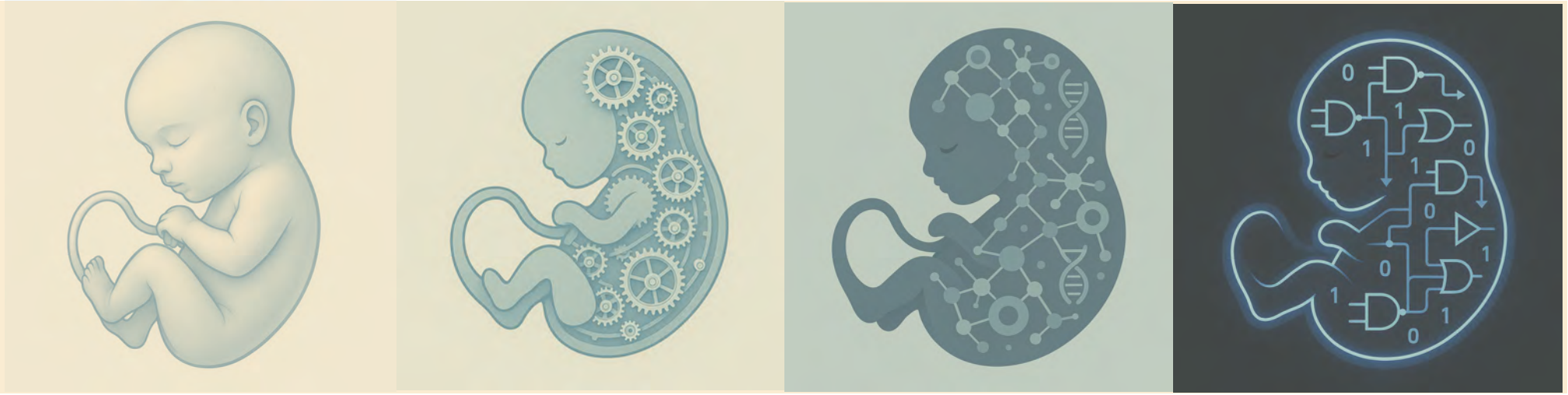
What is biological information? (II)



Course 2: Complexity and information during development

Thomas Lecuit

chaire: Dynamiques du vivant

COLLÈGE DE FRANCE
1530

Summary of previous course

Information provides a language to address the *meaning* and *logic in living systems*

There is a need to account for:

- Function
- Logic

Biological processes can be usefully described as **information processing**: they involve the manipulation of symbolic representations according to formal rules to achieve a functional outcome.

This perspective shifts the focus from the specific chemical substrates (e.g., DNA, proteins) to the logical and algorithmic principles governing the system.

Summary of previous course

A framework to disentangle:

- *Purpose*(why): *computation*
- *Strategy* (how): *algorithm*
- *Biology/physics* (what): *implementation*

Unity of algorithmic level for different implementations
Diversity of algorithmic level representations and solutions for a given function/computational problem

VISION

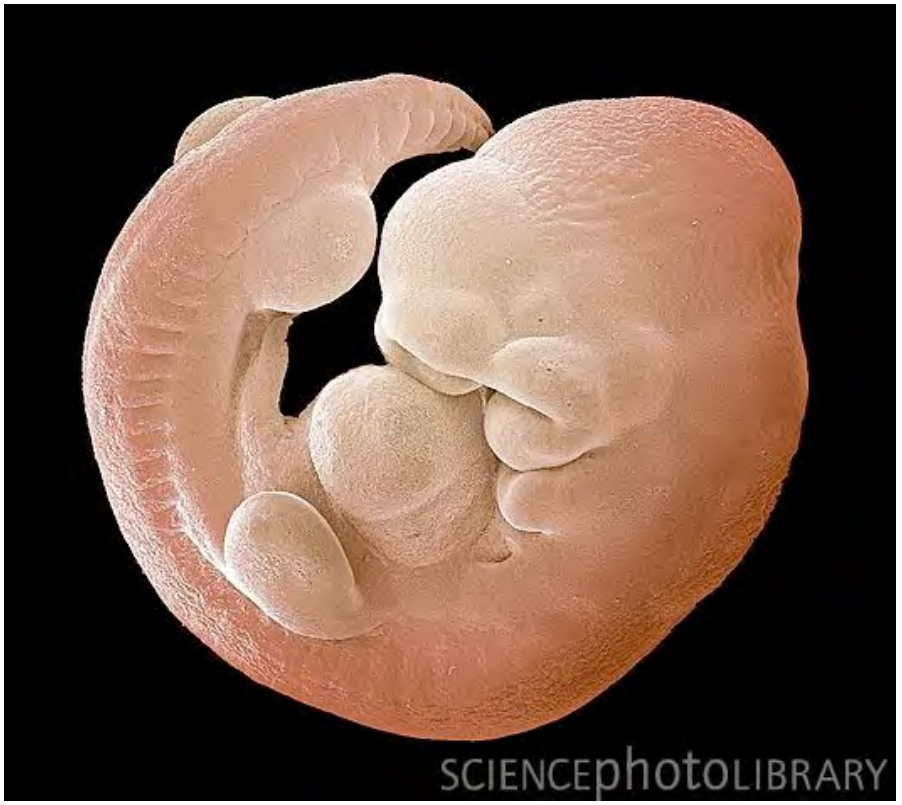
A Computational Investigation into the Human Representation and Processing of Visual Information

David Marr (1945-1980)

1982, Vision, David Marr
W. H. Freeman and Company
2010: *MIT press* (re-published)

Paradox of complexity and information during development

- Complexity increases
- Information cannot increase



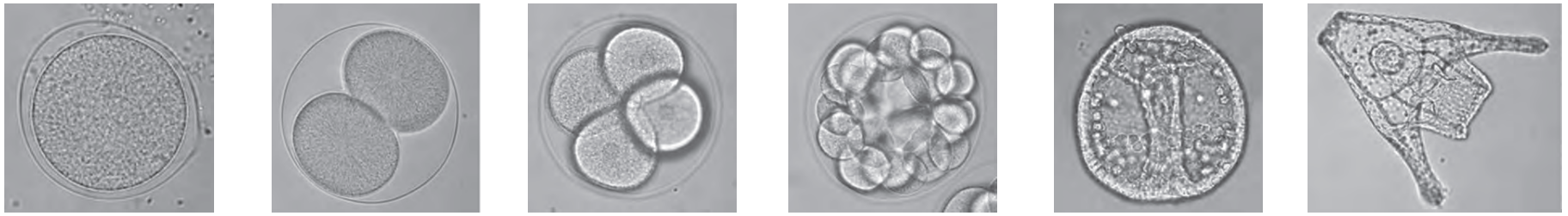
Plan of course

1. Complexity increases during development
2. How to measure complexity?

Complexity increases during development

Qualitative, descriptive approach to complexity

- Organisation/Shape complexity
  - Cell number
  - Cell shape
  - Tissue number
  - Tissue shape
  - Organ shape
- State complexity
  - Gene expression
  - Cell type
  - Interaction/network

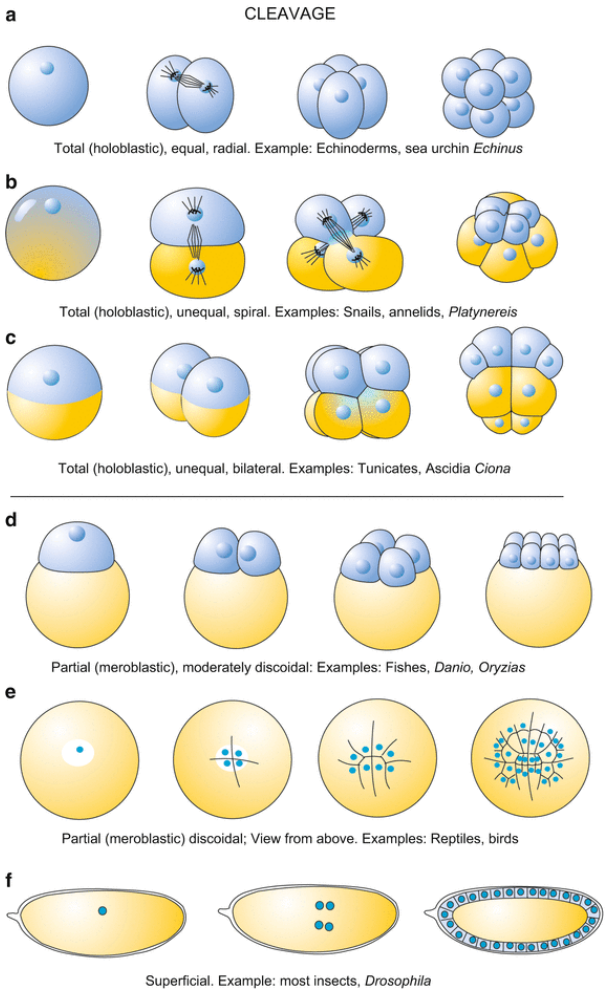


Cell number | Cavity (inside/outside) | Shape

Complexity increases during development

Cell number

Sea Urchin Embryo



Complexity increases during development

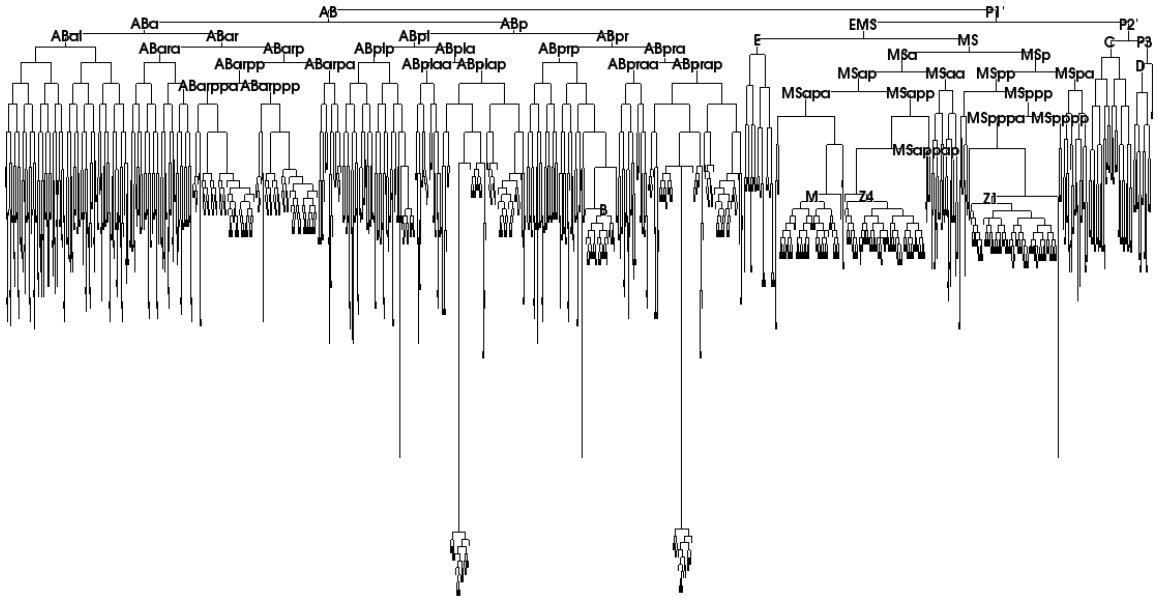
Cell number: 959-1033

Number of cell division ~10 (2^{10} ~1000)

L ~50µm

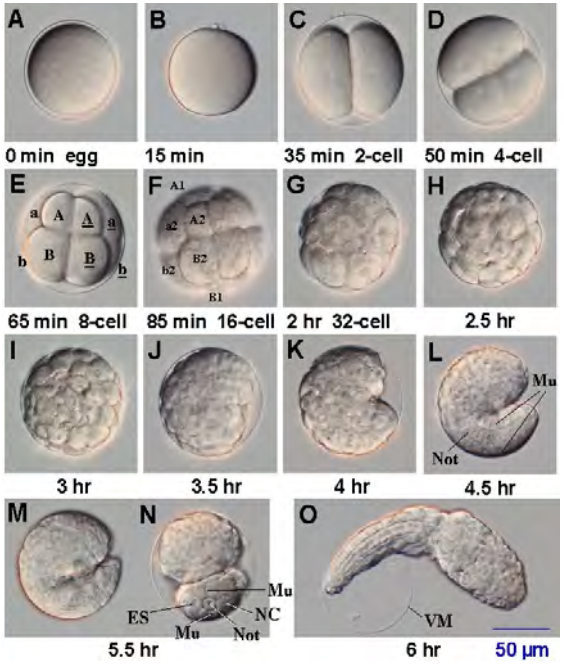
Length L ~1 mm

Lineage in *C. elegans*

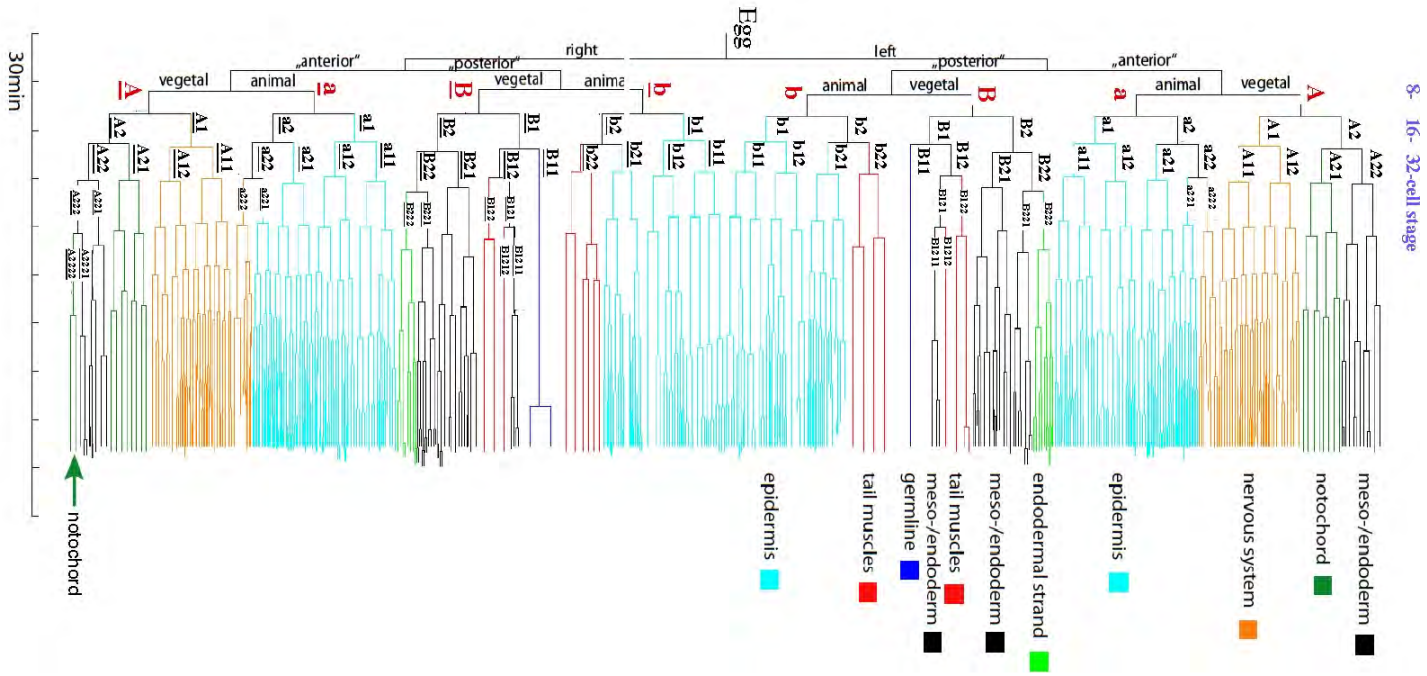


Complexity increases during development

Cell number increases via cleavage division

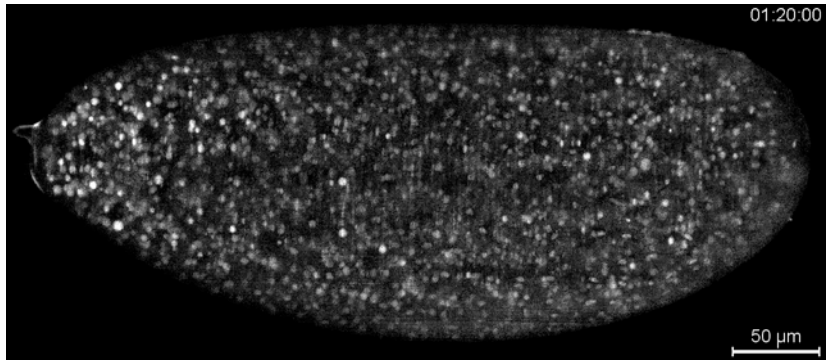


Oikopleura dioica (Tunicate, Urochordate)

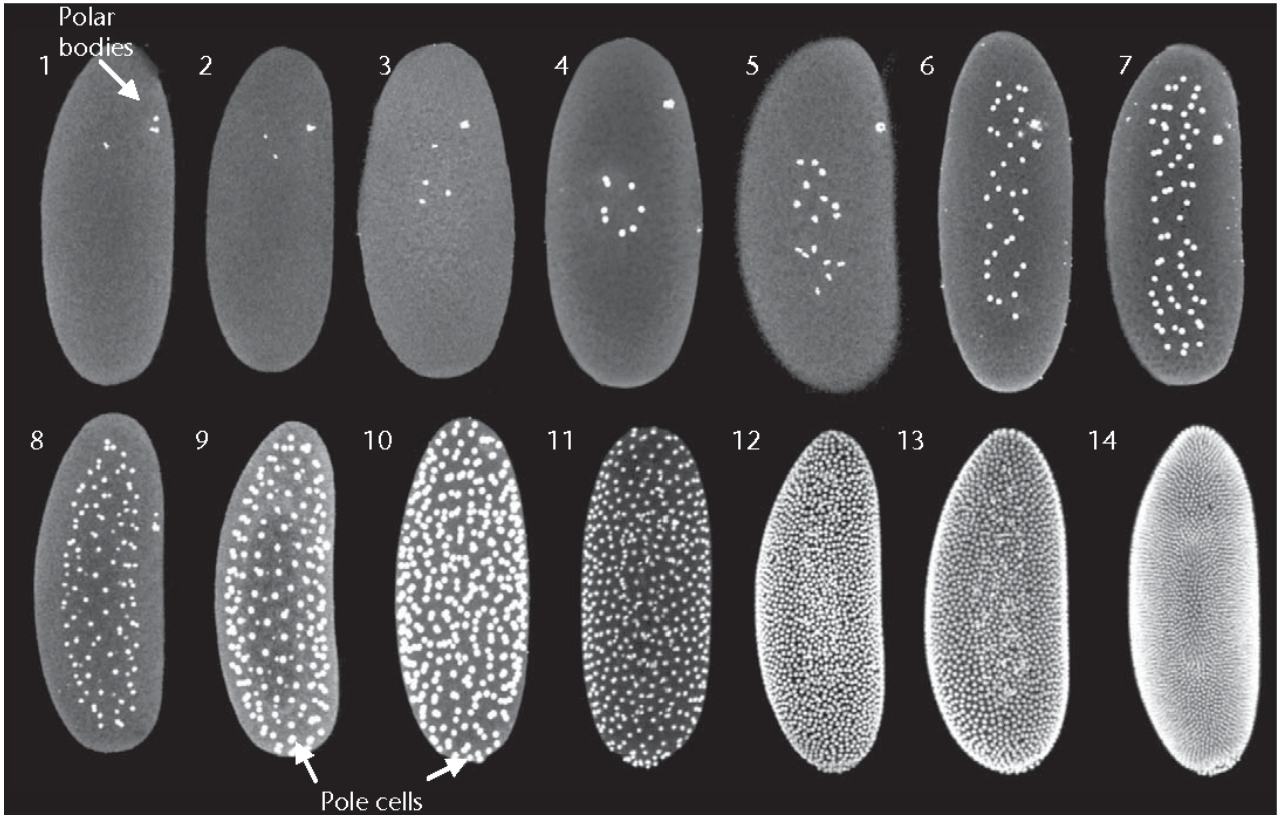


Complexity increases during development

Cell number: 6000 cells



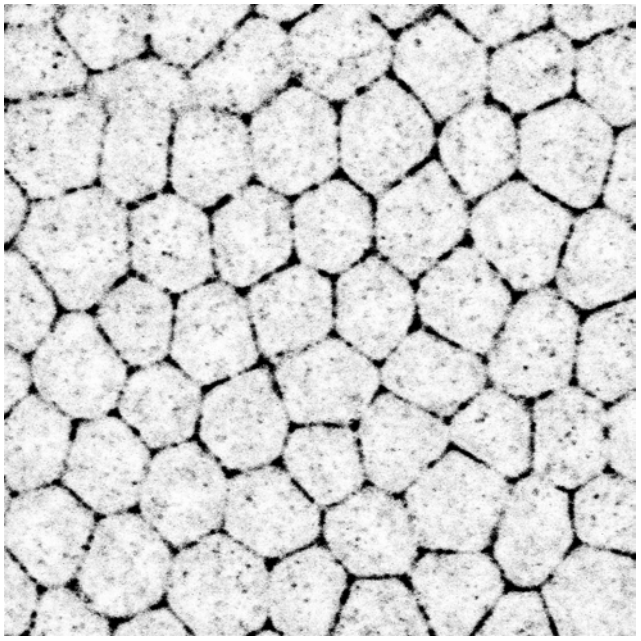
Drosophila early embryogenesis



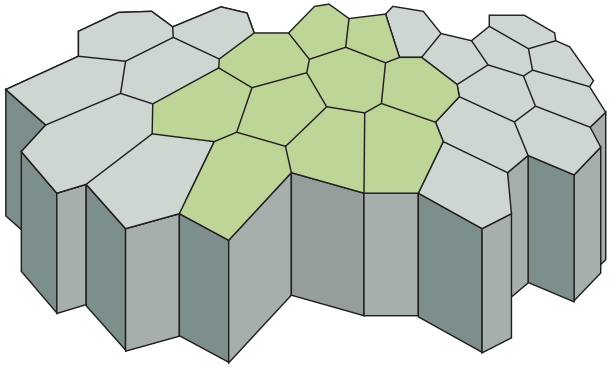
Complexity increases during development

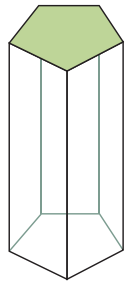
Cell shapes: regular shapes and trajectories

- Relatively stable shapes
- Local interactions

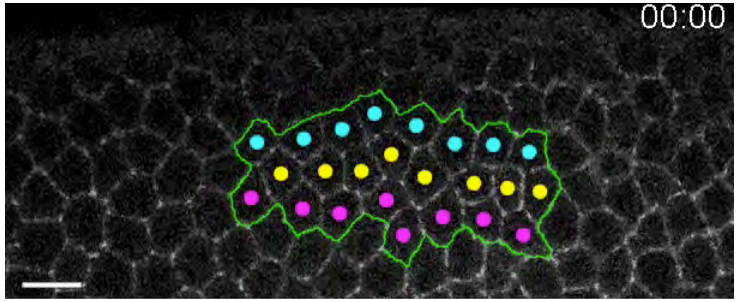


Drosophila Epithelial Cells





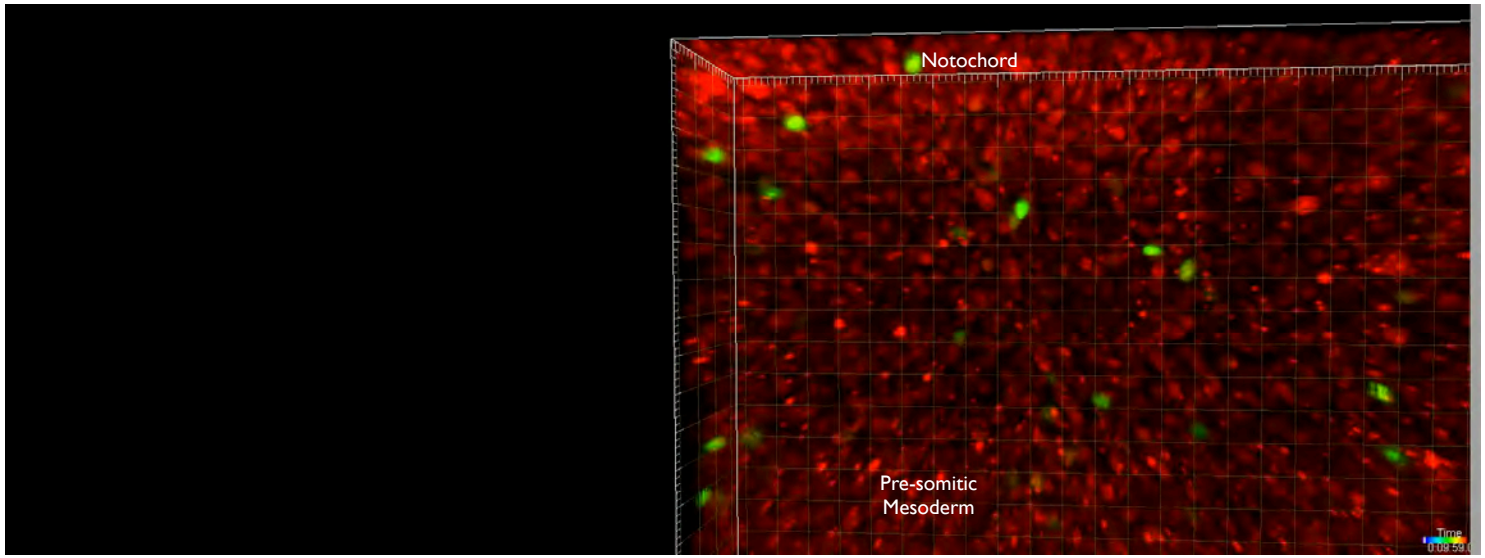
prism



Complexity increases during development

Cell shapes: Complex morphologies and trajectories

- rapid changes in cell shape and position

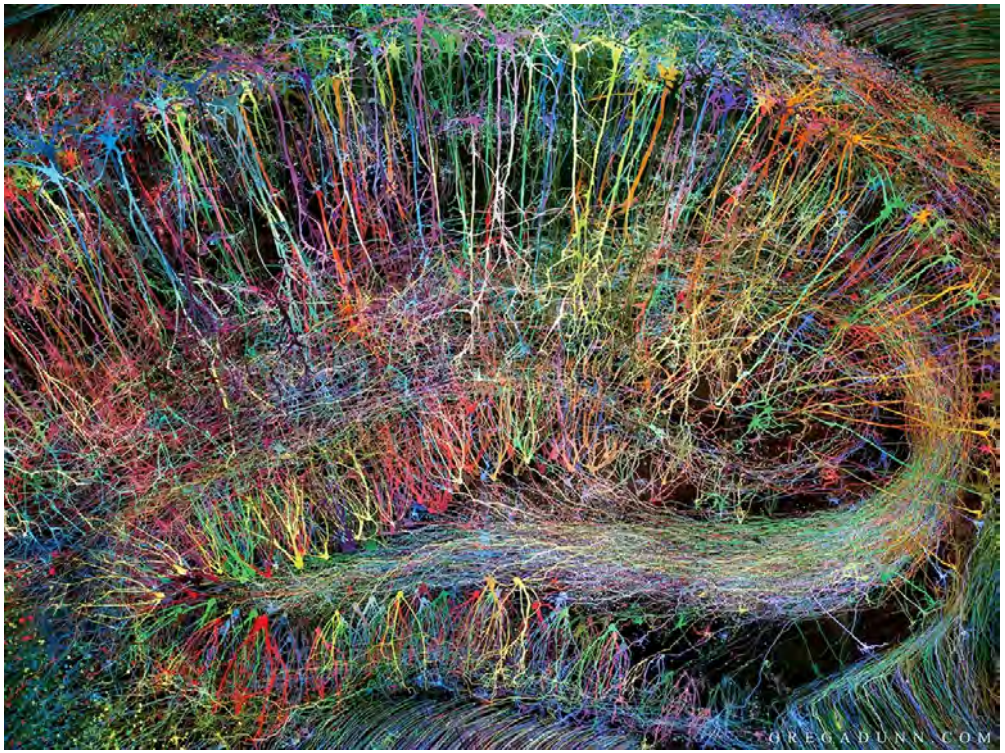


- Mesenchymal cells in presomitic mesoderm

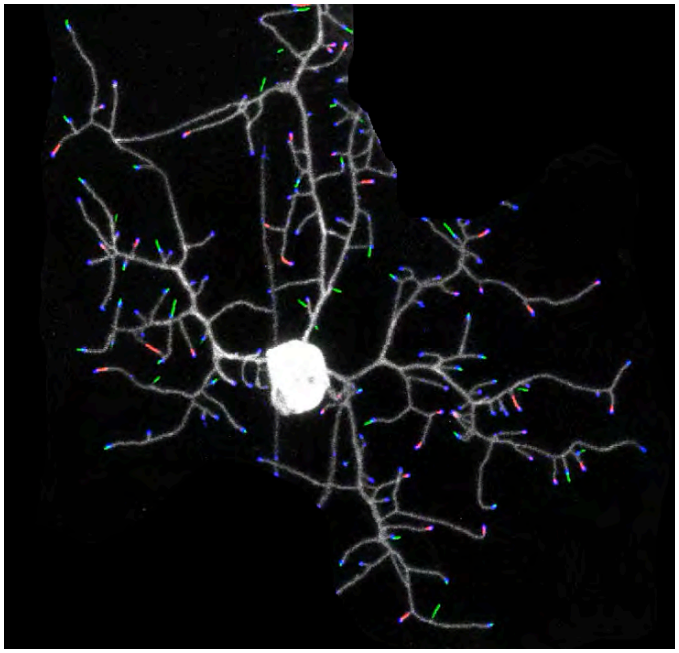
Bénazéraf B. et al, Pourquie O. *Nature*. 466:248. 2010

Complexity increases during development

Cell shapes: Complex morphologies



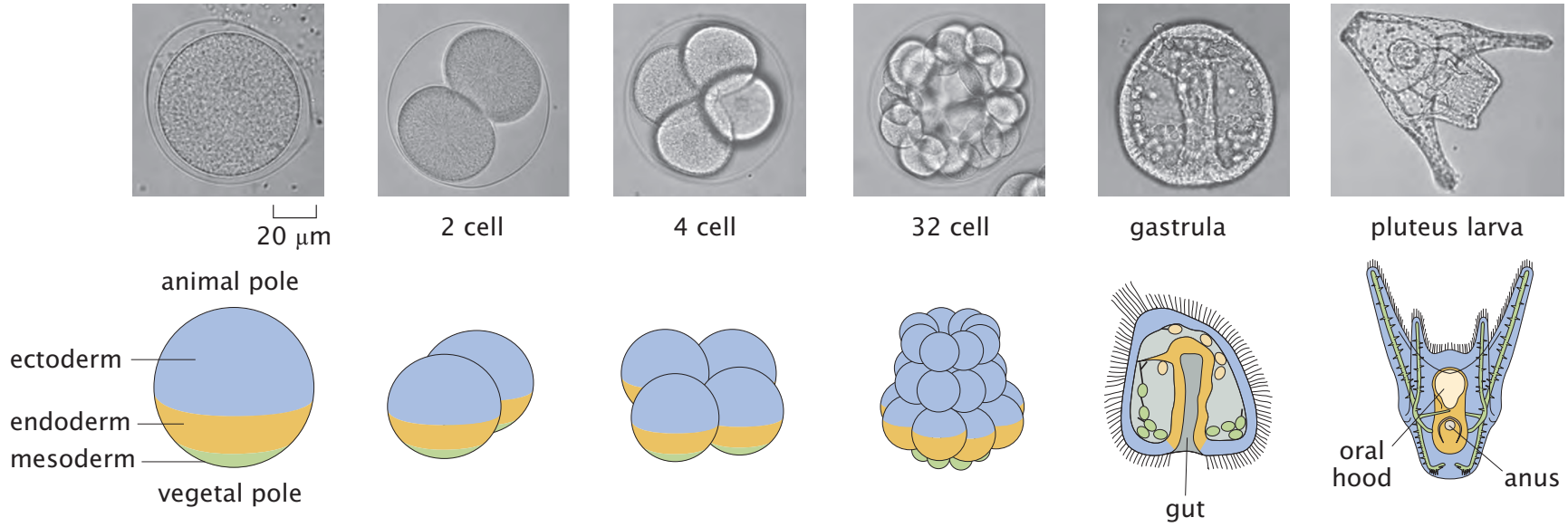
- Very rapid changes in cell shape



Complexity increases during development

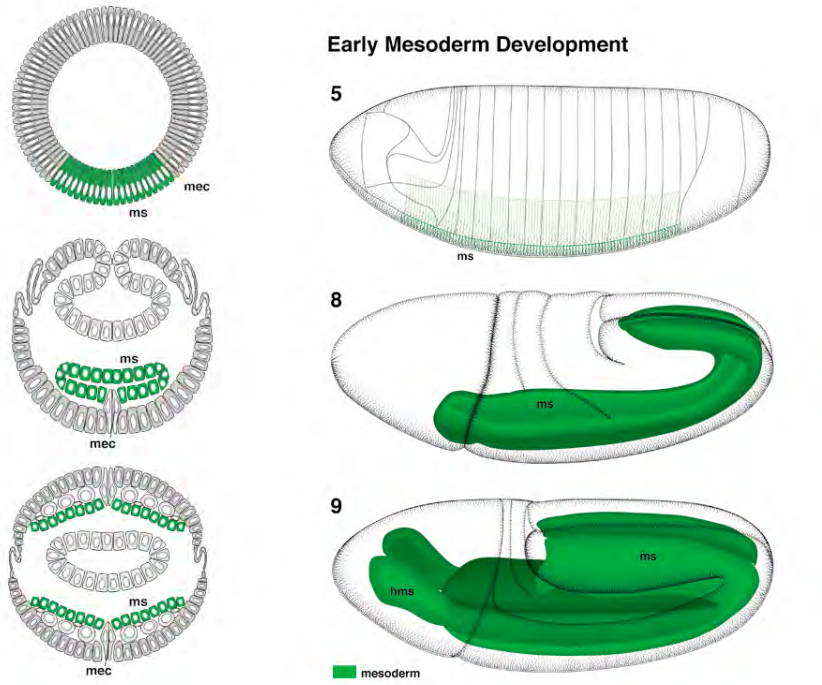
Tissue types

- *First*: Stem cells
- *Second*: Tissue layers (ectoderm, mesoderm, endoderm)
- *Third*: Tissue derivatives: skin, nervous system, lung, muscle, heart, bones etc

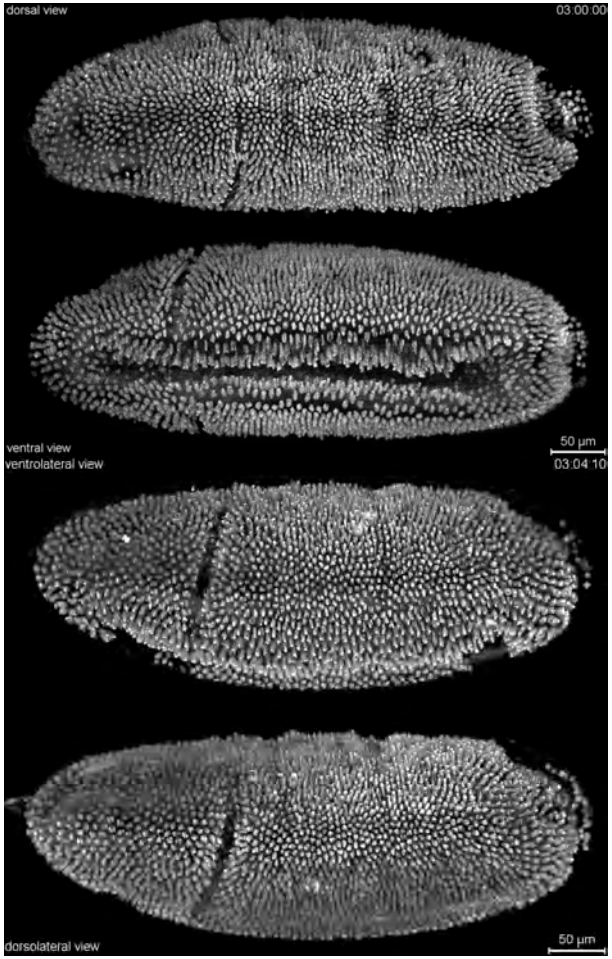


Complexity increases during development

Tissue shapes: Gastrulation (inside/outside)



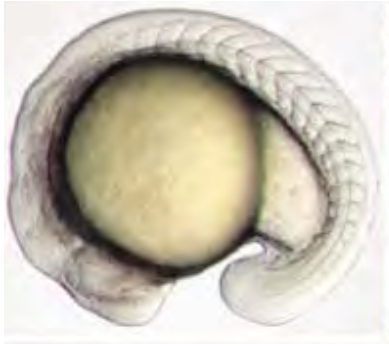
Atlas of Drosophila development, V. Hartenstein (CSHLP 1993)



Philpp Keller, Janelia Research Campus

Complexity increases during development

Tissue shapes



Zebrafish Embryo

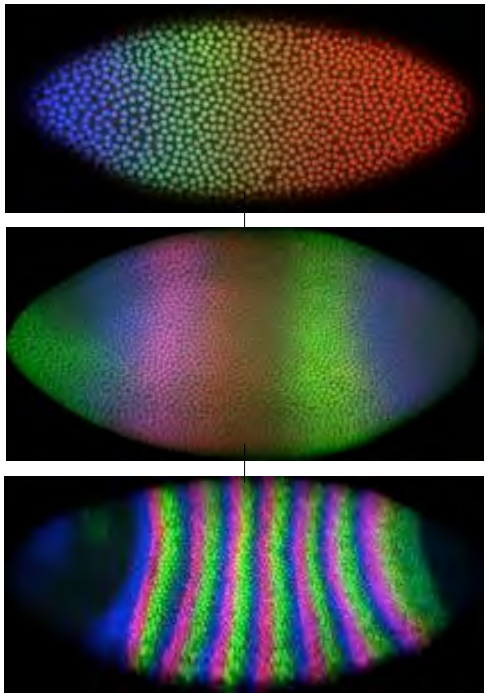
Complexity increases during development

Cell state: Gene expression

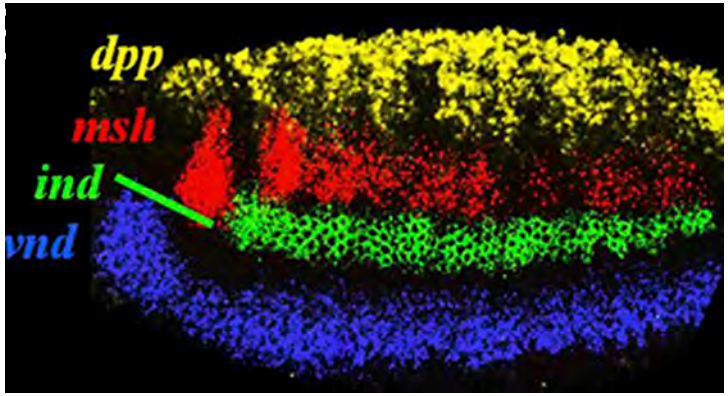
- Diversity of spatial patterns

- In order of N~20.000 genes
- Each cell as an N-dimensional vector $\mathbf{X}(g_1,\ldots, g_i,\ldots g_N)$

A. Boettiger
2009 Nikon Photomicrography competition



D. Kosman, S. Small & J. Reinitz
Dev Genes Evol (1998) 208:290–294

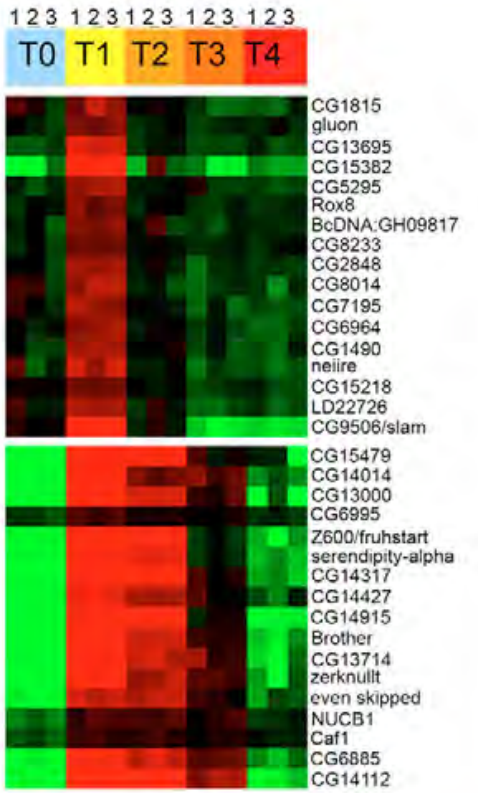


CM. Mizutani et al & E. Bier. *PLoS Biol* 4(10): e313.
https://doi.org/10.1371/journal.pbio.0040313

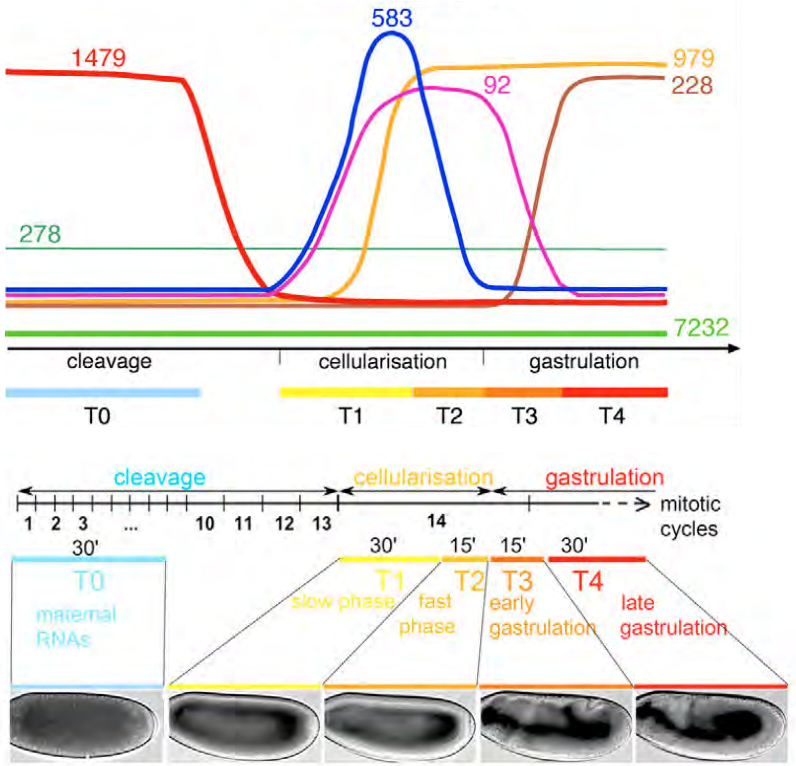
Complexity increases during development

Cell state: Gene expression

- Diversity of temporal patterns



- In order of N~20.000 genes
- Each cell as an N-dimensional vector $\mathbf{X}(g_1,\ldots, g_i,\ldots g_N)$

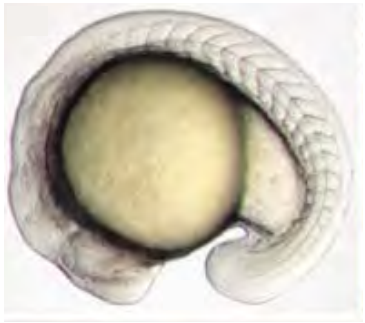


F. Pilot et al, T. Lecuit. *Development* 133, 711-723 doi:10.1242/dev.02251

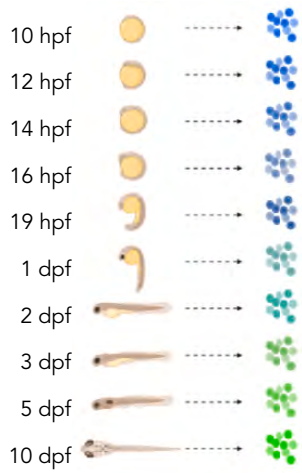
Complexity increases during development

Cell state: Gene expression

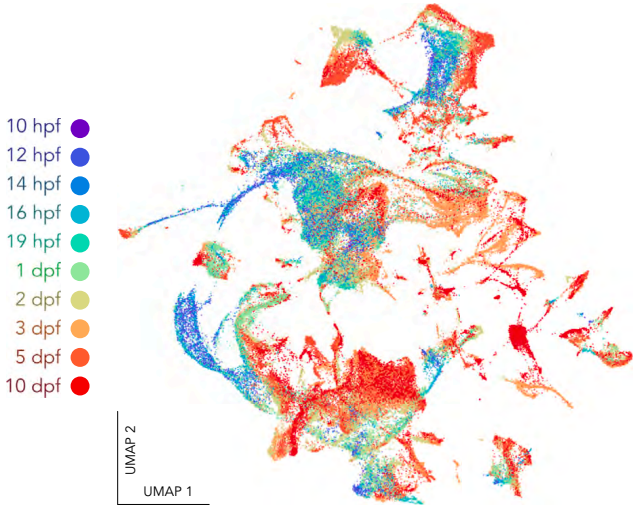
- In order of N~20.000 genes
- Each cell as N-dimensional vector $\mathbf{X}(g_1,\ldots, g_i,\ldots g_N)$
- Diversity of spatial & temporal patterns



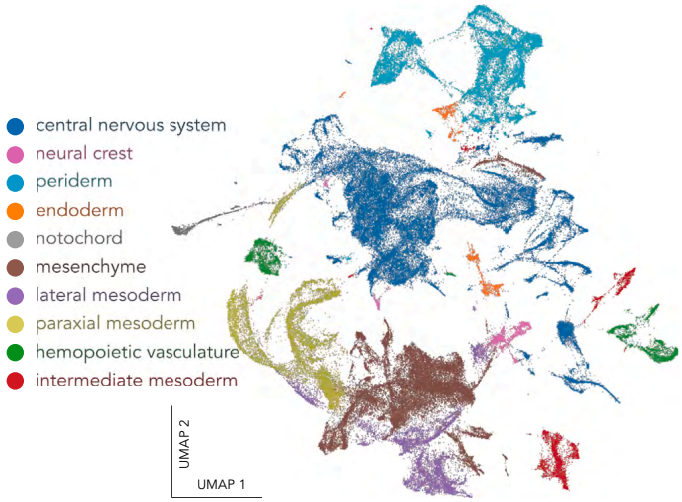
single-embryo dissociation



- scRNAseq time course (UMAP)



- Cluster annotation



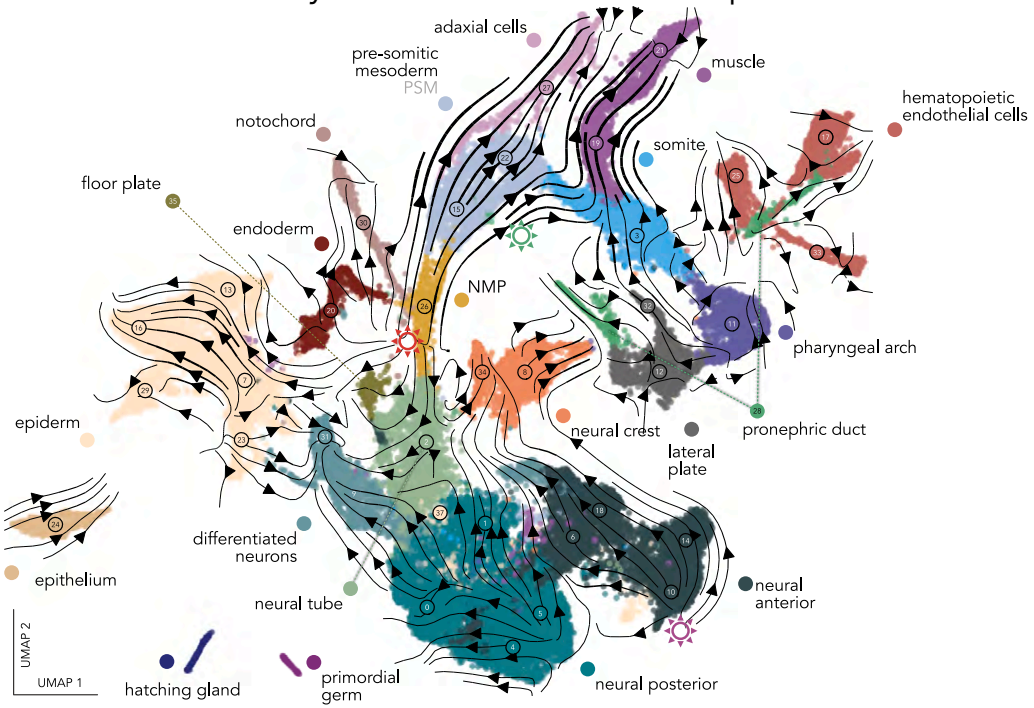
Merlin Lange et al. and Loic Royer, 2024, *Cell* 187, 6742–6759

Complexity increases during development

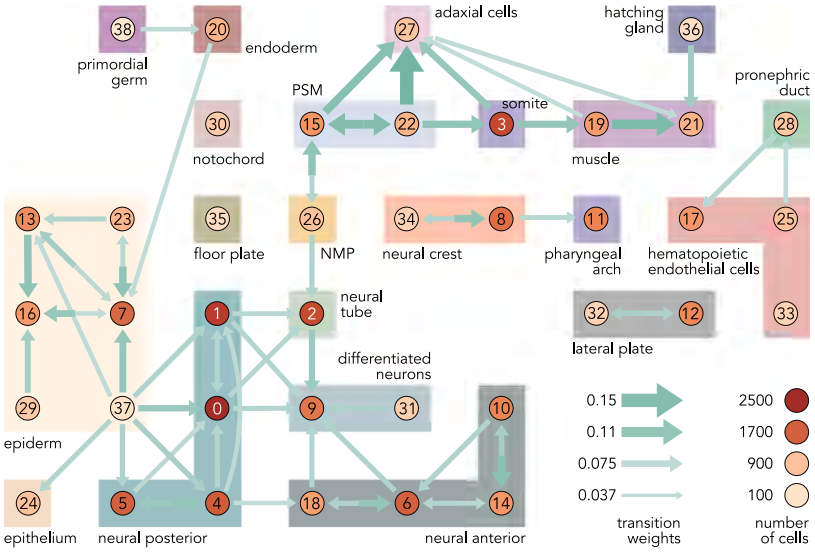
Cell state: Gene expression

- In order of N~20.000 genes
- Each cell as N-dimensional vector $\mathbf{X}(g_1,\ldots, g_i,\ldots g_N)$

- RNA velocity vector field on scRNAseq UMAP



- Cell state transition graph



Merlin Lange et al. and Loic Royer, 2024, *Cell* 187, 6742–6759

Complexity increases during development

Gene expression networks

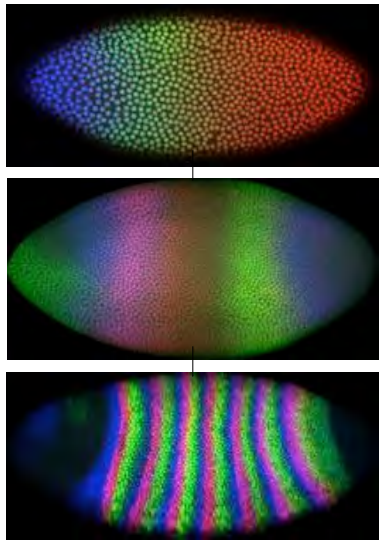
- Cells form interaction networks and compute decision making processes

Antero-Posterior patterning

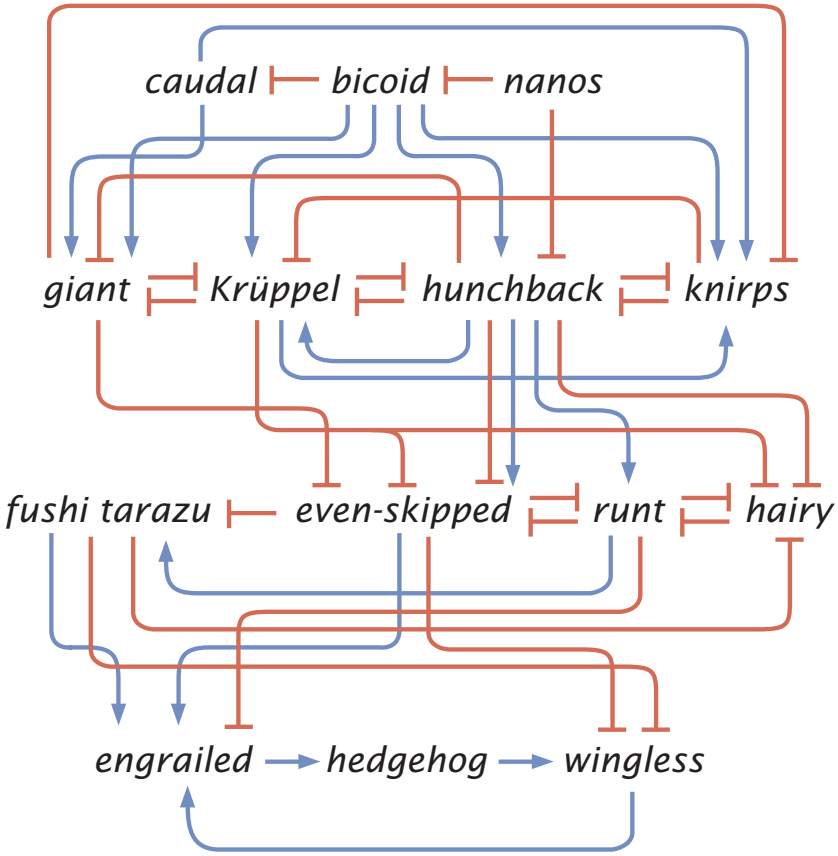
Maternal genes (3)

Gap genes (4)

Pair rule genes (7)



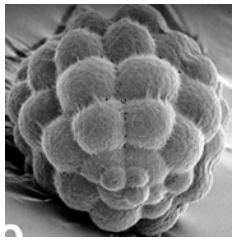
D. Kosman, S. Small & J.

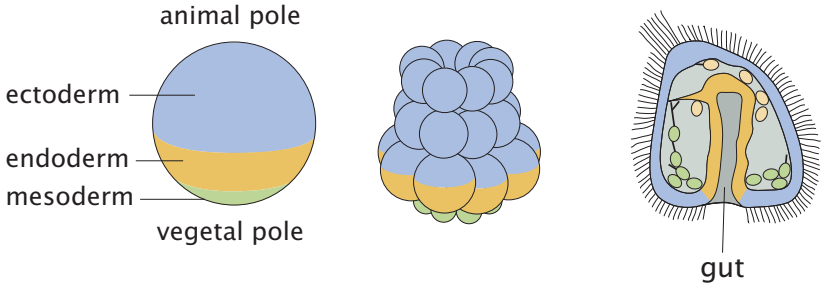


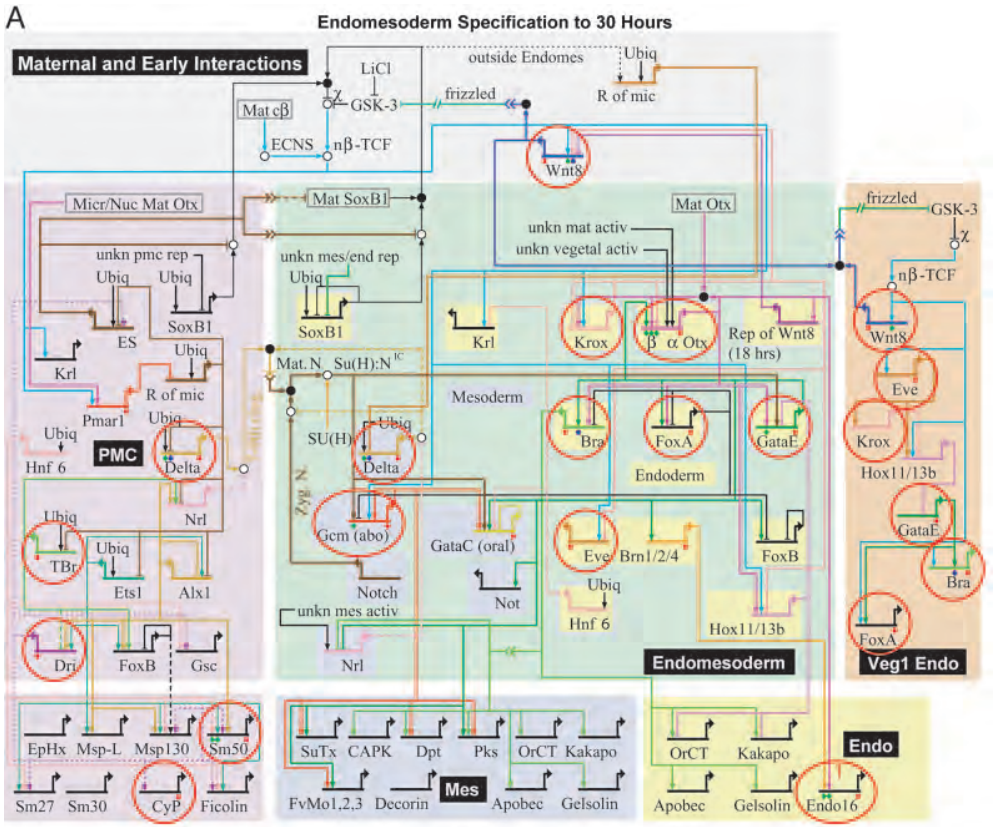
Complexity increases during development

Gene expression networks

- Cells form interaction networks and compute decision making processes





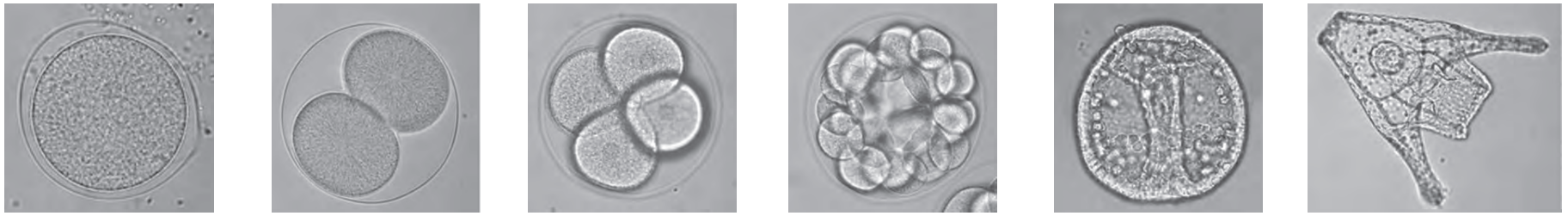


L. Bodenstein. *Mechanisms of Development*, 162 (2020)
https://doi.org/10.1016/j.mod.2020.103606

Complexity increases during development

Qualitative, descriptive approach to complexity

- Organisation/Shape complexity
  - Cell number
  - Cell shape
  - Tissue number
  - Tissue shape
  - Organ shape
- State complexity
  - Gene expression
  - Cell type
  - Interaction/network



Cell number | Cavity (inside/outside) | Shape

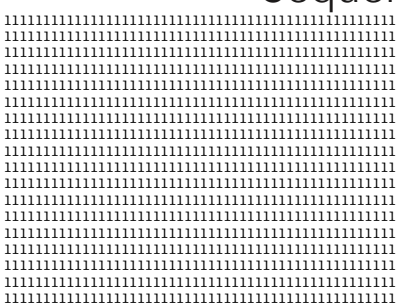
Plan of course

1. Complexity increases during development
2. How to measure complexity?

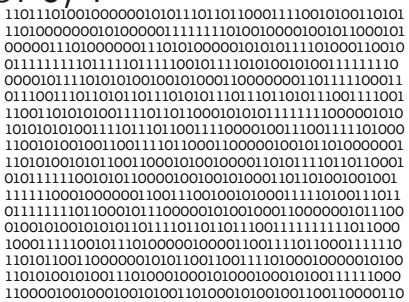
How to measure complexity?

Descriptive complexity: Statistical complexity

Sequence of 0, 1

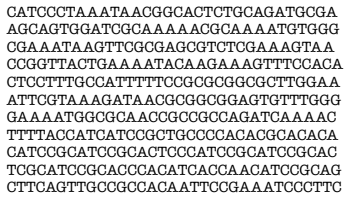


uniform



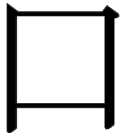
random

Nucleotide sequence



bicoid 1-352

Kanji: number of strokes

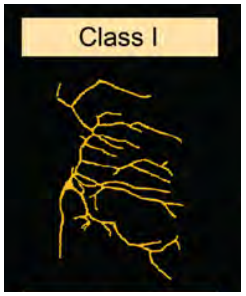
mouth 

4

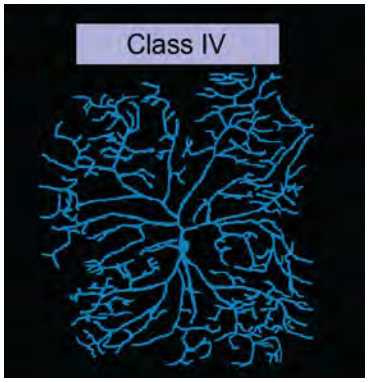
84

daito
« A cloudy dragon in flight »

Neurons: number of branches



~20



>100

How to measure complexity?

Descriptive complexity: How concisely can the data be described?

- **Intuition:** complexity is low for a perfectly ordered or uniform pattern, and maximal for a random pattern. And biological complexity lies in between (organisation, ordering yet not a cristal).
- **Kolmogorov complexity/Algorithmic Information content.**
  - The algorithmic information content (AIC) also referred to as Kolmogorov complexity $KU(s)$ of a string of symbols s is the length of the shortest program for a universal Turing machine (ie. digital computer) U to produce the output s and then halt.

Random number X

10101101111110110110011111011110111110110101110111
10110011001101011100011101111100111100110111101010

$KU(X)$ is high

π

11001001000011111101101010100010001000010110100011
00001000110100110001001100011001100010100010111000

$KU(X)$ is low

M. Gell-Mann, S. Lloyd. Complexity, 1996
https://doi.org/10.1002/(SICI)1099-0526(199609/10)2:1<44::AID-CPLX10>3.0.CO;2-X

How to measure complexity?

Descriptive complexity: How concisely can the data be described?

- **Kolmogorov complexity/Algorithmic Information content.**
  - The algorithmic information content (AIC) also referred to as Kolmogorov complexity $\boldsymbol{KU(s)}$ of a string of symbols s is the length of the shortest program for a universal Turing machine (ie. digital computer) $\boldsymbol{U}$ to produce the output s and then halt. Ex: π has low $\boldsymbol{KU}$.

    M. Gell-Mann, S. Lloyd. Complexity, 1996
    https://doi.org/10.1002/(SICI)1099-0526(199609/10)2:1<44::AID-CPLX10>3.0.CO;2-X
  - **AIC/***KU(s)* is uncomputable.
  - **Yet AIC/***KU(s)* **is the theoretical limit of lossless compression.** The shortest program is the perfectly compressed version of the string of symbols (eg. sequence). So the size of the compressed string *C(s)* is a good approximation of its AIC ($KU(s)$). For instance, size of compressed genome sequence.

  - Use in phylogenomics to build tree based on sequences.

Normalized compression distance (NCD) computed from length of compressed data files.

C(x) = compressed size of the DNA sequence of species X.
C(y) = compressed size of the DNA sequence of species Y.
C(xy) = compressed size of the *concatenated* sequences of X and Y.

$$\mathrm{NCD}(x,y) = \frac{C(xy) - \min\{C(x), C(y)\}}{\max\{C(x), C(y)\}}.$$

Rudi Cilibrasi and Paul M. B. Vitányi, IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. 51, NO. 4, APRIL 2005
Clustering by compression

Complexity of neurons and Algorithmic Information

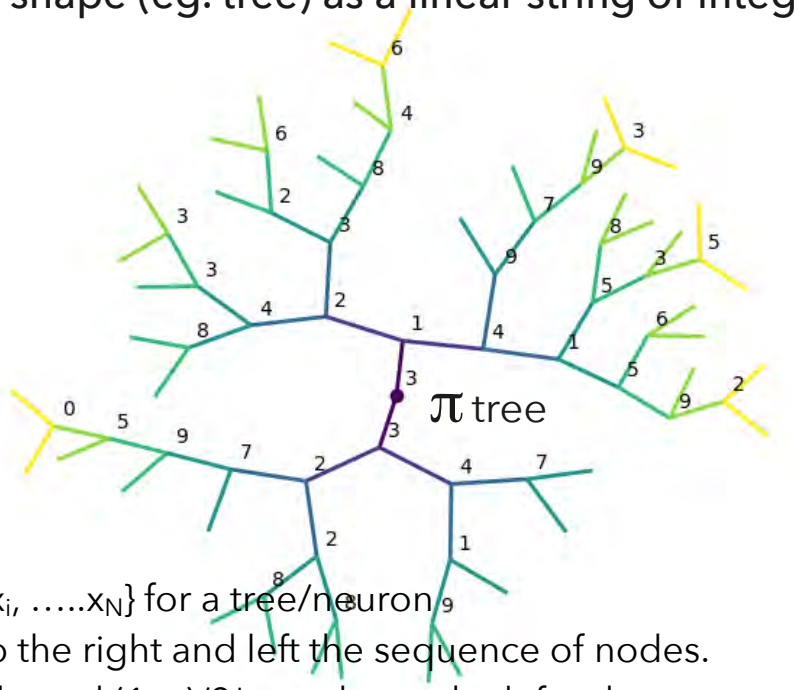
Representation of a shape (eg. tree) as a linear string of integers

Algorithmic description of tree

Asymmetry at branching points

{0,1,2,3,4,5,6,7,8,9}
0: symmetric
9: asymmetric

Encoding of π tree: {31415926538….}



Marc-Eric Perrin (IBDM)

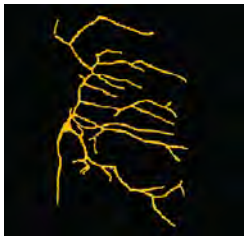
Algorithm: We have a sequence $S_N = \{x_0, x_1, \ldots, x_i, \ldots . x_N\}$ for a tree/neuron

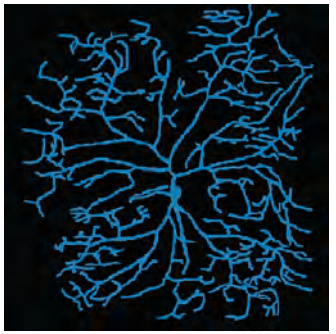
- For each node *i*, calculate $a_i = x_i/9$ and split to the right and left the sequence of nodes.
- Put $(1+a_i)/2*n$ nodes in the subtree to the right and $(1-a_i)/2*n$ nodes to the left subtree.
  - Ex: if $x_i=0$, $a=0$ and $(1+/-a_i)/2*n=n/2$, the subtrees are symmetric.
  - If $x_i=9$, $a=1$ and $(1+ a_i)/2*n=n$ and 0 to the left, the subtrees are maximally asymmetric.
- 2 new sequences are generated and distributed to L and R: $S_R = \{x_0, x, \ldots, x_{((1+ai)/2*n)}\}$ and $S_L = \{x_{((1+ai)/2*n+1)}, \ldots, x_N\}$.
- Repeat at each next node this iterative process.

Complexity of neurons and Algorithmic Information

Description (statistics)
e.g. number of branch points

low

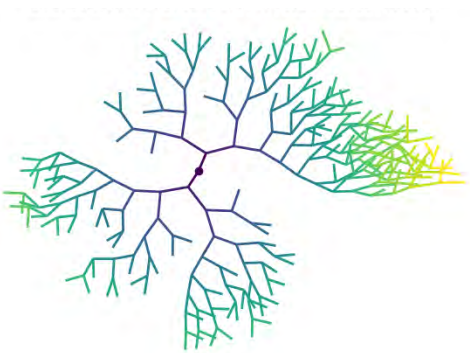


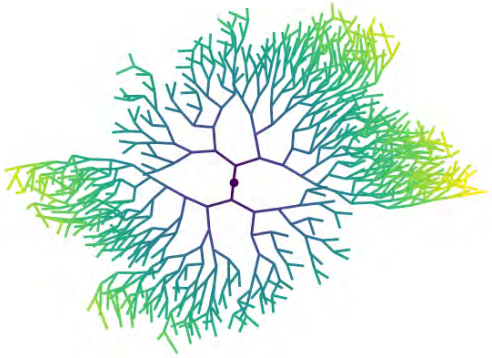


high

Algorithmic information

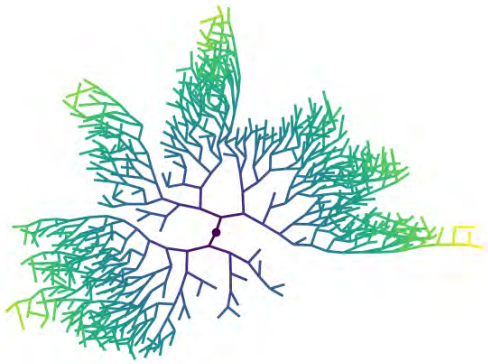
High AIC

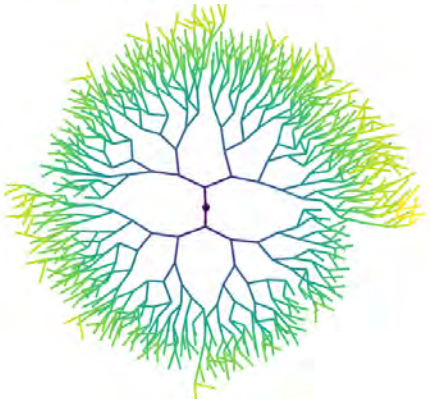




Low AIC

sequence length 400
compressed sequence length 191
'3141592653589793238462643383279502884197...'





sequence length 400
compressed sequence length: 10
'1111111111111...1'

How to measure complexity?

Generative complexity: How much computational effort is needed?

(to create the data from its concise description)

- **Logical depth:** The run time of the minimal algorithm to compute/generate x
  - A shallow structure: perfectly ordered or random (a print function generates this quickly)
  - A deep structure: requires long computation (Ex: π, though it depends on the algorithm)

On the Nature and Origin of Complexity in Discrete, Homogeneous, Locally-Interacting Systems

Charles H. Bennett[1]

Received March 12, 1986

The observed complexity of nature is often attributed to an intrinsic propensity of matter to self-organize under certain (e.g., dissipative) conditions. In order better to understand and test this vague thesis, we define complexity as "logical depth," a notion based on algorithmic information and computational time complexity. Informally, logical depth is the number of steps in the deductive or causal path connecting a thing with its plausible origin. We then assess the effects of dissipation, noise, and spatial and other symmetries of the initial conditions and equations of motion on the asymptotic complexity-generating abilities of statistical–mechanical model systems. We concentrate on discrete, spatially-homogeneous, locally-interacting systems such as kinetic Ising models and cellular automata.

Logical depth has the right intuitive properties for a complexity measure: subjectively trivial structures (such as a string of zeros, or a random string of zeros and ones produced by coin tossing) are shallow; on the other hand, deep structures can be produced slowly, though not quickly, by a universal computer. It may at first seem surprising that random objects are shallow, but a typical random output x typically cannot be produced at all by a small program, and so a "print program," containing a *verbatim* description of the desired output, is fast-running and of near-minimal size. Armed with the notion of depth, we can say that a system self-organizes if it produces, with high probability, increasingly deep local configurations in the course of its evolution, configurations that could not have been produced quickly by any other evolution with a simple initial condition and simple laws of motion.

Descriptive and generative complexity

Descriptive complexity: *How concisely can the data be described? Algorithm*

Generative complexity: *How long is the computation?*

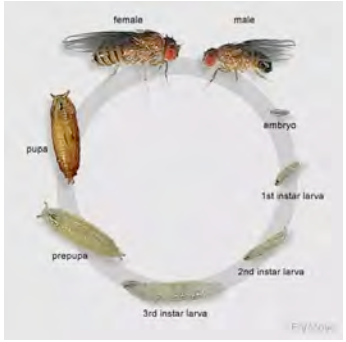
Application to Biological systems:

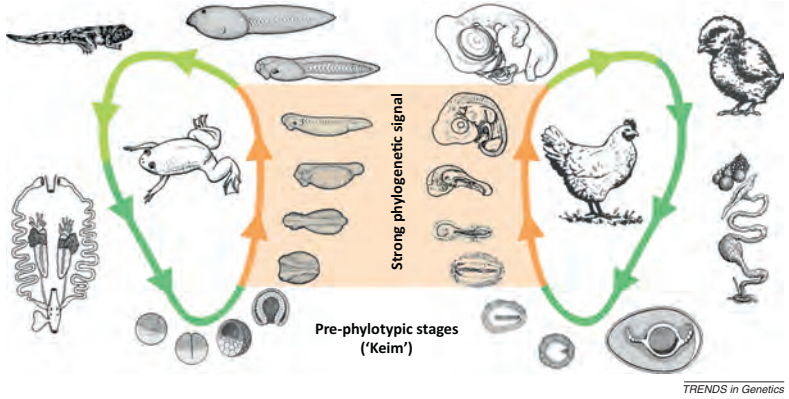
- Simple algorithmic level description capture key properties of a system (eg. Turing instabilities). Simplicity can be found.
- Yet, biological processes usually take a long time to unfold and reach their end point (eg. developmental processes, such as morphogenesis, cell differentiation, or even more evolution…).
- *Question*: is developmental speed constrained by computational complexity? Or by other processes such as energetics of process? Or both

few hours in *Drosophila*

~10 days in mouse

~ 5 weeks in human

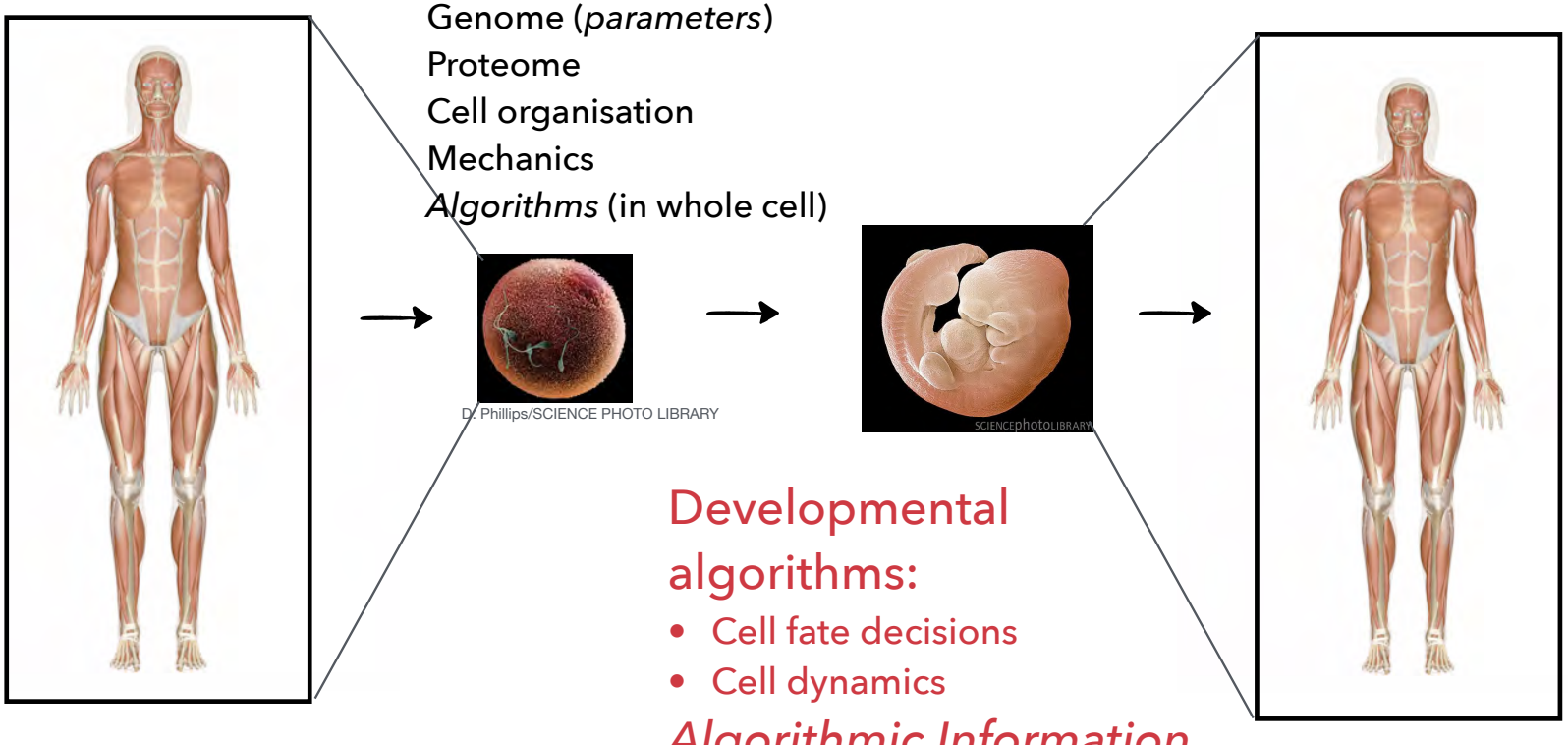




The egg as a« compressed information state » of the future organism

compression decompression

Genome (*parameters*)
Proteome
Cell organisation
Mechanics
Algorithms (in whole cell)



Developmental algorithms:

- Cell fate decisions
- Cell dynamics

Algorithmic Information

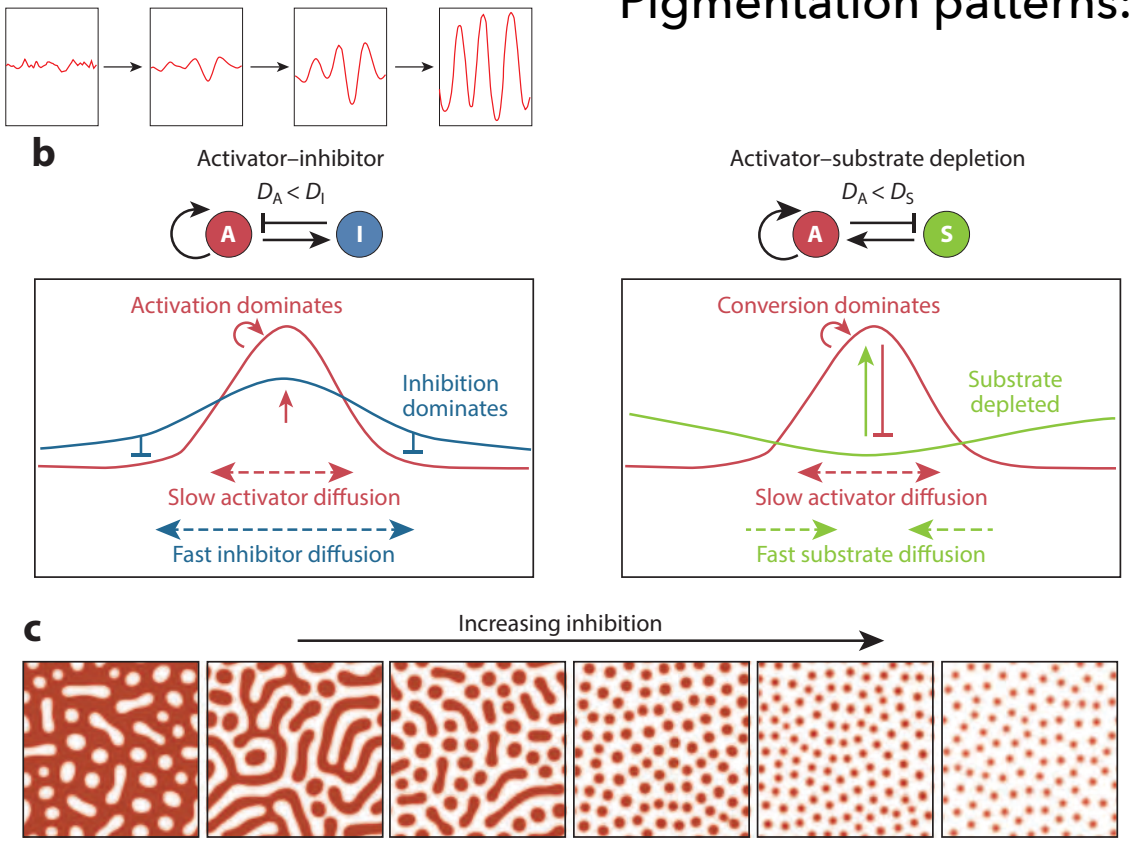
Developmental complexity - algorithmic view

How complex are these tissues?
Emergence of seemingly complex structures from simple generative rules

- Complex to describe
- But actually relatively simple to recapitulate (« simple algorithm » to generate the pattern or shape)

« Simple » emergence of complex patterns

Pigmentation patterns: Turing instability



Inelegant (ie. complex) or elegant (ie. simple) patterns

DROSOPHILA DEVELOPMENT

Making stripes inelegantly

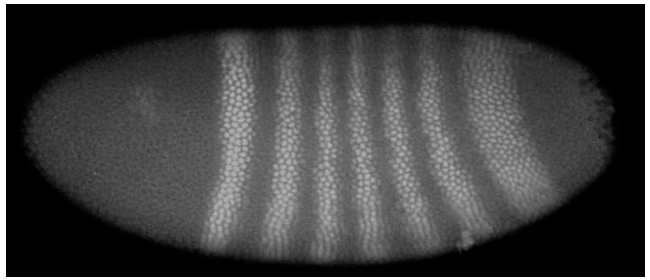
Michael Akam

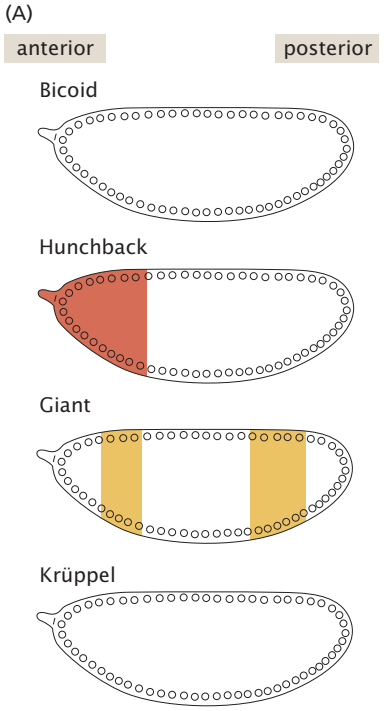
NATURE · VOL 341 · 28 SEPTEMBER 1989

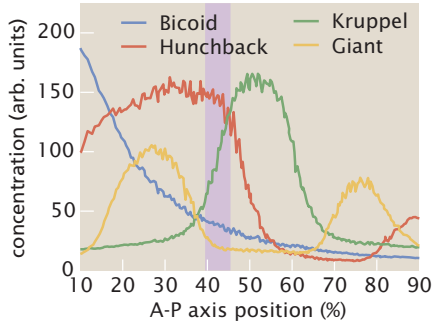
Periodicity might be generated in one of two ways. An elegant mechanism, favoured by model builders[3,4], would use an intrinsically periodic pattern-generating system, comprising the pair-rule genes and their products. This would only need to be triggered by local stimuli from the gap genes. Alternatively, unique instructions could be generated by the gap-gene proteins to define the position of each pair-rule stripe.

If each pair-rule stripe is generated by the interpretation of a unique instruction, then the apparent simplicity of the repeating segment pattern is deceptive. A 'specific instruction' mechanism could generate an arbitrarily complex set of uneven stripes. The simplicity of the natural pattern must then be attributed to a process of selection needed to produce a

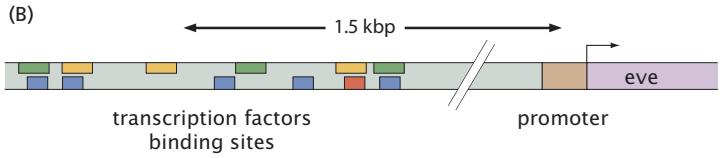
Eve protein







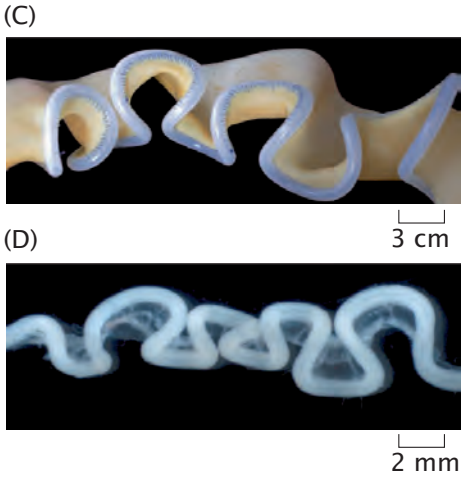
(D)

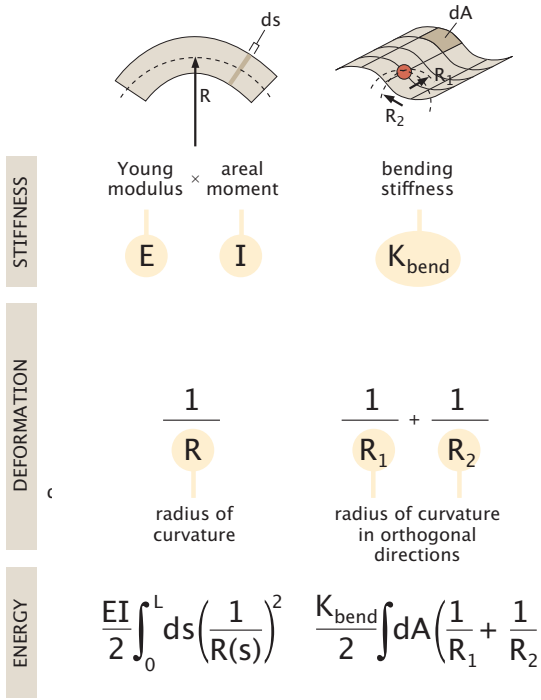


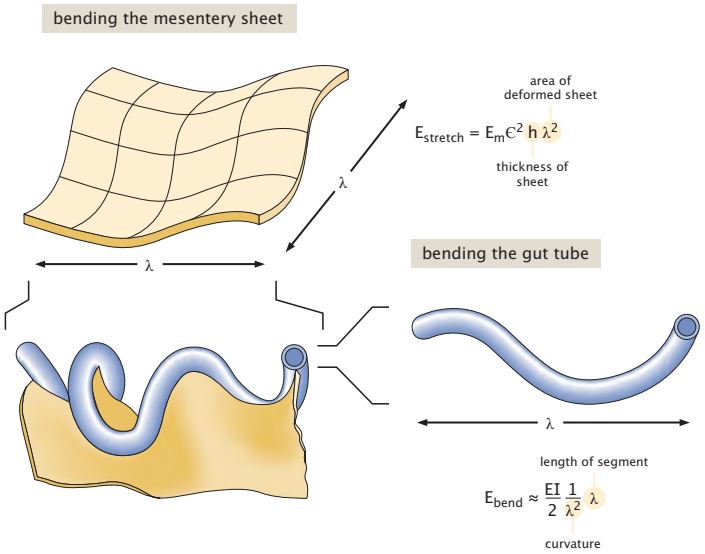
Rob Philipps, Jane Kondev, Julie Theriot, Hernan G. Garcia.
Physical Biological of the Cell (Garland Science)

« Simple » emergence of complex shapes

Gut looping: elasticity theory, buckling







from Rob Phillips (CalTech) and Christina Hueschen (Stanford Univ.)

Thierry Savin, et al, L. Mahadevan and Cliff Tabin. *Nature* (2011) 476:57-62.

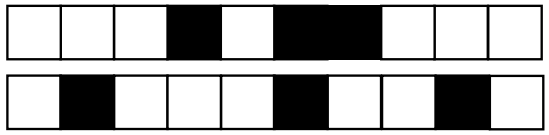
Cellular Automata and emergence of complexity from simple rules

- A Cellular Automaton (CA) is a discrete model used as a framework for **simulating complex systems using very simple rules.**

A CA is defined by four key components:

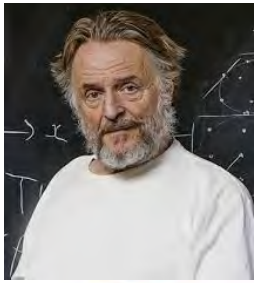
1. **The Grid:** A lattice of cells. This can be 1-dimensional (a line), 2D (a square grid, most common), or higher dimensions.
2. **States:** Each cell can be in one of a finite number of states (e.g., 0 or 1, dead or alive, on or off).
3. **Neighborhood:** The set of surrounding cells that influence the state of a given cell. The most common 2D neighborhood is the von Neumann neighborhood (up, down, left, right) and the Moore neighborhood (includes diagonals, 8 cells total).
4. **Rules:** A set of deterministic or probabilistic rules that define how each cell updates its state synchronously based on its current state and the states of its neighbors.

000	001	010	011	100	101	110	111
1	0	0	1	0	0	0	1



Cellular Automata and emergence of complexity from simple rules

- A Cellular Automaton (CA) is a discrete model studied used as a framework for **simulating complex systems using very simple rules.**



- **1940s - Origins:** The concept was pioneered by **John von Neumann** and **Stanisław Ulam** at Los Alamos National Laboratory. Ulam suggested using a discrete grid for calculations (hydrodynamics), and von Neumann wanted to model self-replication.

- **1970s - The Game of Life: John Conway**, a mathematician at Cambridge created the **"Game of Life"**, a 2D CA with very simple rules that give rise to very complex behaviors

- **1980s - Stephen Wolfram** conducted a systematic, long-term study of elementary (simple 1D) CAs. He classified them into four behavioral classes (fixed, periodic, chaotic, complex). His work laid the groundwork for CAs as models of complex systems and natural phenomena.

Cellular Automata and emergence of complexity from simple rules

Theory of
Self-Reproducing Automata

JOHN VON NEUMANN

edited and completed by Arthur W. Burks

University of Illinois Press
URBANA AND LONDON 1966

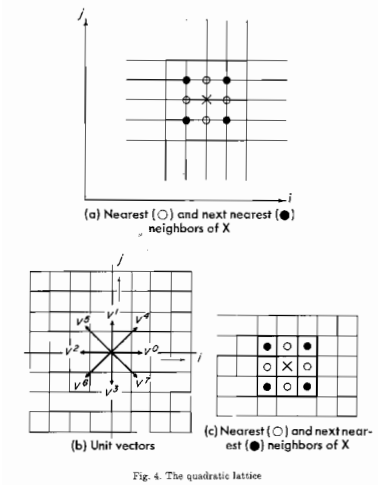


Fig. 4. The quadratic lattice

- 29-state cells and a complex rule, was the first CA to possess a self-replicating pattern

Chapter 2

A SYSTEM OF 29 STATES WITH A GENERAL TRANSITION RULE

2.1 Introduction

2.1.1 The model: states and the transition rule. In this chapter we will develop the first model that possesses the potentialities of logical and constructive universality and of self-reproduction (cf. questions (A)–(E) in Sec. 1.1.2.1), as well as the other attributes evolved in the course of the discussion of these in Chapter 1. This model is based on a crystalline medium (cf. Secs. 1.3.3.1–1.3.3.3); we will be able to construct it in two dimensions and to use there the quadratic[1] (regular) lattice {cf. the end of Sec. 1.3.3.3, in particular questions (P) and (R)}. Each lattice point of this crystal will be able to assume a finite number of different states (say N states) and its behavior will be described (or controlled) by an unambiguous *transition rule*, covering all transitions between these states, as affected by the states of the immediate neighbors.

We will, then, perform the major constructions called for by questions (A)–(E) in Section 1.1.2.1 (and the relevant subsequent discussions of Ch. 1) for a specific model defined along these lines.

2.1.2 Formalization of the spatial and the temporal relations. At this point we introduce some rigorous concepts and notations.

The lattice points of the quadratic crystal (cf. Sec. 2.1.1) are designated by two integer-valued coordinates, i, j. It is natural to treat the crystal as unlimited in all directions, at least as long as there does not emerge some definite reason for proceeding differently. This determines the ranges of i, j:

$$i, j = 0, \pm 1, \pm 2, \cdots . \tag{1}$$

[It does not matter which lattice point is selected as the origin (0, 0).] The pair i, j thus represents a point in the plane, but it is also convenient to view it as a vector, i.e., to treat it as an additive quan-

[1] [The lattice points of a quadratic crystal lie at the corners of squares.]

132

1.1.2.1 The main questions: (A)–(E). Within the above limitations, however, we will deal with problems that are rather central—at least for the initial phases of the subject. We will investigate automata under two important, and connected, aspects: those of logics and of construction. We can organize our considerations under the headings of five main questions:

(A) Logical universality. When is a class of automata logically universal, i.e., able to perform all those logical operations that are at all performable with finite (but arbitrarily extensive) means? Also, with what additional—variable, but in the essential respects standard—attachments is a single automaton logically universal?

(B) Constructibility. Can an automaton be constructed, i.e., assembled and built from appropriately defined "raw materials," by another automaton? Or, starting from the other end and extending the question, what class of automata can be constructed by one, suitably given, automaton? The variable, but essentially standard, attachments to the latter, in the sense of the second question of (A), may here be permitted.

(C) Construction-universality. Making the second question of (B) more specific, can any one, suitably given, automaton be construction-universal, i.e., be able to construct in the sense of question (B) (with suitable, but essentially standard, attachments) every other automaton?

(D) Self-reproduction. Narrowing question (C), can any automaton construct other automata that are exactly like it? Can it be made, in addition, to perform further tasks, e.g., also construct certain other, prescribed automata?

(E) Evolution. Combining questions (C) and (D), can the construction of automata by automata progress from simpler types to increasingly complicated types? Also, assuming some suitable definition of "efficiency," can this evolution go from less efficient to more efficient automata?

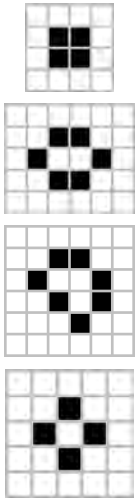
Cellular Automata and emergence of complexity from simple rules

- **Rules:**
  - A cell with 0 or 1 neighbour dies
  - A cell with 2 or 3 neighbour survives
  - A cell with 4-6 neighbours dies
  - A dead cell with 3 neighbours becomes live
- **Initial conditions lead to different outcomes.**

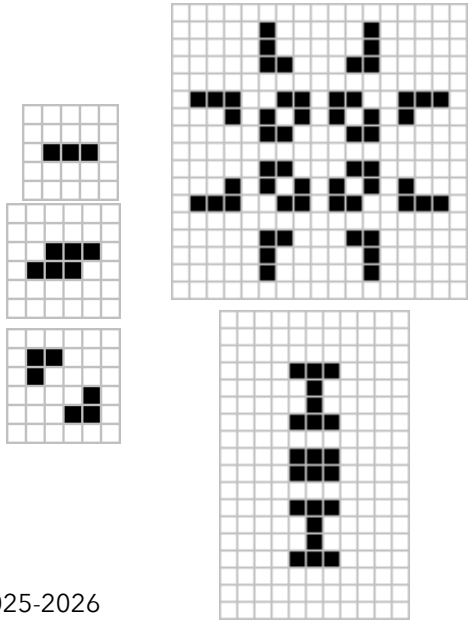
https://conwaylife.com/

- Dart synthesis

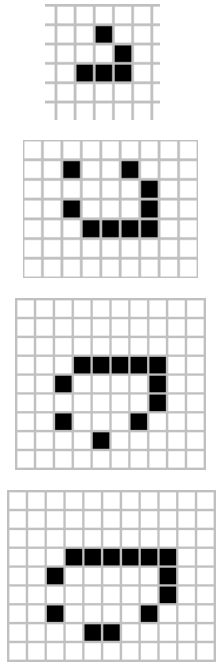
Stable forms



Oscillators

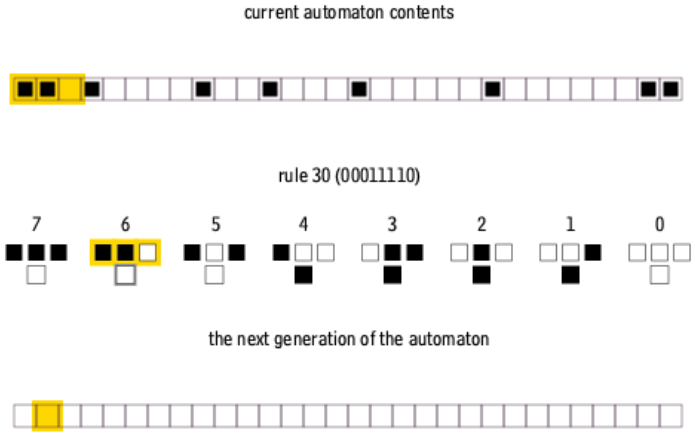


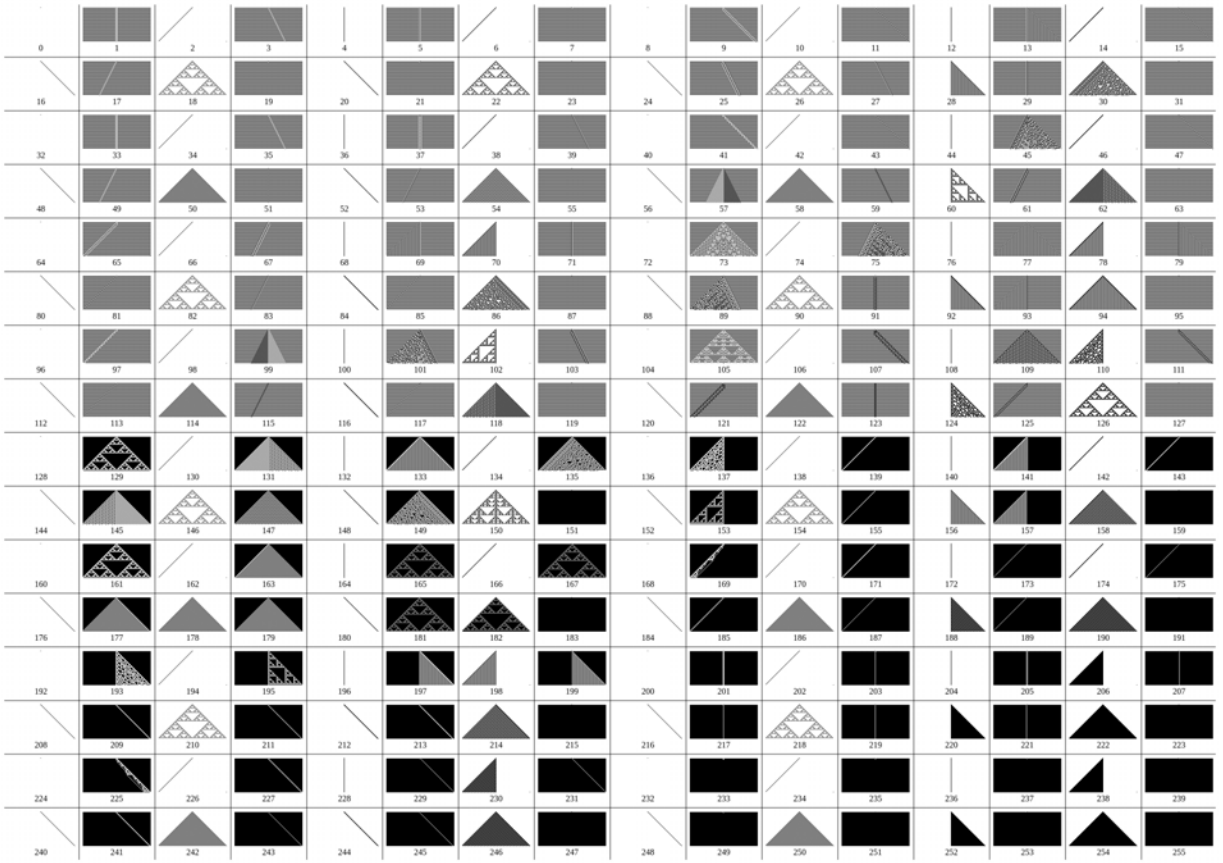
Space ships



Cellular Automata and emergence of complexity from simple rules

Classification: Wolfram classes I–IV:
class I: quiescent; class II: periodic; class III: chaotic; class IV: complex).



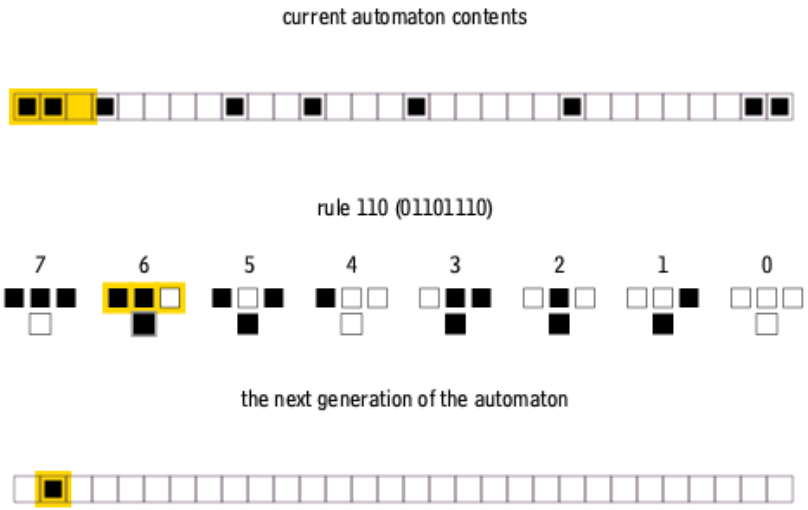


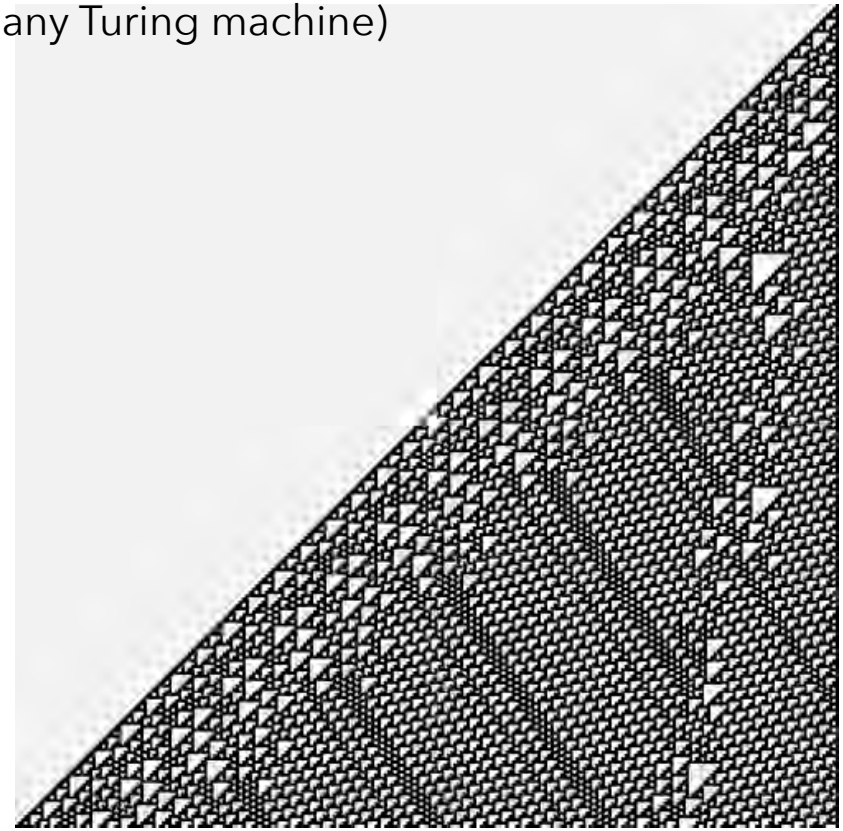
8 motifs and 2^8=256 rules

Cellular Automata and emergence of complexity from simple rules

Classification: Wolfram class IV – complex

Rule 110: Turing complete (can simulate any Turing machine)

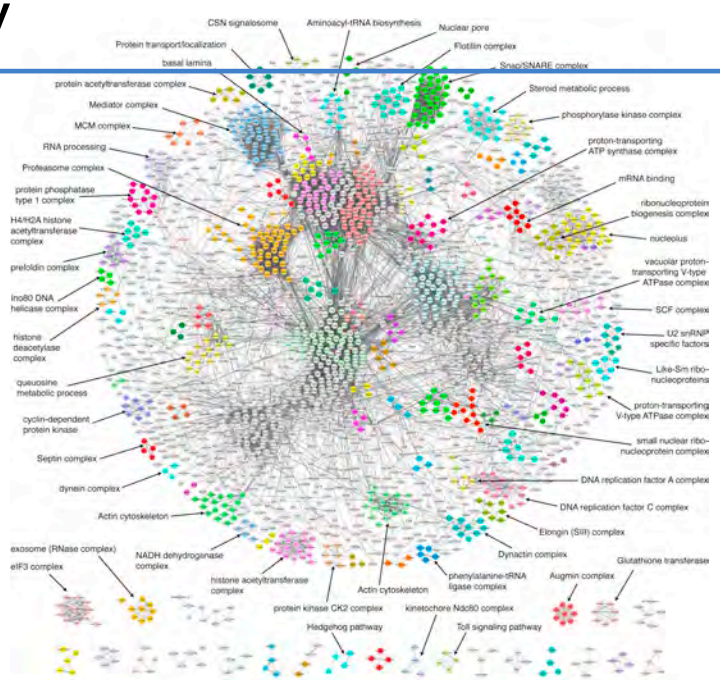




Biological complexity

1. **Heterogeneity:** multiple different agents (~10^4 different proteins/cell)

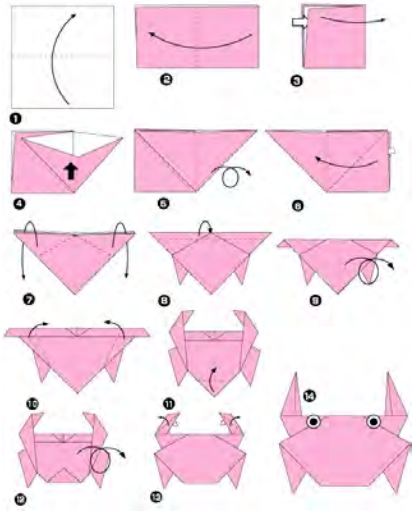
   out of ~10^{10} total proteins/cell



Drosophila proteome

Guruharsha et al, *Cell* 147, 690–703 (2011)

2. **Sequentiality:** new steps bring new contexts which change the course of events.

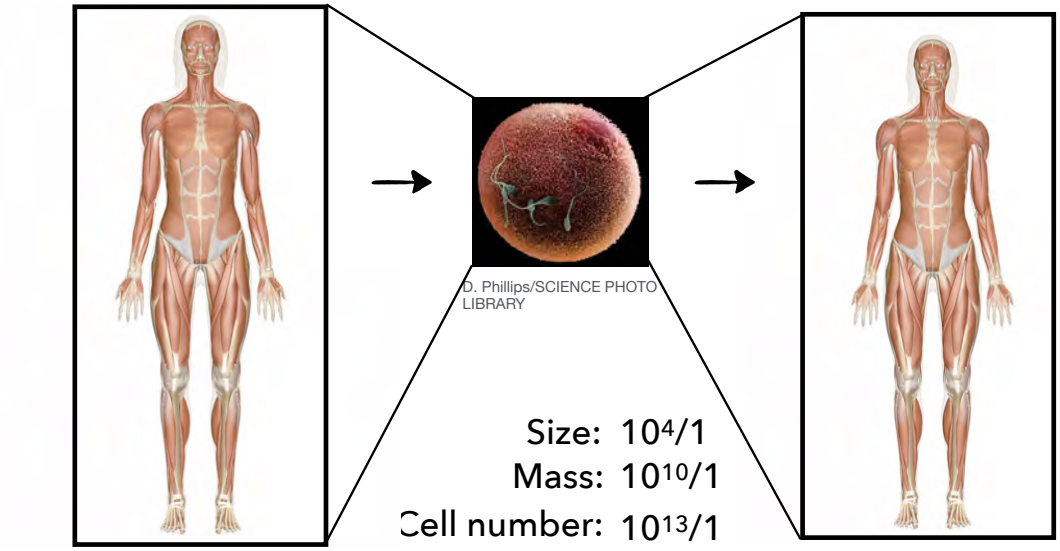


Plan of course

1. Complexity increases during development
2. How to measure complexity?
3. Information can not increase in a closed system

The egg as a« compressed information state » of the future organism

- The complexity of an adult is seemingly compressed/represented in a single cell (considering information as the set of algorithms and parameters required for this process)
- *Questions*:
  - does the egg contain all the information needed to rebuild a new organism?
  - **does the increasing complexity during development require new information?**



Shannon information theory

- **Basic architecture of a communication system**
  « The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. »

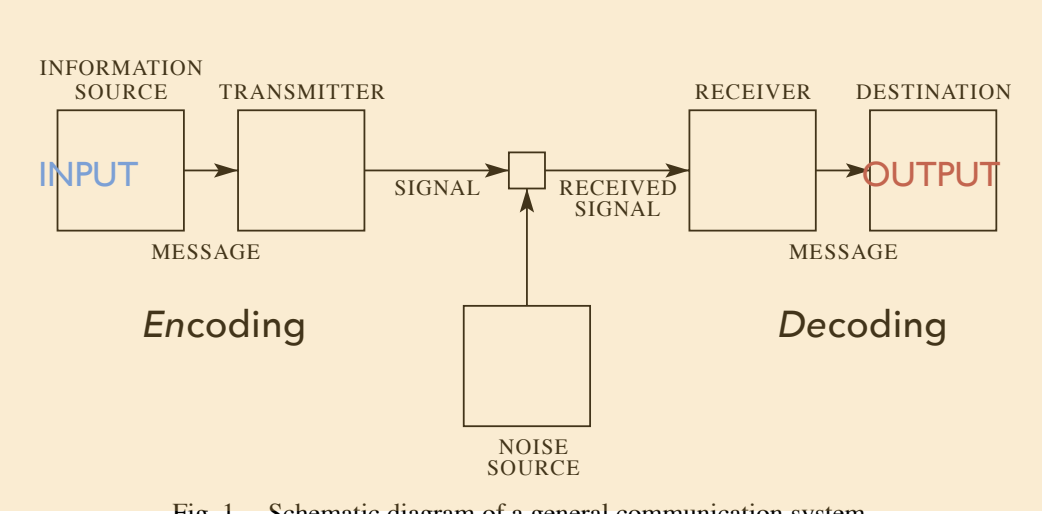
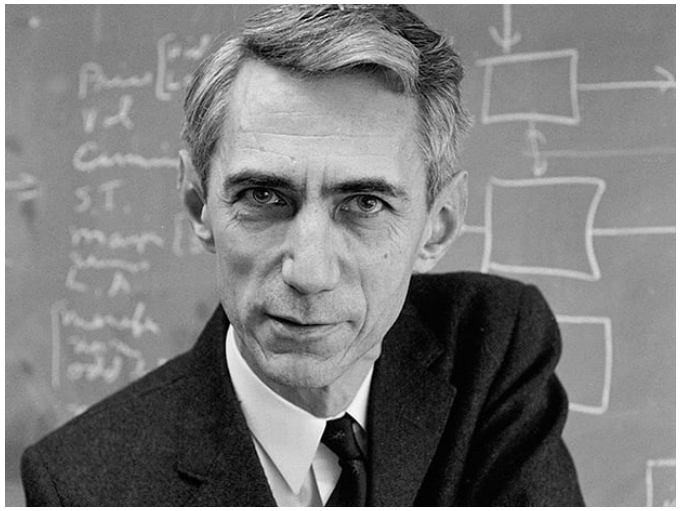


Fig. 1 — Schematic diagram of a general communication system.



Claude Shannon (1916-2001)

Shannon information theory

- Definition of Shannon entropy: $H = -\sum p_i \log p_i$

- The probability-weighted average of information content of events (IC= $-\log p_i$)

  - H **is a measure of uncertainty or hidden information.** The more uncertainty the greater the information gained per choice (surprise)
  - H **has the form of an entropy**

  - H is a number, with unit *bit* with $\log_2$ base.
  - Can be interpreted as the number of Yes/No questions required to fully resolve the uncertainty about a state (discriminate between N possible states).

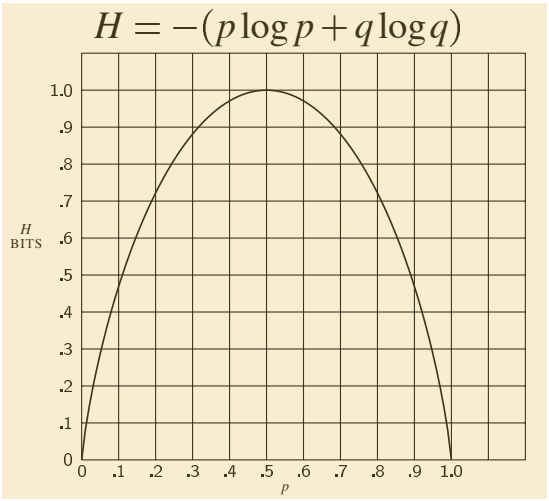


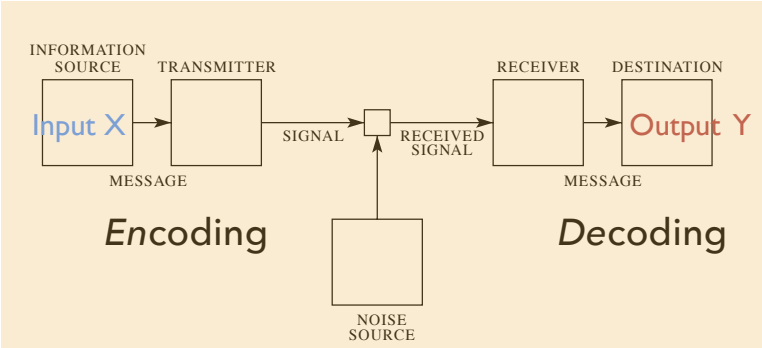
Fig. 7 — Entropy in the case of two possibilities with probabilities p and $(1-p)$.

Shannon information theory

- Definition of mutual information between variables X (input) and Y (output)

$$I(X,Y) = H(X) + H(Y) - H(X,Y)$$

or equivalently: $I(X,Y) = H(X) - H_Y(X) = H(Y) - H_X(Y)$



- I(X;Y) measures the extent of shared information or dependency between variables X and Y.
- Measures the amount of information that one random variable reveals about another
- **The reduction in entropy when measuring Y compared to before measuring is the information that Y provides about X.**

$$R = H(x) - H_y(x)$$
$$= H(y) - H_x(y)$$

Capacity of noisy channel:

$$C = \mathrm{Max}\left(H(x) - H_y(x)\right) = \mathrm{Max}\ I(x,y)$$

ie. the amount of information sent less the uncertainty of what was sent

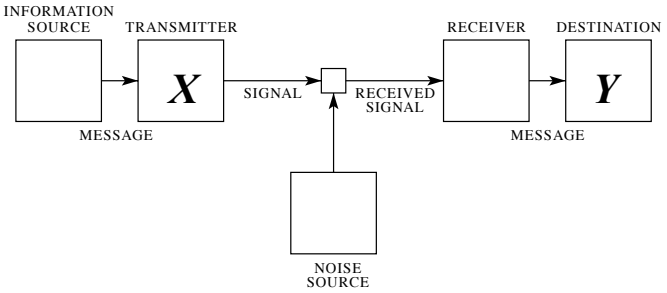
ie. the amount of information received less the part due to noise

Data processing inequality (DPI)

- Intuition & basic principle:
- It is not possible to create new information by processing data. It may be lost or maintained only

- Consider a Markov chain: $X \to Y \to Z$
  - ie. Z is conditionally dependent on Y but not on X

$$P(Z=z|Y=y,X=x)=P(Z=z|Y=y)$$



- Statement: mutual information between X, Y and Z satisfies:

$$I(X;Y) \geq I(X;Z) \quad \text{or} \quad I(Y;Z) \geq I(X;Z)$$

Z cannot have more information about X than Y

In terms of entropy: $H(X|Y) \leq H(X|Z)$

- Interpretation: noise irreversibly corrupts information
- Consequence: no new information can be created in a communication channel

Shannon information theory

- Consequence of data processing inequality in biology

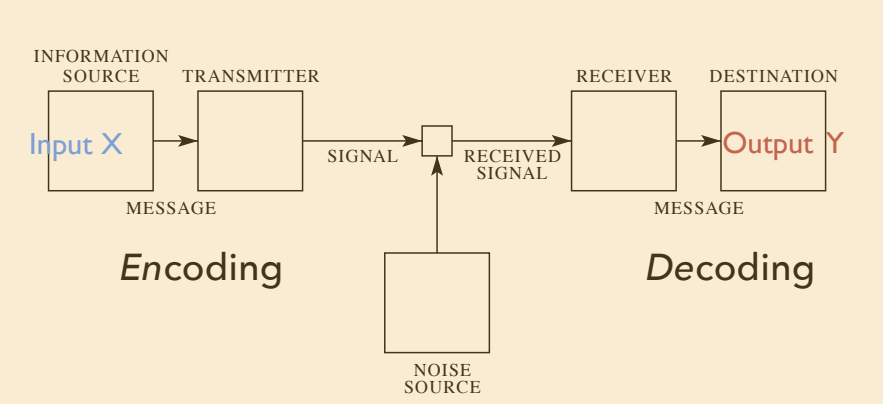
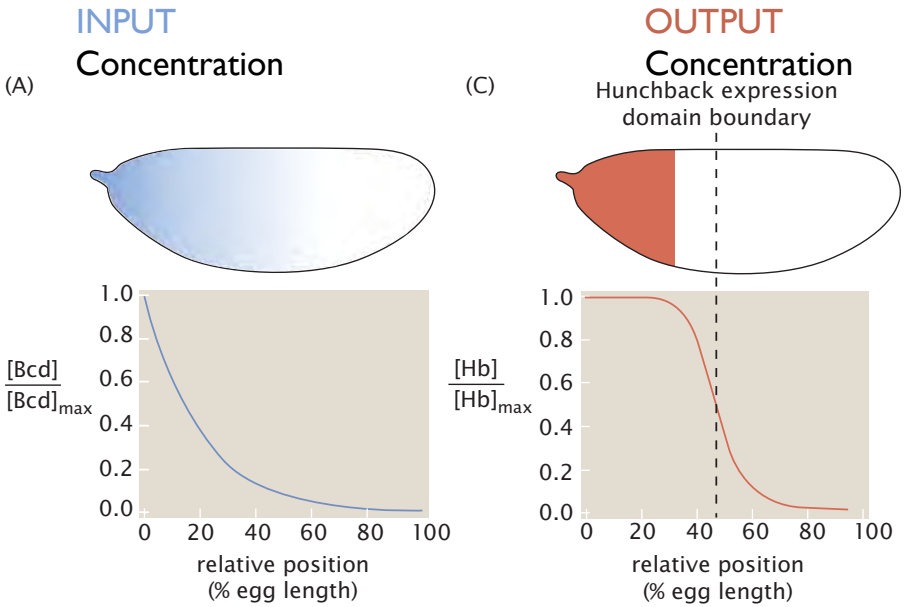


Fig. 1 — Schematic diagram of a general communication system.

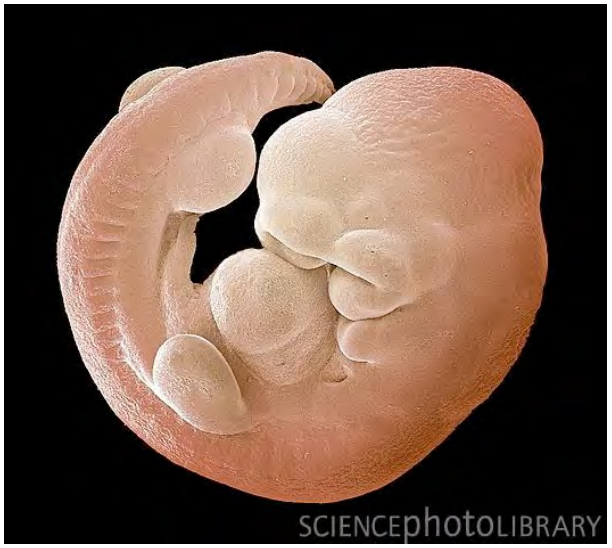


Many bits ~1 bit

R. Phillips, J. Kondev, J. Thériot & H. Garcia.
Physical Biology of the Cell (Garland Science) 2012

Where is the « missing information »?

- Is all the information needed to produce a complex embryo present in the fertilised egg?
- If so what kind of information?
    - Yes: consider algorithmic information content, not just cues (X, Y, Z etc)
    - Also consider all forms of information: chemical, mechanical, structural …
- How to reconcile data processing inequality and increase biological complexity?



Addressing the paradox of information and complexity

- How to reconcile data processing inequality with increased biological complexity?

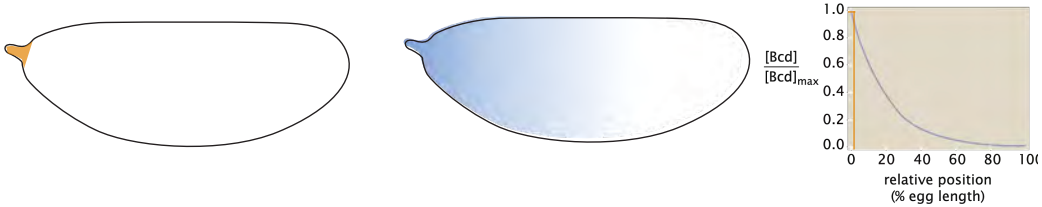
1. Some cues/info are present but are *unaccounted for/hidden*.
2. Some cues need to be *decoded and recoded* (await induction)
3. Some cues (inputs) are present but i*nactive or not interpretable*: require a *new context* (co-factor or new configuration)

Where is the « missing information »?

1. Some information is present but is unaccounted for/hidden

- **Paradox**:
- Onset: A molecule is relocalised at a source. This is described by a step function: $c(x,0) = \begin{cases} c_0 & \text{if } x = 0 \\ 0 & \text{if } x > 0 \end{cases}$ (1 bit)
- Then: the molecule diffuses, is degraded and forms a gradient. If cells can discriminate N concentrations: $I \leq \log_2 N$ bits.

  Yet, information cannot increase (DPI).



- **Solution**:

Apparent "information gain" reflects hidden information reservoirs (geometry, boundary conditions, physical laws ie the algorithms) that are mobilised during the process.

1. Fick's law + degradation: $\frac{\partial c}{\partial t} = D\frac{\partial^2 c}{\partial x^2} - \lambda c.$ Steady state solution: $c(x) = c_0 e^{-x/\ell}, \quad \ell = \sqrt{\frac{D}{\lambda}}.$

Algorithmic Information content (AIC): the exponential profile is not arbitrary—it is generated by a short description (diffusion + decay equation). The hidden information is the "algorithm", not the profile itself. This information is contained in the channel. It is a property of the channel

Where is the « missing information »?

1. Some information is present but is unaccounted for/hidden

- **Solution**:

Apparent "information gain" reflects hidden information reservoirs (geometry, boundary conditions, physical laws) that are mobilised during the process.

2. Geometric information

- Impact of boundary conditions in solution of diffusion equation:

Geometry of space: inherited information (from mother)

- Positional information (PI): encoding of space/geometry by chemical concentration

$$PI := I(X;C) = H(X) - H(X|C), \qquad I(X;C) = \int dx\, p(x) \int dc\, p(c|x) \log_2 \frac{p(c|x)}{p(c)}.$$

Where is the « missing information »?

2. Some cues need to be decoded and recoded (await induction)

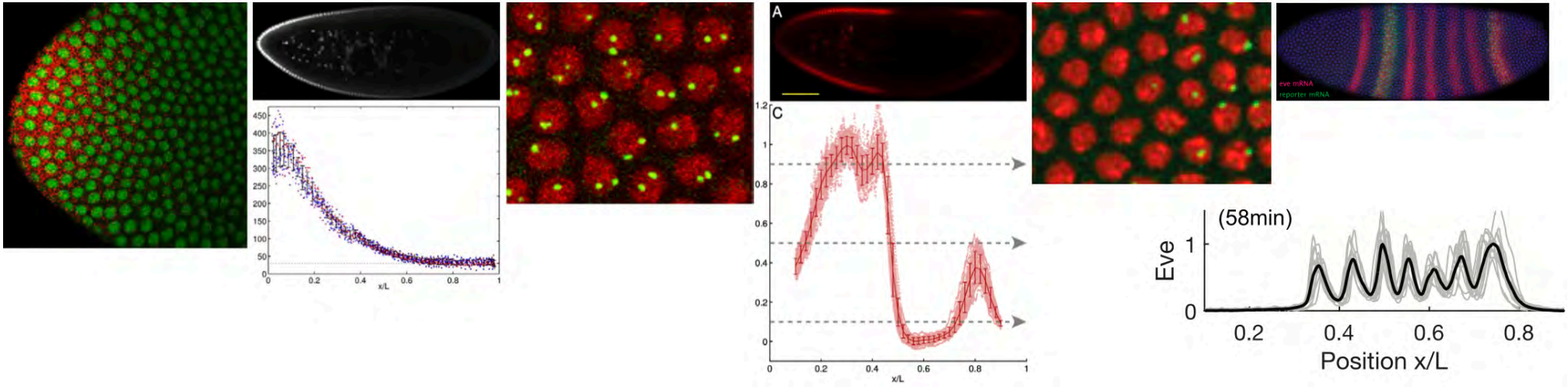
Example: Temporal sequence of gene regulation

Context dependent regulation & decisions

2. Some cues need to be decoded and recoded (await induction)

Example: Temporal sequence of gene regulation

1) *Encoding*	2) *PI* in *I*(Bcd;x)	3) *Recoding*	4) *PI* in *I*(Hb;x)	5) *Recoding*	6) *PI* in *I*(Eve;x)
Mechanisms of Bcd gradient formation		Mechanisms of Hb regulation by Bcd		Eve regulation by Gap genes	

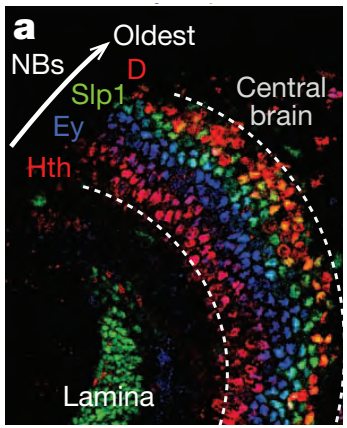


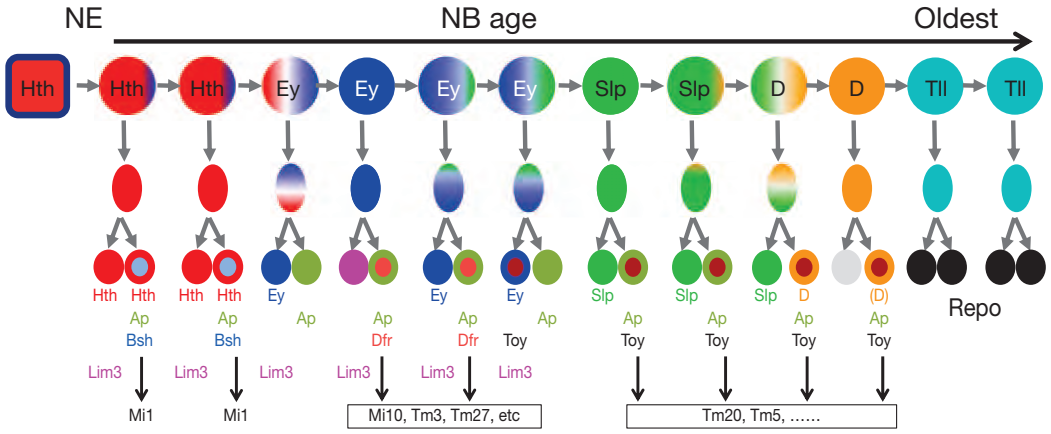
G. Tkacik and T. Gregor. *Development* 148, dev176065. doi:10.1242/dev.176065 (2021)
J. Bothma, H. Garcia et al, T. Gregor and J. Levine, *PNAS*, 111:10598 (2014)
Petkova, M.D., Tkačik,G., Bialek,W., Wieschaus, E.F. and Gregor,T. *Cell* 176, 844-855 (2019)

Context dependent regulation & decisions

2. Some cues need to be decoded and recoded (await induction)

Example: Temporal sequence of gene regulation





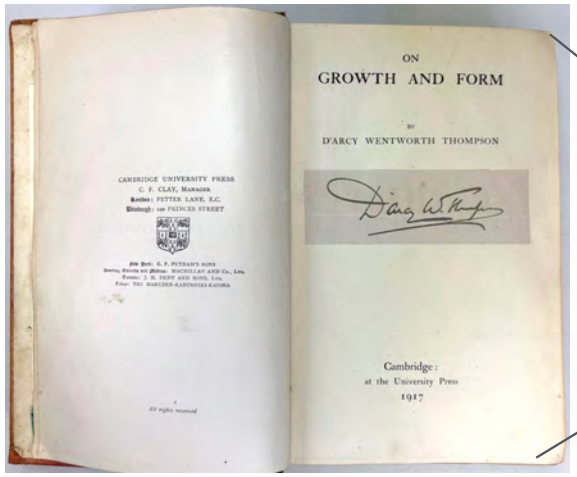
X. Li et al., and C. Desplan. *Nature* 498, 456–462 (2013).

C. Doe *Annu. Rev. Cell Dev. Biol.* 2017. 33:219–40

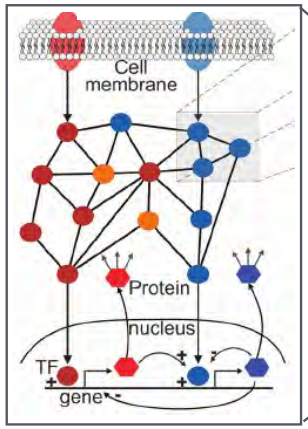
Where is the « missing information »?

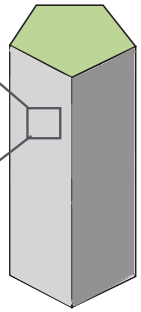
3. Some cues (inputs) are present but require new context (co-factor or new configuration) to become active or interpretable

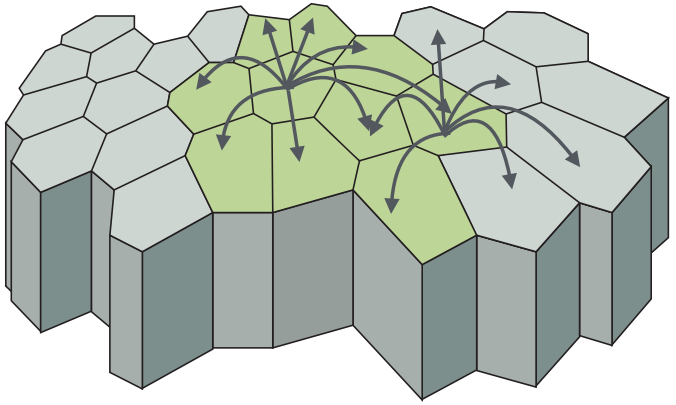
What is the effect of cell number on information and complexity?



- Increasing the copy number of a book does not increase the information content

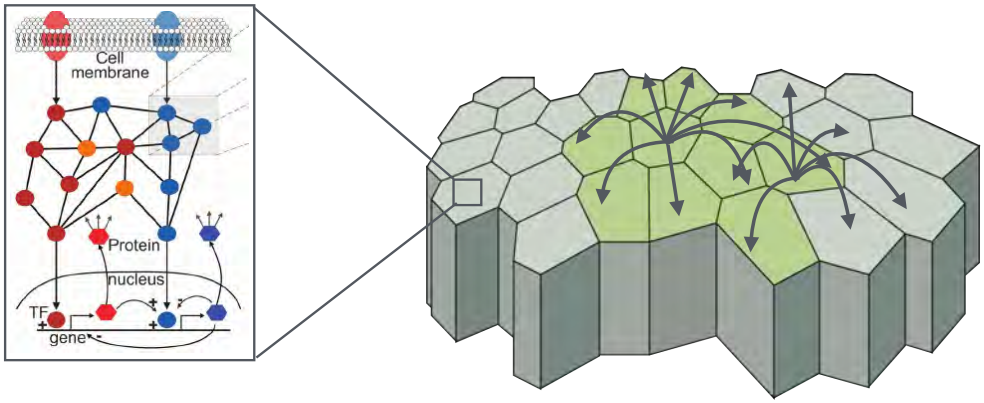






- Increasing the copy number of cells leads to increased complexity
- This stems from cell-cell interactions

What is the effect of cell number on information and complexity?



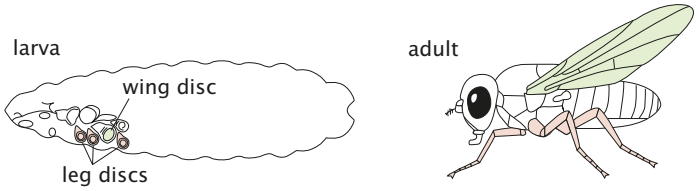
- New signals (eg. protein) become available as a consequence of cell interactions (eg. fluctuations and bistability as in lateral inhibition)
- Cells respond by inducing a new signal that controls downstream processes.
- The early embryo does not have all the signals present at the onset.
- Sequentiality of signals emerges from Algorithms that processes input to output

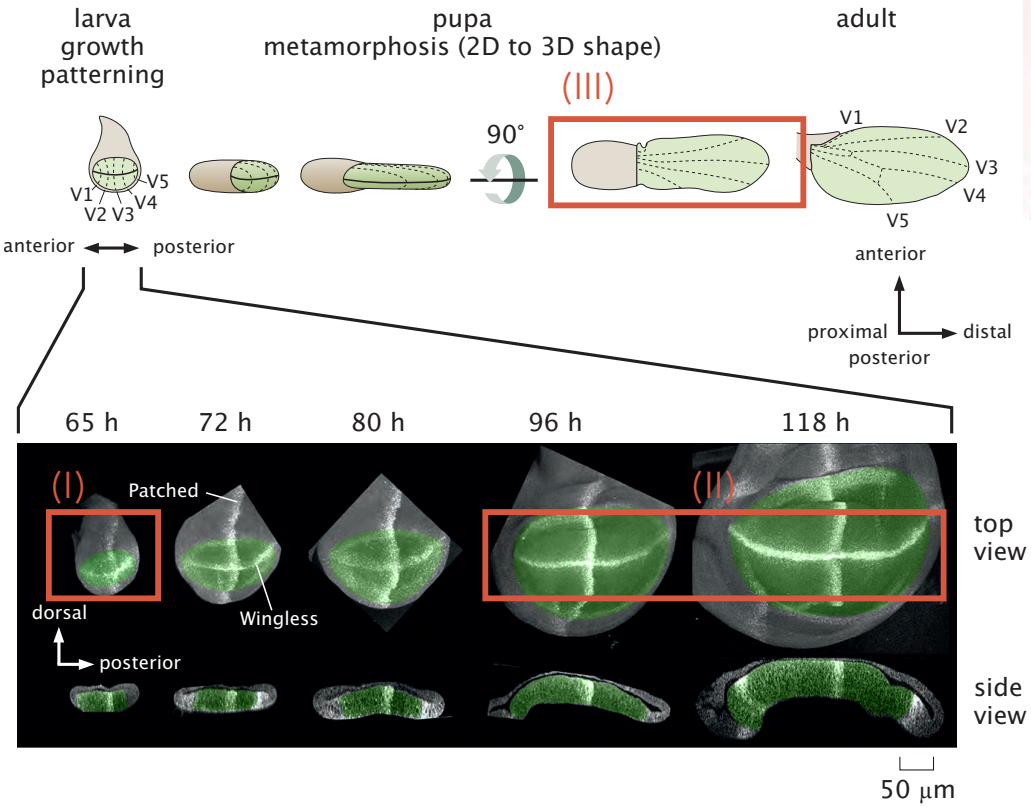
- The potential for their subsequent expression is present in the form of an algorithm that computes induction at a later stages (see course #4 « *Dynamics of information processing* »)

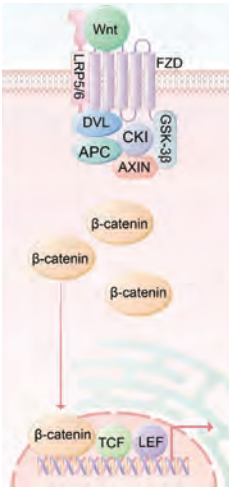
Context dependent regulation & decisions

Morphogenesis: context dependent function of Wnt signalling

- Wing development in *Drosophila* illustrates how signalling by a given ligand (Wnt/Wg) depends on the biological context.
- The same signalling pathway, induced by the ligand Wnt/Wg:
  - **(I)** specifies the wing tissue fate at early stages.
  - **(II)** promotes growth of the wing
  - **(III)** finally Wg induces the specification of the wing margin.





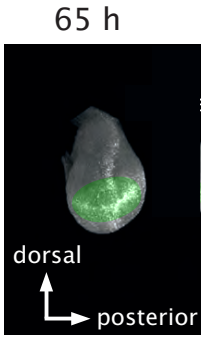


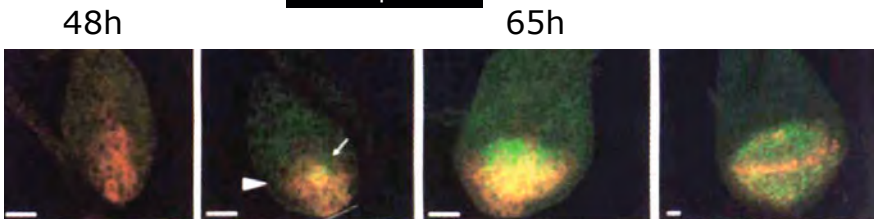
DOI: 10.14218/JCTH.2024.00401

Context dependent regulation & decisions

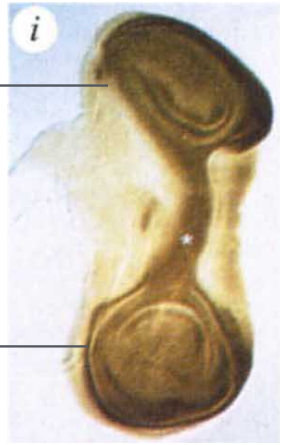
Morphogenesis: context dependent function of Wnt signalling

- At early stage of larval wing imaginal disc development, Wg induces the transcription factor Nubbin and specifies the wing fate (marked in green).





- Removal of Wg at these early stages blocks formation of the wing (hence the gene name).
- Ectopic Wg expression at these early stages induces formation of an ectopic wing (2 wings)

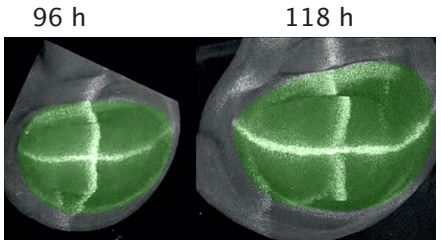


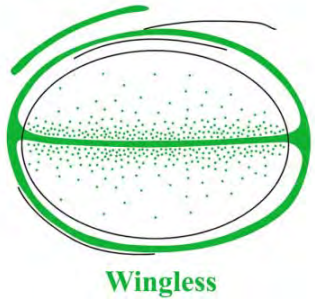
M. Ng et al and S. Cohen. *Nature*, 381: 316-318 (1996)

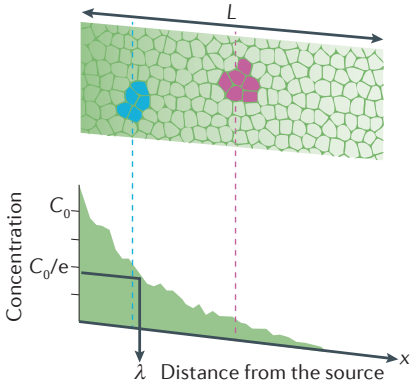
Context dependent regulation & decisions

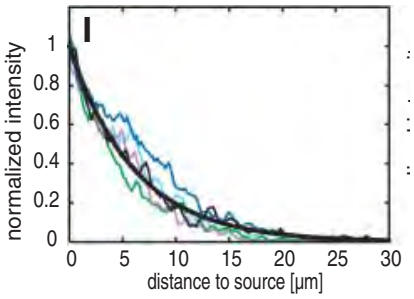
Morphogenesis: context dependent function of Wnt signalling

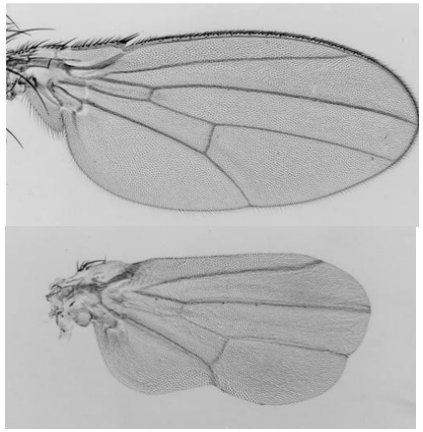
- At later stages of development, Wg is expressed only at the future (aka presumptive) wing margin.
- Wg protein diffuses away from the margin and forms an exponential concentration gradient (morphogen).
- Wg at a distance from the wing margin supports growth of the tissue.









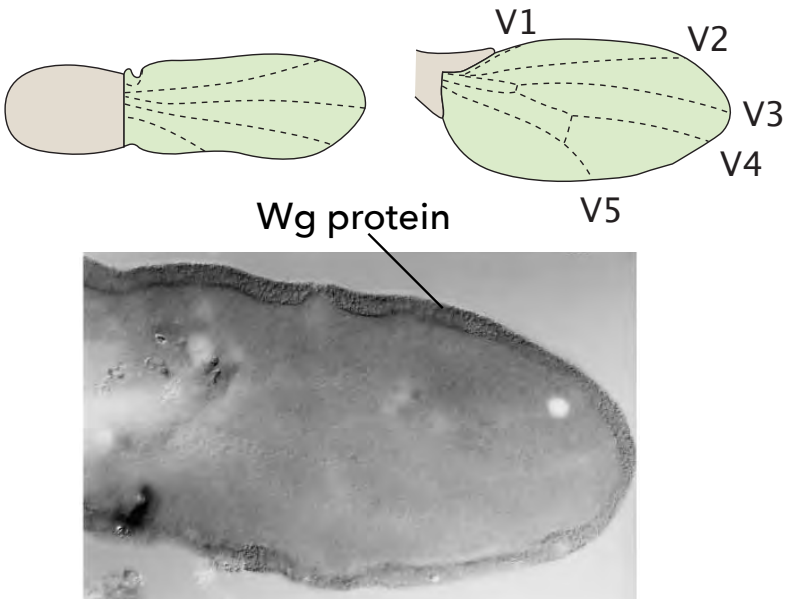


Vincent and Dubois *Developmental Cell*, 3, 615–623 (2002)
Kicheva et al, *Science*, 315, 521 (2007);
Wartlick et al *Nature Rev MCB*, 12, 594 (2011)
Couso et al, *Development* 120, 621-636 (1994)

Context dependent regulation & decisions

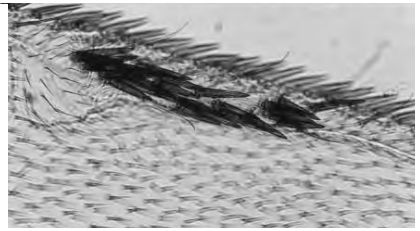
Morphogenesis: context dependent function of Wnt signalling

- During pupal development, metamorphosis the wing imaginal discs acquires its final shape
- Cell differentiation in veins and the wing margin
- Wg specifies the differentiation of bristles at the wing margin

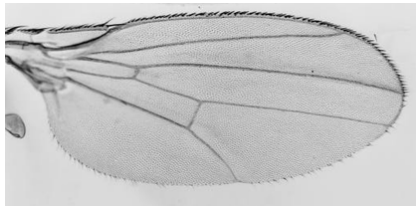


control

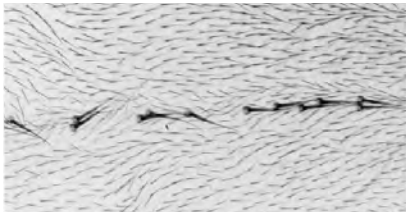
Activated Wg signalling
(GSK3ß - clone)



wg^{ts}



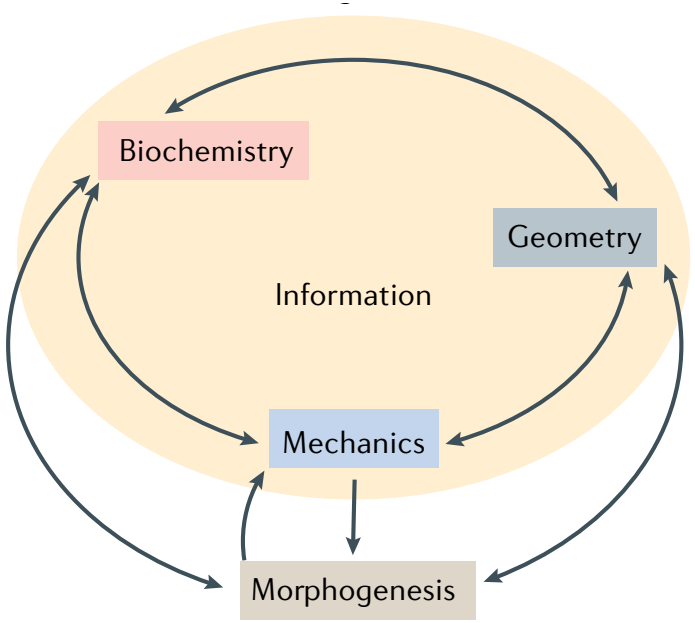
Reduced Wg signalling
(ts shift in pupa 0-4hAPF)



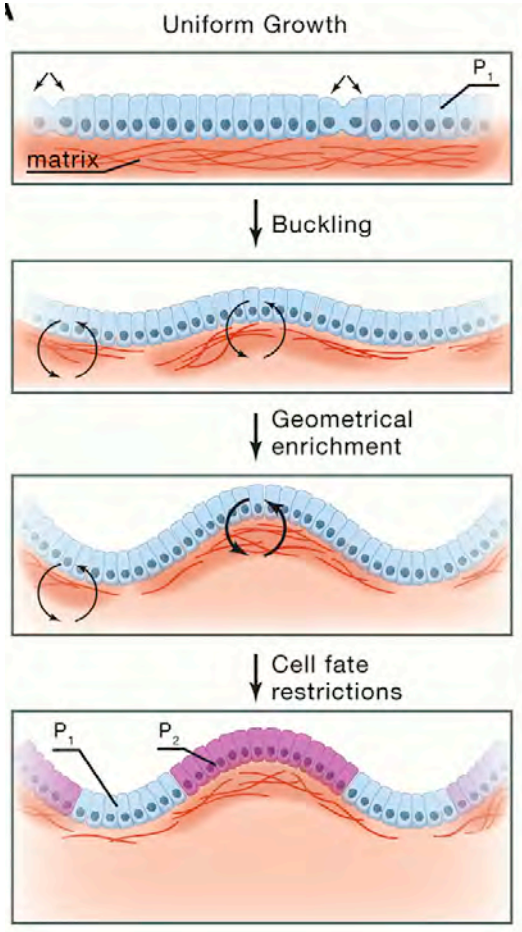
JP. Couso et al, *Development* 120, 621-636 (1994)

Context dependent regulation & decisions

Morphogenesis: change in shape induces signalling



Collinet C. & Lecuit T. *Nature Rev. Mol. Cell Biol.*, 2021

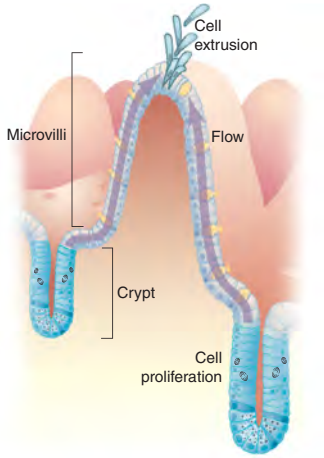


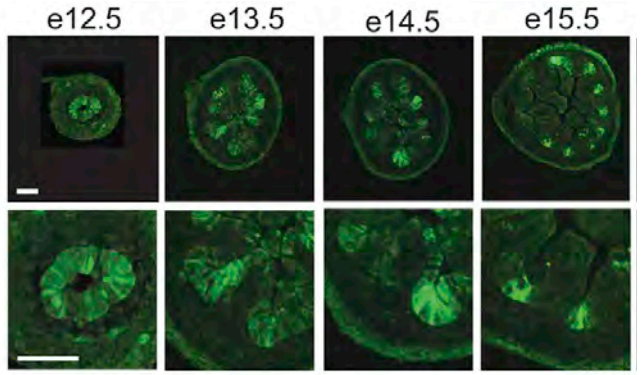
Hannezo and Heisenberg, *Cell* 178, 13-25 (2019)

Context dependent regulation & decisions

Morphogenesis: change in shape induces signalling

- Proliferating intestinal stem cells (ISC) form a niche at the base of villi (crypt).
- Lgr5, a marker of ISC, is first expressed in the entire epithelium, and is later restricted to regions at the base of villi.

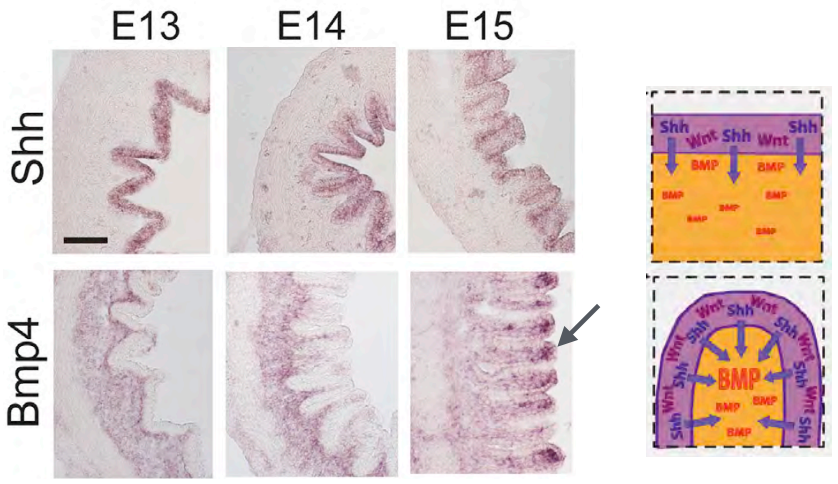




Lgr5

- Shh, is expressed uniformly in the intestine epithelium during formation of the villi.

- BMP4, another growth factor expressed in the underlying mesenchyme, is first expressed uniformly, but later on is restricted to the distal tip of the villi.
- BMP4 subsequently represses ISC induction at the villi tip.

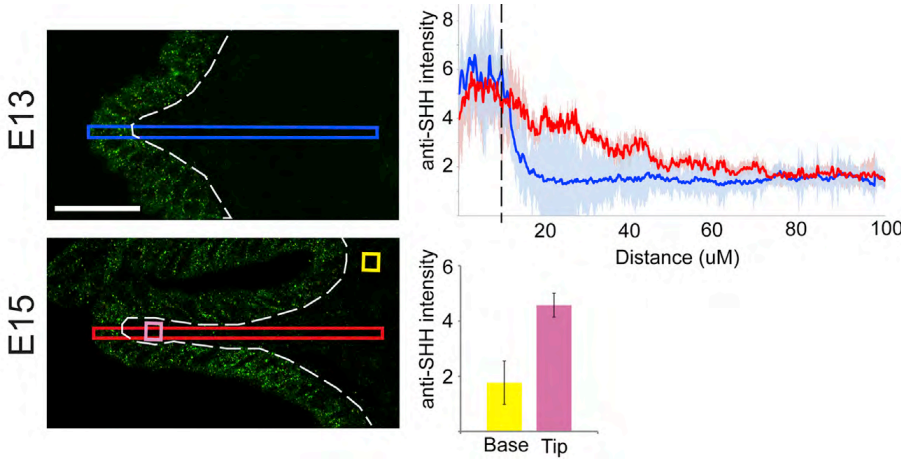


A. Shyer et al. and C. Tabin, *Cell* 161, 569–580 (2015)

Context dependent regulation & decisions

Morphogenesis: change in shape induces signalling

- *Hypothesis*: **Impact of surface to volume ratio on concentration of Shh.** Shh concentrates in mesenchyme surrounded by a higher surface of epithelium.
- The concentration profile of Shh changes as the tissue folds and [Shh] increases at the tip.



A. Shyer et al. and C. Tabin, *Cell* 161, 569–580 (2015)

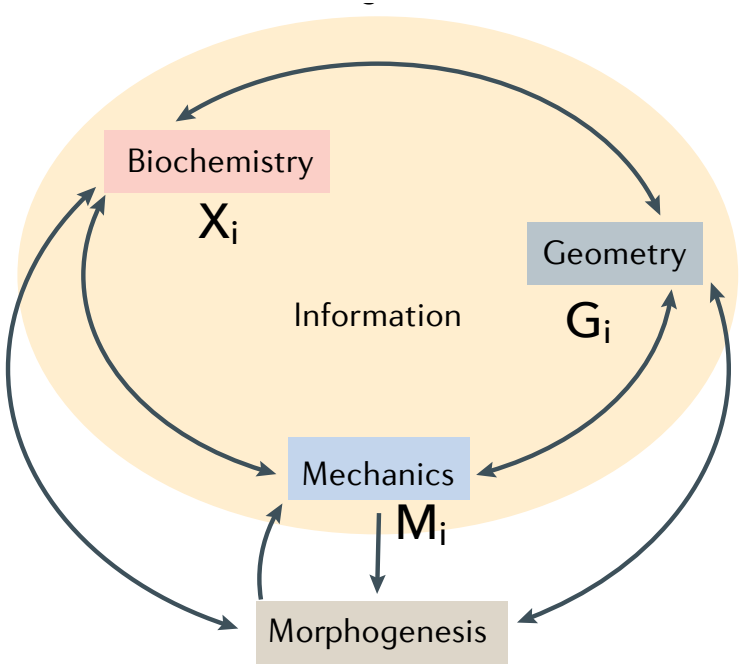
Where is the missing information

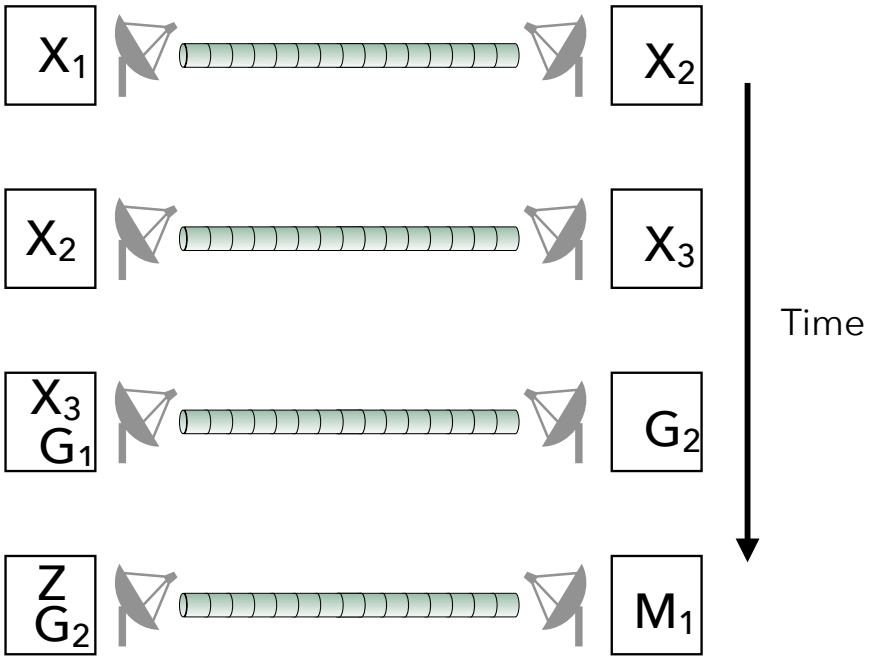
- How to reconcile data processing inequality with increased biological complexity?

1. Some cues/info are present but are unaccounted for/hidden.
2. Some cues need to be decoded and recoded (await induction)
3. Some cues (inputs) are present but inactive or uninterpretable: require a new context (co-factor or new configuration)

Where is the « missing information »?

This amounts to new channels (Shannon) being open step by step as the dynamics of physicochemical processes unfold

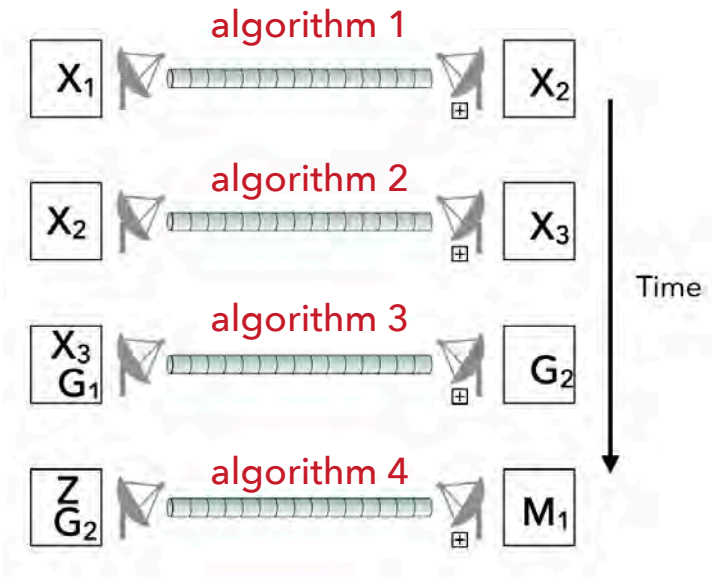


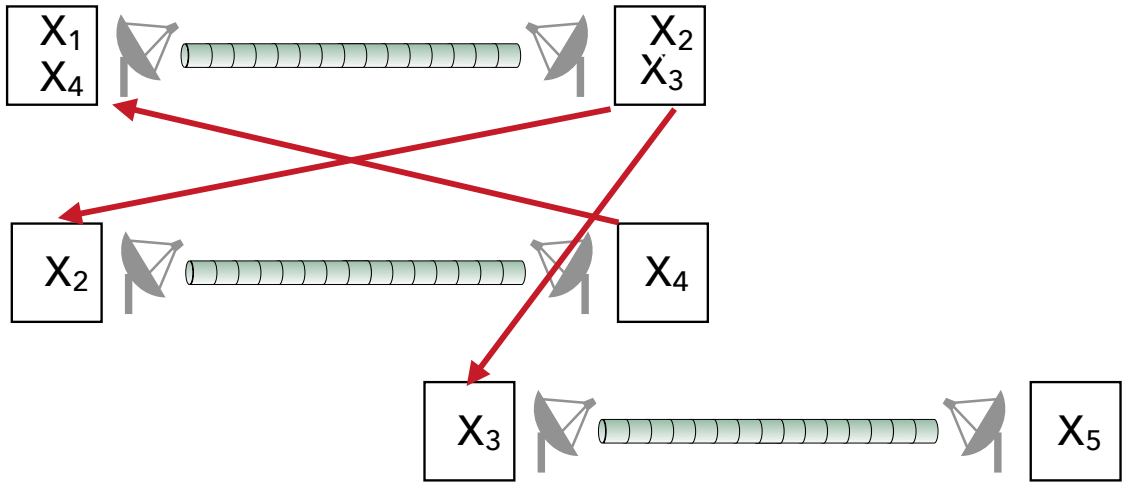


Where is the « missing information »?

Consider also **Algorithmic Information Content**:

- within each channel
- among channels connected in a network



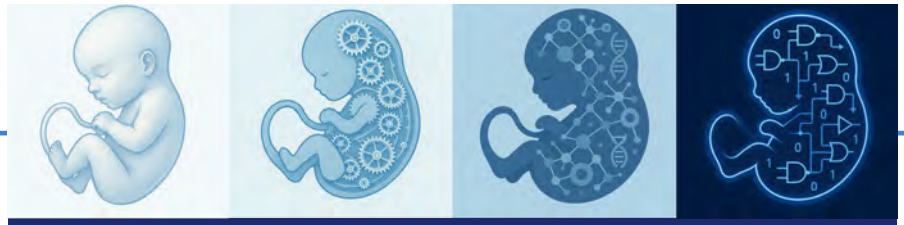


Conclusions

1. Complexity increases during development

2. Distinguish different kinds of complexity:
   - Statistical complexity: description independent of mechanism/algorithm
   - Descriptive complexity: *length of algorithm* that describes/recapitulates the system
   - Generative complexity: *duration* of computation.

3. Algorithmic level of analysis can be encapsulated in Algorithmic information content (Kolmogorov complexity). Good approximation via compression.
AIC underlies complexity
4. Data processing inequality: information can only be lost in a channel.
5. Shannon information theory during development:
   - « Unaccounted for » information: inactive or uninterpretable cues (structure, mechanics etc) and channels.
   - Algorithmic information in channels, between channels.



Thomas LECUIT, chaire Dynamiques du vivant

Qu'est-ce que l'information biologique ? (II)

COURS : 20 novembre > 18 décembre 2025

COURS

Le jeudi de 10 h à 12 h
Amphithéâtre Guillaume Budé

Jeudi 20 Novembre 2025
Introduction :
approche computationelle du vivant

Jeudi 27 Novembre 2025
Complexité et information
au cours du développement

Jeudi 4 Décembre 2025
Vision logique des flux d'information

Jeudi 11 Décembre 2025
Vision dynamique des flux d'information

Jeudi 18 Décembre 2025
Apprentissage non neuronal
dans un système biologique

COLLOQUE

De 9h à 18h
Amphithéâtre Maurice Halbwachs

Vendredi 26 juin 2026
Information flow and computation in living systems

Les cours et colloques
sont gratuits, en accès libre,
sans inscription préalable.

Image générée par une I.A. © T. Lecuit.

COLLÈGE DE FRANCE 1530

Thomas Römer, administrateur du Collège de France
11, place Marcelin-Berthelot, 75005 Paris
www.college-de-france.fr

Année académique 2025/2026