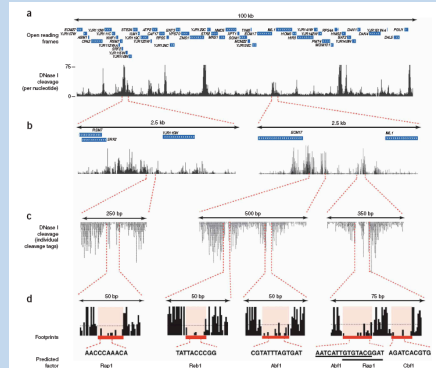
### Chromatin immunoprecipitation-high throughput sequencing (ChIP-seq)

### Chromatin immunoprecipitation-high throughput sequencing (ChIP-seq)

### Issues with ChIP-seq

- Difficult to carry out on thousands of transcription factors (need specific antibodies (preferentially 2/factor))
- Need a sufficient amount of starting material (difficult to carry out on minor cell populations).
- Alternatives:
  - Identification of all binding sites, irrespective of the factors they bind (DNase1-seq)
  - Search for clusters of Binding sites, conserved between related animals

### Genomic DNase 1 footprinting

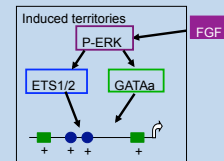
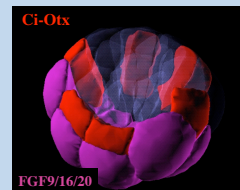


### Ciona Otx as a model system for the genome-wide computational identification of metazoan enhancers



Pierre Khoeiry

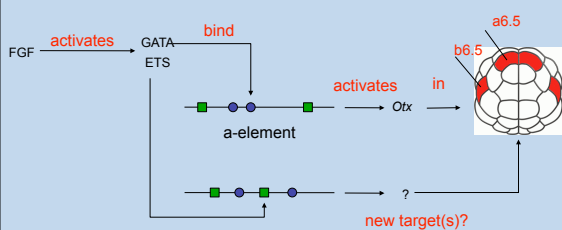
### Activation of Otx during neural induction



*C. intestinalis* TTGGAICTGAAGCTCGT **TATC**TCTAAC **GGAA**GTTTTCGAAA **GGA**ATTGTTCAATATCTAA **GAT**GGGA  
*C. savignyi* AATAATCT **TATC**TCGATTGCAA **GGAA**GTTTTCGAAA **GGAA**GCTAC---GATATCTAC **GATA**ACC

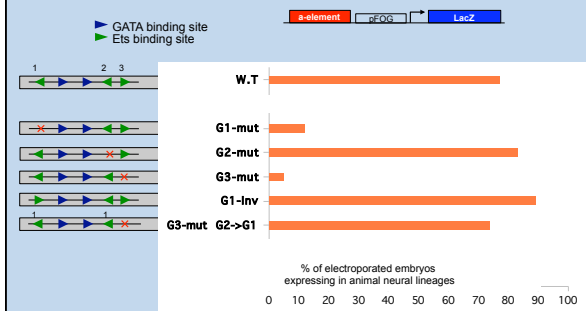
Bertrand et al., 2003

### Principle of the search for additional targets



Pierre Khoeiry

### Only 2 GATA sites are required in the Otx a-element



### Challenges of cluster searches

>800000 and >600000 binding sites for GATA and HGGAWR (GATA and ETS Respectively)

36000 clusters of 2 GATA (GATA) and 2 ETS (HGGAWR) sites in a sequence window of 80 bp

**Syntactic Search**  
Select clusters that match the Otx cis-regulatory syntax (spacing, order and orientation of ETS and GATA sites)

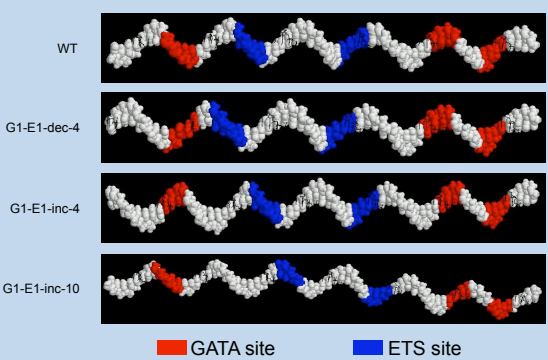
**Phylogenetic footprinting**  
Select clusters evolutionary conserved in orthologous non-coding regions of *Ciona intestinalis* and *Ciona savignyi*.

Pierre Khoueiry

### Is there a syntax in the a-element?

- There are different types of enhancers:
  - Bill board model: as long as you have the correct binding sites no matter how they are arranged, the enhancer works.
  - Enhanceosome: the arrangement of binding sites matters for the activity: eg the Transcription factors bind cooperatively or have to interact between each other.

### Altering the spacing between TF binding sites.



WT

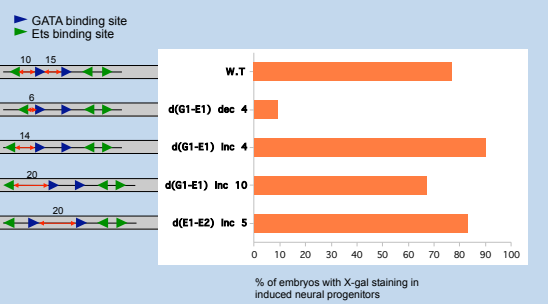
G1-E1-dec-4

G1-E1-inc-4

G1-E1-inc-10

■ GATA site      ■ ETS site

### A precise spacing of the ETS and GATA sites is not required



▶ GATA binding site  
▶ Ets binding site

Configuration	% of embryos with X-gal staining in induced neural progenitors
W.T	~80
d(G1-E1) dec 4	~10
d(G1-E1) Inc 4	~90
d(G1-E1) Inc 10	~70
d(E1-E2) Inc 5	~85

% of embryos with X-gal staining in induced neural progenitors

### Bioinformatics approach Flowchart

2 GATA (GATA) 2 ETS (HGGAWR) in a window of 80bp

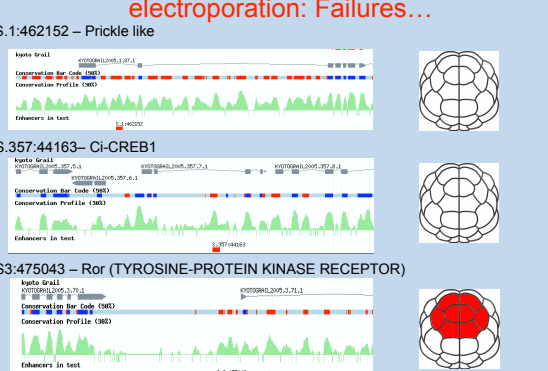
- Pan genomic identification of individual binding sites for TFs of interest (856586 GATA and 191709 MGAAR sites)
- Binding sites clustering in reference genome (5085 clusters in a window of 80 bp)
- Exclude clusters with inter-site distance smaller than allowed minimum (5 bases) (2334 clusters)
- Select clusters that can be aligned to the 2 Genomes (508 clusters)
- Select clusters having THE REQUIRED NUMBER OF SITES in orthologous location in second genome (9 clusters)

### Experimental validation of 26 clusters by electroporation: Failures...

S.1:462152 – Prickle like

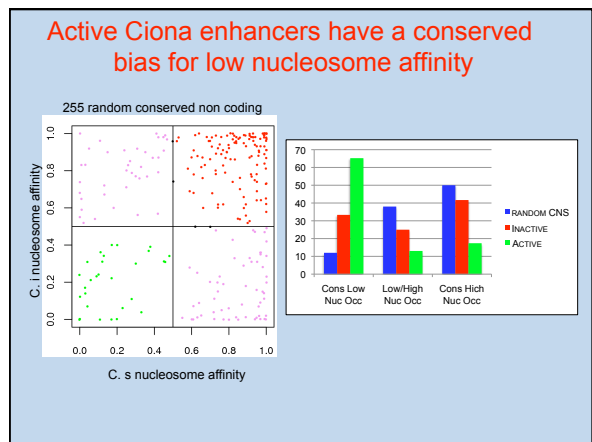
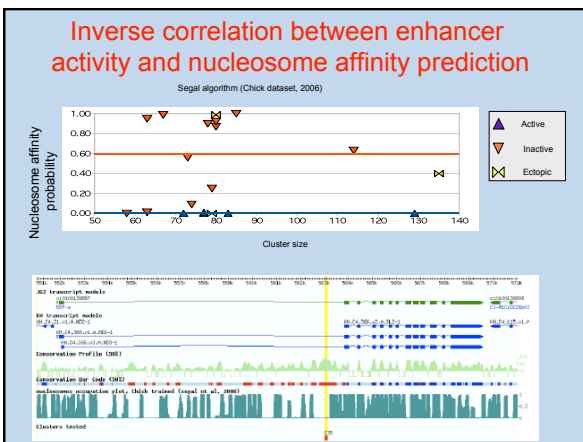
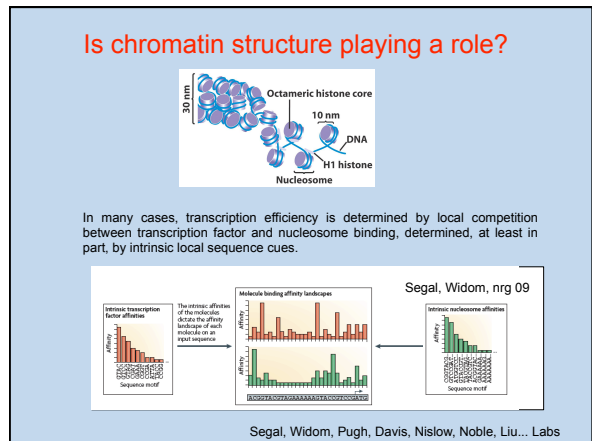
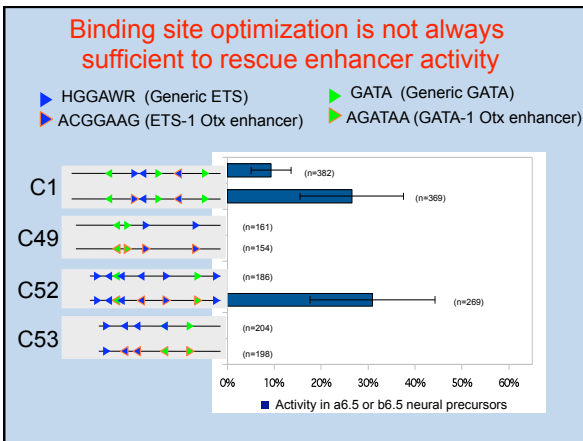
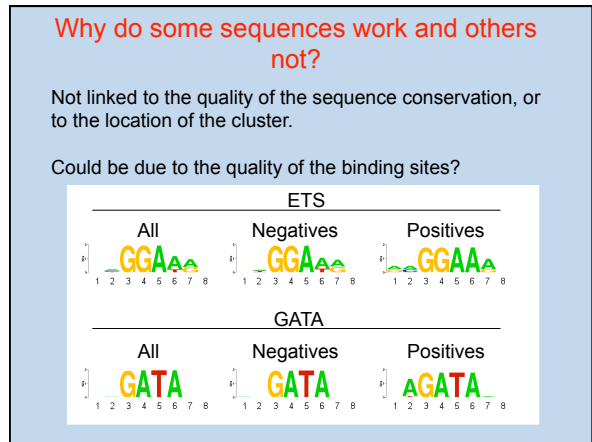
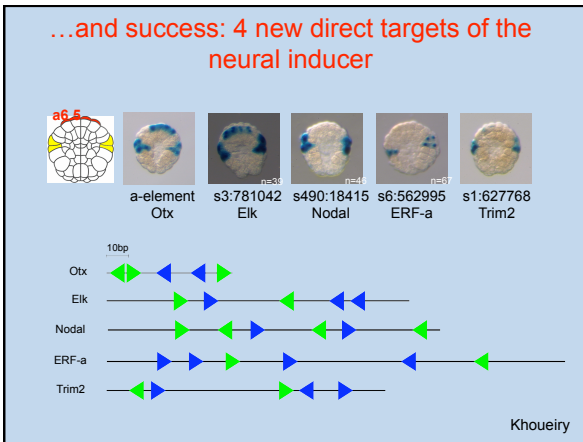
S.357:44163– Ci-CREB1

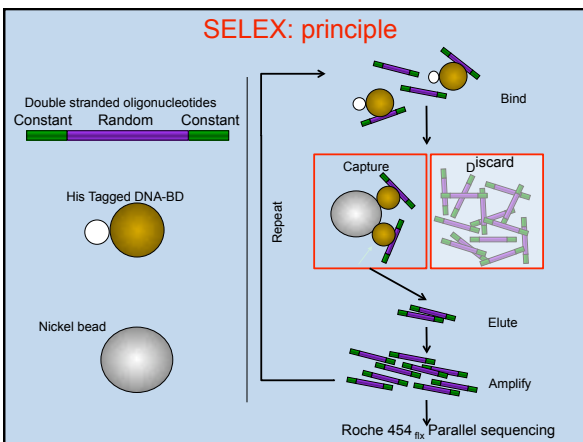
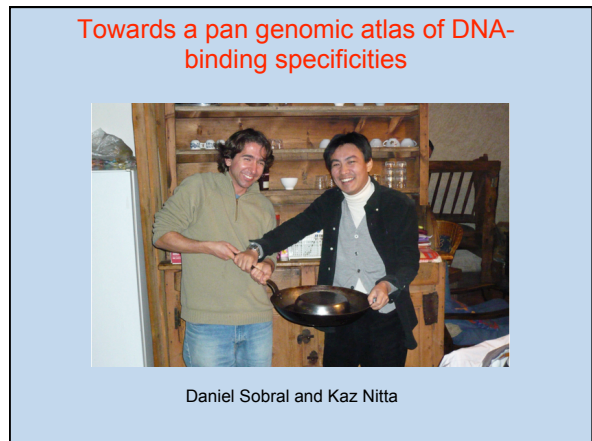
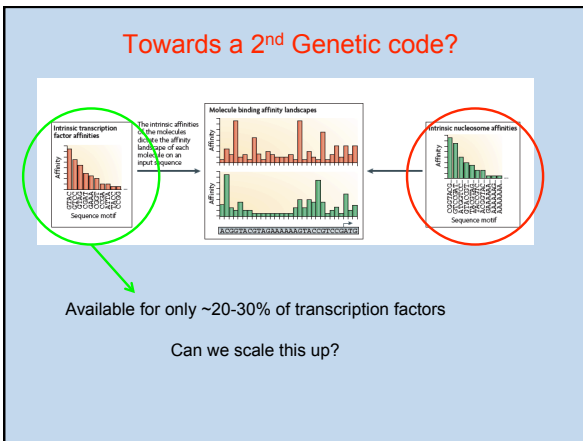
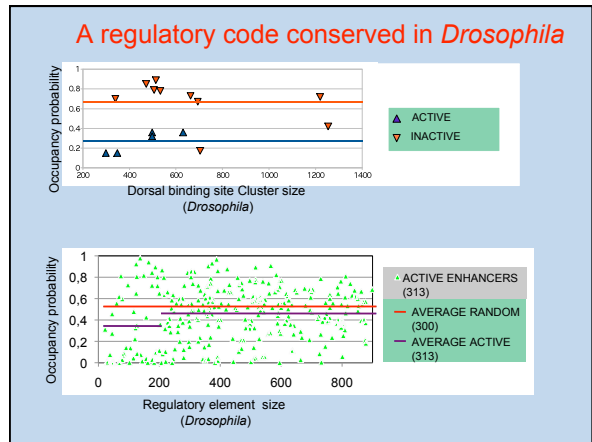
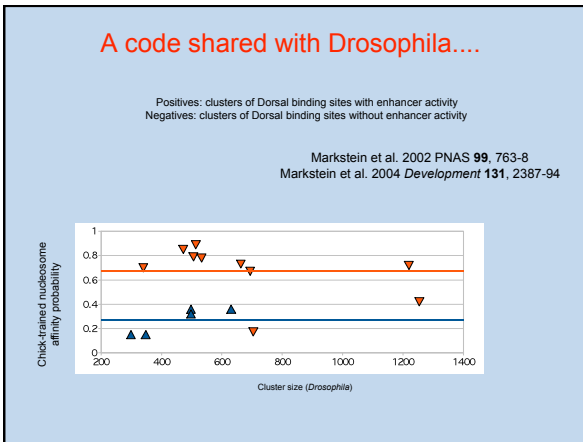
S3:475043 – Ror (TYROSINE-PROTEIN KINASE RECEPTOR)



Pierre Khoueiry

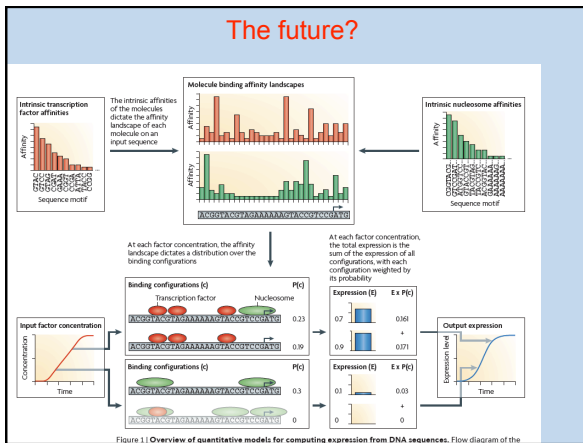
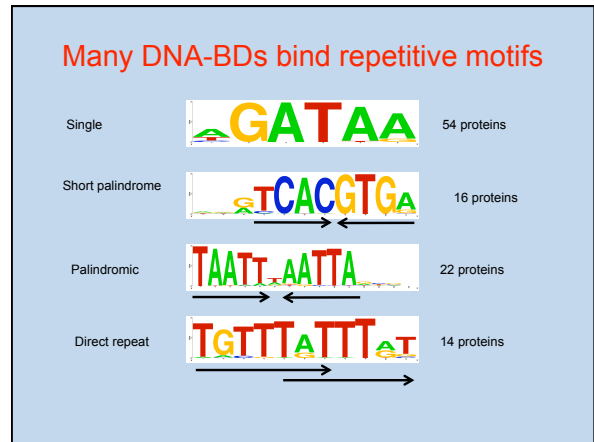
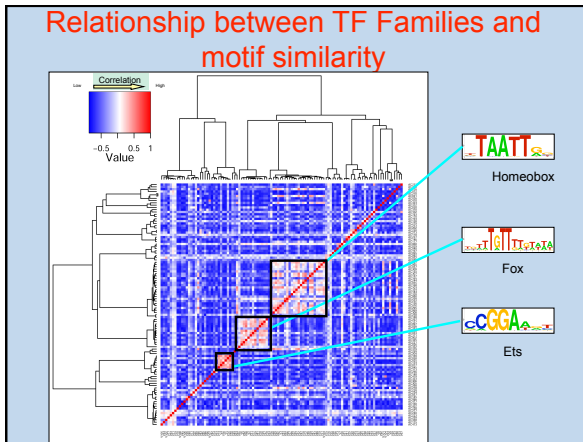






### Selex screen summary

100%		358 Core TFs + 147 possible TFs detected in the Genome
89%	89%	458 cloned DNA-BD
53%	62%	273 recombinant soluble DNA-BDs
19%	36%	103 Motifs for 103 proteins



- ### Suggested further reading
- **Ascidians:** [Ascidians and the plasticity of the chordate developmental program](#), Lemaire P, Smith WC, Nishida H. *Curr Biol*. 2008 Jul 22;18(14):R620-31. Review.
  - **Gene regulatory networks and evolution**
    - [The evolution of hierarchical gene regulatory networks](#), Erwin DH, Davidson EH. *Nat Rev Genet*. 2009 Feb;10(2):141-8. Epub 2009 Jan 13. Review.
    - [Gene regulatory networks and the evolution of animal body plans](#), Davidson EH, Erwin DH. *Science*. 2006 Feb 10;311(5762):796-800. Review.
  - **Transcriptional regulation**
    - [Transcription regulation and animal diversity](#), Levine M, Tjian R. *Nature*. 2003 Jul 10;424(6945):147-51. Review.
    - [Insights from genomic profiling of transcription factors](#), Farnham PJ. *Nat Rev Genet*. 2009 Sep;10(9):605-16. Epub 2009 Aug 11. Review.
    - [From DNA sequence to transcriptional behaviour: a quantitative approach](#), Segal E, Widom J. *Nat Rev Genet*. 2009 Jul;10(7):443-56. Review.