

The many bits of positional information

Gašper Tkačik, Thomas Gregor

▶ **To cite this version:**

Gašper Tkačik, Thomas Gregor. The many bits of positional information. Development (Cambridge, England), 2021, 148 (2), 10.1242/dev.176065 . pasteur-03213416

HAL Id: pasteur-03213416
https://pasteur.hal.science/pasteur-03213416

Submitted on 30 Apr 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

© 2021. Published by The Company of Biologists Ltd | Development (2021) 148, dev176065. doi:10.1242/dev.176065

REVIEW

The many bits of positional information

Gašper Tkačik[1] and Thomas Gregor[2,3,*]

ABSTRACT

Half a century after Lewis Wolpert's seminal conceptual advance on how cellular fates distribute in space, we provide a brief historical perspective on how the concept of positional information emerged and influenced the field of developmental biology and beyond. We focus on a modern interpretation of this concept in terms of information theory, largely centered on its application to cell specification in the early *Drosophila* embryo. We argue that a true physical variable (position) is encoded in local concentrations of patterning molecules, that this mapping is stochastic, and that the processes by which positions and corresponding cell fates are determined based on these concentrations need to take such stochasticity into account. With this approach, we shift the focus from biological mechanisms, molecules, genes and pathways to quantitative systems-level questions: where does positional information reside, how it is transformed and accessed during development, and what fundamental limits it is subject to?

KEY WORDS: Developmental biology, Pattern formation, Shannon information, Drosophila embryo, Lewis Wolpert

Introduction

How and when cells in a developing organism know what they are and where they are, are questions that are almost synonymous with the definition of developmental biology (Kirschner and Gerhart, 1997; Lawrence, 1992). In metazoans, different cells have to perform different tasks. They therefore need to interpret cues that steer them towards the correct fates (Ephrussi and St. Johnston, 2004). Evolution had the possibility to act on both the 'cues' and the machinery that performs the 'interpretation' of these cues. Wolpert's concept of positional information (PI) elegantly touches on both of these aspects.

The idea that cells adopt different fates by 'sensing' the presence or absence of chemicals, so called fate-determining factors or 'determinants' (Conklin, 1905; Wilson, 1904), dates back to the early 20th century. Experiments on sea urchin embryos suggested that developmental patterns could be determined by opposing 'gradients' (Boveri, 1901a,b), while regeneration experiments on flat worms postulated the existence of 'formative substances' that influence the developmental plan of the embryo (Morgan, 1904, 1905). The notion of chemical gradients acting at large distances to affect developmental patterning has an even longer history (Lawrence, 2001), but it was not until the middle of the 20th century that Turing postulated that concentrations of specific chemicals, called 'morphogens', might instruct cell fates and thus the emergence of shape and form in a developing organism (Turing, 1952).

[1]Institute of Science and Technology Austria, Am Campus 1, AT-3400 Klosterneuburg, Austria. [2]Joseph Henry Laboratories of Physics and the Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ 08544, USA. [3]Department of Developmental and Stem Cell Biology, UMR3738, Institut Pasteur, FR-75015 Paris, France.

*Author for correspondence (tg2@princeton.edu)

G.T., 0000-0002-6699-1455; T.G., 0000-0001-7692-8348

The next big idea was the inclusion of space and the notion that spatial fields of chemicals could lead to developmental patterning and cellular differentiation (Crick, 1970; Lawrence, 1970; Wolpert, 1969). Key to this idea is a predetermined initial symmetry-breaking event, often triggered by asymmetrically localized factors. For example, morphogens are produced in cells that are located in spatially restricted regions and they diffuse along a central axis of an egg or tissue, thereby establishing a gradient. Wolpert eloquently postulated that cells could determine their fate by interpreting local concentrations of these graded profiles, and he coined the abstract notion that these profiles thus contain 'positional information' (Wolpert, 1969, 1971). This was one of the solutions he proposed for the 'French Flag Problem' of patterning (Wolpert, 1969), which later became colloquially known as the 'French Flag' model (Sharpe, 2019). Here, adjacent groups of cells are delineated by a concentration threshold, which defines a boundary. Fate determination in this model is due to an additional step, in which cells 'interpret' the concentration of the morphogen. 'Information' is thus contained in the nominal value of the concentration at a given position, and in the molecular apparatus that transforms this value into a cellular response. Thus, morphogen concentrations of two orthogonal gradients could act as positional coordinates, defining a two-dimensional spatial fate map. Individual cells measure and interpret the local morphogen concentration and determine the appropriate fate choice for that position, as manifested experimentally by Spemann's famous grafting experiments (Spemann and Schotté, 1932) and by the arrangement of chick wing digits (Saunders and Gasseling, 1968).

Conceptually, Wolpert's postulate was indeed a big leap forward, as evidenced by the significant gap before its experimental manifestation and its subsequent molecular proof. The framework of PI found immediate popularity and was put to use, e.g. by Postlethwait to interpret his famous Antennapedia *Drosophila* mutant, in which a pair of head antennae is converted into legs. Postlethwait postulated 'that perhaps all appendages may have the same PI and that what makes one appendage different from another is the response of cells with a different determination to the same set of proximodistal, mediolateral positional cues' (Postlethwait and Schneiderman, 1971), which turned out to be the case for Hox genes in all animals (Akam, 1989).

In 1974, the existence of cytoplasmic determinants was undoubtedly proven by transplantation experiments in *Drosophila* (Illmensee and Mahowald, 1974). Fifteen years later, the first morphogen molecule was finally discovered, with the anterior determinant Bicoid in the *Drosophila* embryo displaying all the characteristics of Wolpert's concept (Driever and Nüsslein-Volhard, 1988a,b; reviewed by Lawrence, 1988; Wolpert, 1989). This discovery was immediately followed by the demonstration that a frog growth factor determines differential cell fates according to concentration thresholds (Green and Smith, 1990; Green et al., 1990; reviewed by Green and Smith, 1991). Subsequently, many more PI-carrying morphogens were discovered (Neumann and Cohen, 1997), including in vertebrates such as zebrafish (Chen and Schier, 2001) and chick (McMahon et al., 2003).

Development (2021) 148, dev176065. doi:10.1242/dev.176065

The concept of PI has since had enormous success in shaping our understanding of spatial patterning in developing organisms (Fasano and Kerridge, 1988; Lacalli and Harrison, 1991; Moses and Rubin, 1991; Reinitz et al., 1995; Tomlinson et al., 1987; Wolpert, 1969, 1971; see review by Wolpert, 1996). Given its intuitively physical nature, the concept of PI also lent itself swiftly to quantitative questions. For example, the number of different thresholds that can be set reliably by a given concentration gradient could be estimated using straightforward calculations (Lewis et al., 1977). Moreover, the idea of PI has been applied to understand precision and reproducibility in development. Specific morphological features during early development have been studied in great detail and have been shown to occur reproducibly and precisely across wild-type embryos (Crauk and Dostatni, 2005; Gregor et al., 2005; Houchmandzadeh et al., 2002; Jaeger and Reinitz, 2006; Jaeger et al., 2008; Lecuit et al., 1996), while perturbation experiments have revealed systematic shifts of these features (Capovilla et al., 1992; Kraut and Levine, 1991; Rivera-Pomar et al., 1995). These findings have thereby established a causal – but not quantitative – link between the PI encoded in morphogens and the resulting body plan.

To sharpen the use of PI and to elevate its usefulness as a quantitative tool, we propose here a mathematical definition that is based on the concepts of Shannon's information theory (Box 1). We first introduce the mathematical framework that allows us to formalize the colloquial concept whereby 'a cell determines its position from noisy patterning cues in the form of low-concentration molecular gradients'. We next highlight how the combination of precise data and mathematically rigorous PI quantities helped us revisit key biological questions. Finally, we end by formulating several unsolved puzzles to motivate future research.

In search of a mathematical framework for PI

Initial efforts towards a quantitative interpretation of PI relied mainly on indirect, system-specific quantities. Some of the measured quantities were based on the necessity for precision and reproducibility in the patterning process (Bollenbach et al., 2008; Desponds et al., 2016; Gregor et al., 2007; He et al., 2010; Morishita and Iwasa, 2009, 2011), whereas others were based on the idea that special shapes of morphogen profiles, 'sharp' gene expression boundaries, or a 'stripe' of gene expression, are intrinsically favored for successful patterning and are thus selected for by evolution (Briscoe and Small, 2015; Crauk and Dostatni, 2005; Erdmann et al., 2009; Fujioka et al., 1995; Houchmandzadeh et al., 2002; Jaeger et al., 2004; Meinhardt and Gierer, 1980; Sokolowski et al., 2012). Interestingly, both intuitions contain a partial, yet incomplete, characterization of PI. However, a unifying mathematical framework that could consistently merge the two was missing.

Ideally, a mathematical formalization of PI should satisfy the following properties: (1) PI should be independent of specific biological mechanisms that establish or read out primary morphogen gradients or patterns; (2) PI should be a numerical measure that can be experimentally determined; (3) PI should be defined without *a priori* assumptions about pattern shape, and thus should be applicable to any arbitrarily complex spatial gene expression pattern; (4) PI should be applicable and generalizable to multiple concentration fields of patterning molecules; and (5) PI should allow for theoretical first principle derivations, and lend itself to the establishment of a predictive theory for biological patterning.

These five desired properties can all be fulfilled simultaneously when information about the physical position (i.e. the coordinates) of a cell within an organism is encoded in noisy spatiotemporal concentration profiles of morphogen molecules. Here 'encoding' signifies the biological processes that establish spatially graded molecular profiles (Fig. 1). The mechanistic implementation of this encoding could be complex, consisting of a variety of biological steps that are only partially known: maternal cues, gene regulatory and signaling networks, cell-cell communication, diffusion, etc. However, PI should only be a function of the resulting spatiotemporal concentration profiles, regardless of the processes that establish them, as these profiles are by definition the sole quantities that determine subsequent morphological events. In addition, PI should be equivalently applicable to both classical graded profiles of signaling molecules (morphogen gradients) and spatiotemporal expression patterns of developmental genes; for simplicity, we therefore use the term 'morphogen' broadly to refer to both of these cases.

Importantly, the issue of how PI is read out or decoded is separate from the measure of how much information is present in the pattern. Here 'decoding' stands for the biological processes that estimate the physical location of a cell in a tissue or determine its discrete cell fate based on readout or measurement of noisy local morphogen concentration levels (i.e. the processes that 'interpret the positional cues'). Both, encoding and decoding are mechanism dependent (Fig. 1). Building a general mathematical framework relies on the possibility of separating these mechanisms from the actual representation of PI, which depends solely on directly measurable concentration profiles and is thus mechanism independent.

PI is not only 'established' (e.g. as a morphogen gradient) and then 'read-out' (e.g. via thresholds), but it can also be 'recoded' (Fig. 1). Recoding means that the information present in the morphogen gradient is reformatted or transformed into another internal cellular representation (e.g. for downstream processing convenience). Gap genes in *Drosophila*, for example, carry PI much like their primary maternal morphogen regulators do. This information originates from the primary morphogens and vanishes if they are removed (Petkova et al., 2019). Typically, the process of 'reading out' implies applying an operation on the morphogen gradient that loses PI. Yet gap genes individually (and most likely as a group) encode at least as much

Box 1. An introduction to Shannon's 'information theory'

When a change in random variable, X, leads with some probability to a change in another random variable, Y, we say that X 'has information' about Y. This information would allow us to infer (or predict) the value of Y if we knew the value of X, and vice versa. Claude Shannon identified mutual information, I(X;Y), as the unique measure that mathematically captures such a statistical dependence between X and Y, while satisfying various intuitive expectations (e.g. independent bits of information add) and remaining independent of system-specific assumptions (Shannon, 1948).

Mutual information is derived from a more basic quantity, the 'entropy' $S(X)=-\Sigma P(X) \log_2 P(X)$, where the summation extends over all values of X that happen with probability P(X). Entropy measures the dynamic range of the distribution, and is conceptually related to its variance. Mutual information is I(X;Y)=S(X)+S(Y) – S(X,Y), or the difference in entropy of X and Y taken separately (as if they were statistically independent) and jointly (which captures any correlation between them). Mutual information generalizes the linear correlation coefficient (or regression R^2) to nonlinear dependence between two random variables. Linear correlation can miss statistical dependencies that information will detect. Information will be zero only if X and Y are statistically independent, and thus no inference about one variable is possible from the other. Despite its unusual notation, *I(X;Y)* is not a function but a single non-negative number, the units of which are 'bits' (see Box 4). Larger values imply stronger statistical dependence, less noise and higher predictability between the two variables.

Development (2021) 148, dev176065. doi:10.1242/dev.176065

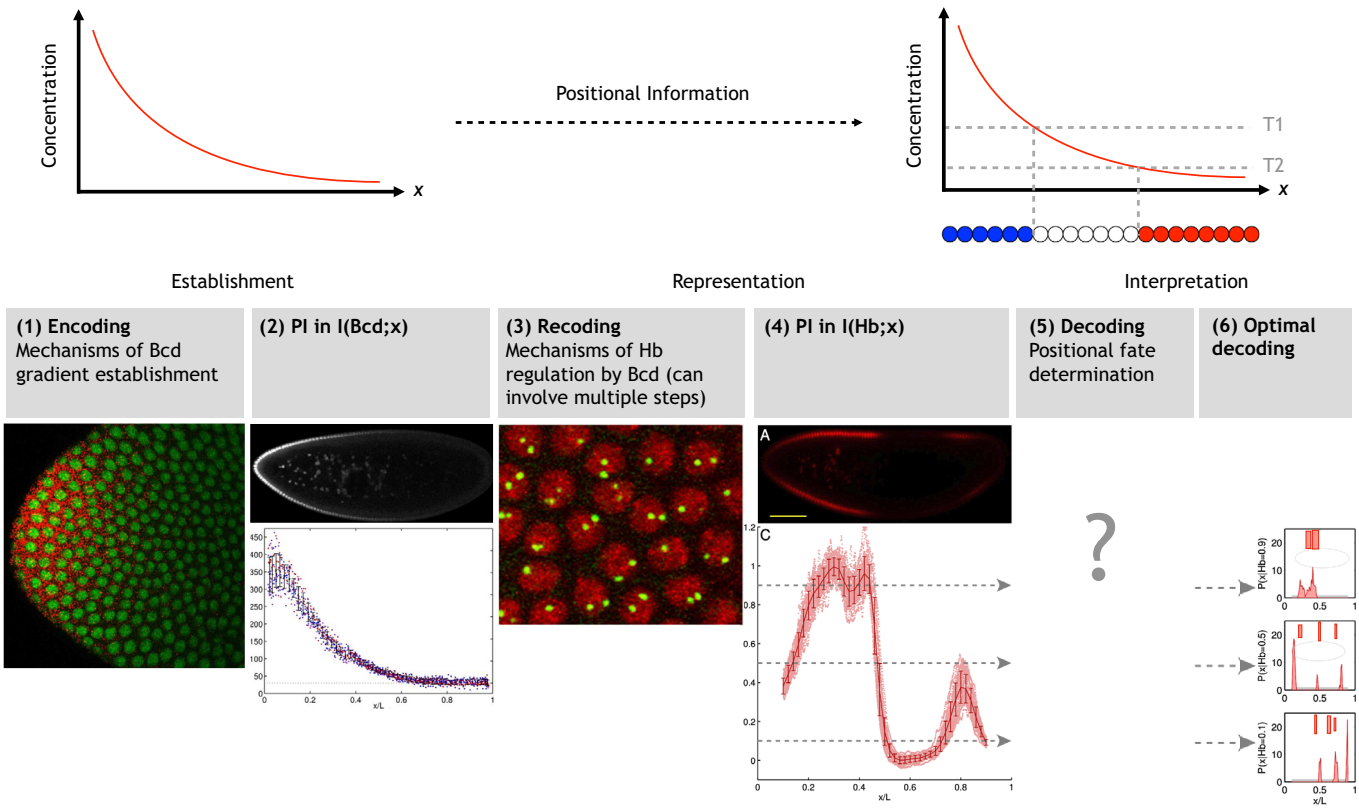


Fig. 1. A framework for positional information. In Wolpert's conception, 'positional cues' are provided by concentration fields of patterning chemicals, depicted here as a single morphogen gradient (top left) extending along the linear *x* dimension. These positional cues are then 'interpreted' by thresholds T1 and T2 (top right) that convey discrete cell identities (blue, white or red), resembling the famous French Flag model. We postulate an intermediate step in which the same information that is present in the morphogen gradient is put into other forms of 'representation', of which there can be several layers. Thus, PI undergoes multiple transformations, from establishment to recoding (potentially multiple times) to decoding. Steps that depend on biological mechanisms ('encoding', 'recoding' and 'decoding') can be separated from mechanism-independent abstractions (here, 'PI' and 'optimal decoding'). As an example, we use the gene expression cascade that patterns the anterior-posterior axis of the *Drosophila* embryo. During encoding (1), a gradient of Bicoid (green) is established from maternally deposited mRNA (red) at the anterior. Once established, it is possible to estimate the amount of PI in the Bcd gradient (2: top, Bcd-GFP-expressing embryo; bottom, nuclear concentration measured in individual nuclei along the AP axis) in a way that depends solely on the measured gradient but not on the mechanisms underlying its establishment. Bcd then regulates expression of the gap gene *Hunchback* (3: *Hb*, yellow: red, nuclei) and, as a result, information about position is transformed (or 'recoded') into the *Hb* profile. Once established, it is again possible to estimate the amount of PI in the Hb profile in a mechanism-independent way (4). Gap gene expression profiles are then somehow 'decoded' (5) by cells to determine their positions or cell fates in a way that depends on biological 'decoding' mechanisms; however, there is a single mathematically optimal way, which is mechanism independent (6: 'optimal decoding'), to estimate position from the morphogen profiles. Probability distributions (red) for three Hb concentration levels (gray arrows) determine where cells are located along the AP axis.

information as their primary morphogen inputs, and provide a complete 'coordinate system' allowing for precise positional determination. It is thus more pertinent to speak of transforming or recoding of PI that will be read out only at a later stage. Such transformations could happen multiple times, and each successive step should be tracked in a general mathematical framework. The concept of recoding is conceptually loosely related to Wolpert's original idea of 'positional value' (Wolpert, 1989).

A theoretical framework for PI that maps spatiotemporal concentration profiles to position must also consider stochasticity. Although patterning precision and reproducibility can be achieved over very short developmental time spans, using only a few handfuls of genes (Bentovim et al., 2017; Bollenbach et al., 2008; Briscoe and Small, 2015; Gregor et al., 2007; Houchmandzadeh et al., 2002; Patel and Lall, 2002; Petkova et al., 2014; Reeves et al., 2012), the processes underlying patterning are subject to molecular noise (Arias and Hayward, 2006; England and Cardy, 2005; Houchmandzadeh et al., 2005; Hu et al., 2010; Tkačik et al., 2008a; Tostevin et al., 2007; Tsimring, 2014; van Kampen, 2007). Moreover, there is random variability not only within a specimen, but also between specimens, e.g. in the strength of the morphogen sources (Bollenbach et al., 2008; Howard, 2012).

The necessity for a probabilistic approach is best exemplified when considering an undifferentiated cell in a developing organism. The cell experiences a single random realization of an otherwise variable information-carrying profile. When fluctuations between specimens or between adjacent cells of the same specimen are large, differences between cells can no longer be distinguished and PI is lost. This statement is true irrespective of the biological mechanism that reads out the gradient. It is a theoretical statement about what is possible in principle, which no biological (or engineered) system can evade. Thinking about what individual cells can measure locally – as in Wolpert's original concept – sharply contrasts with the typical approach to data analysis in biology, where one identifies 'statistically significant differences' in the mean gradient profile from one cell to the next, or where one disregards stochasticity by looking only at aggregated (averaged) profiles. A theoretical framework appropriate for Wolpert's PI concept therefore must be phrased in terms of probability distributions, not geometrically, as would be appropriate when dealing with shapes and patterns in the absence of noise.

Development (2021) 148, dev176065. doi:10.1242/dev.176065

Establishing a mathematical framework for PI

Information theory is the mathematical treatment of concepts, parameters and rules governing efficient and reliable transmission of messages through communication systems (see Box 1). It has been applied to biological problems (Tkačik and Bialek, 2016) but it was not until the late 2000s that ideas about information transmission appeared for biochemical networks (Bowsher and Swain, 2014; de Ronde et al., 2011; Mugler et al., 2010; Tkačik and Walczak, 2011; Tkačik et al., 2008c; Tostevin and Ten Wolde, 2009; Ziv et al., 2007), specifically for the anterior-posterior (AP) patterning gene network of the early *Drosophila* embryo (Tkačik et al., 2008b). These initial studies focused on computing how well fluctuations in some 'input' chemical signal (morphogen, transcription factor or ligand concentration) are encoded in the resulting 'output' gene expression levels, given that gene expression is necessarily subject to molecular noise of well-understood biophysical origins (Gregor et al., 2007; Tkačik et al., 2008a). At that time, molecular signals were only starting to be experimentally measurable at a single-cell level (Blake et al., 2003; Elowitz et al., 2002; Golding et al., 2005; Ozbudak et al., 2002; Raser and O'Shea, 2004; Rosenfeld et al., 2005).

To introduce information theory in the context of genetic networks, and as a vehicle for a mathematical framework for PI, we focus here on the example of the early *Drosophila* embryo. The general framework we develop can be generalized to other systems in a straightforward manner, but depends on the specific circumstances and constraints imposed by the different experimental setups. In the case of the *Drosophila* embryo, we postulate that it has evolved to 'send' or encode real physical coordinates x of cells or nuclei through a noisy biochemical reaction network that at different x generates different patterning molecule concentrations $\boldsymbol{g}$. Here, $\boldsymbol{g}$ represents morphogen concentrations, either primary gradients or subsequently expressed developmental genes (such as gap or pair-rule genes) – the mathematics remain the same. The concentrations $\boldsymbol{g}$ are denoted in bold face to indicate that there can be multiple relevant concentrations, and thus, formally, $\boldsymbol{g}$ is a vector at every position x. Because of noise, $\boldsymbol{g}$ is not a deterministic function of x, but we have to use a probability distribution $P(\boldsymbol{g}|x)$ that tells us the probability of finding a certain $\boldsymbol{g}$ at x.

Shannon's original formulation of information theory revolved around the concept of a noisy information channel (Shannon, 1948). A 'channel' here represents an evolved biochemical reaction network. It encodes different positions x into concentration levels $\boldsymbol{g}$, probabilistically, as described by $P(\boldsymbol{g}|x)$. Neither the concept of PI nor the channel concept depends on underlying mechanisms, but only on how input signals x are mathematically transformed into outputs $\boldsymbol{g}$. Biological mechanisms inside the channel are *de facto* treated as a black box. Information theory then introduces a general and unique measure of how well information can be sent through such noisy channels, the mutual information $I(\boldsymbol{g}|x)$ (Cover and Thomas, 2006):

$$I(\boldsymbol{g};x) = \left\langle \int d\boldsymbol{g}\, P(\boldsymbol{g}|x)\, \log_2 \frac{P(\boldsymbol{g}|x)}{P_g(\boldsymbol{g})} \right\rangle_x. \qquad \text{(Eqn 1)}$$

Angular brackets indicate an average over all locations x, assuming that cells or nuclei are uniformly distributed over the coordinate x. (See Dubuis et al., 2013b and Tkačik et al., 2015 for straightforward generalizations.) Similarly, $P_g(\boldsymbol{g})=\langle P(\boldsymbol{g}|x)\rangle_x$ is the average of the distribution of morphogen concentrations across all positions x; it represents the probability that a particular combination of concentrations, $\boldsymbol{g}$, can be seen anywhere in the embryo (Fig. 2).

Our key assertion can now be made precise: we claim that the mutual information [a mathematical object of information theory (Cover and Thomas, 2006)] linking position and morphogen concentration, $I(\boldsymbol{g};x)$, is the proper formalization of PI (a concept of developmental biology). The distribution of morphogen concentrations at a given position, $P(\boldsymbol{g}|x)$, can be estimated from experimental data (see Box 2), giving access to empirical measures of PI $I(\boldsymbol{g};x)$, which is mathematically derived from $P(\boldsymbol{g}|x)$ by Eqn 1. Although proper estimation from finite datasets requires care, the technical procedures have been documented elsewhere (Borst and Theunissen, 1999; de Polavieja, 2004; Strong et al., 1998; Tkačik et al., 2015). More pertinent for morphogenesis are the following characteristics of PI (summarized below and expanded in Boxes 3 and 4):

- PI is a unique measure of all statistical dependence between morphogen concentrations and position with important theoretical guarantees. It measures how well any variation of morphogen profile with position (linear or not) can be used to determine positional specification (Dubuis et al., 2013b). Thereby, PI satisfies property 1 (Fig. 3).
- PI is a single number with interpretable units. Intuitively, I bits of information (see Box 4) are necessary and sufficient to distinguish 2^I discrete alternatives with zero error (Hillenbrand et al., 2016); if some degree of positional error is allowed, I bits suffice to specify more alternatives (Tkačik et al., 2015). Thereby, PI satisfies property 2 (Fig. 4).
- PI is applicable to single or multiple morphogen gradients of arbitrary shapes, independently of the biological system and mechanistic detail. The framework does not single out particular profile shapes, positional markers or special positions. Thereby, PI satisfies properties 3 and 4 (Tkačik et al., 2015), also enabling a theoretical search through the space of all possible morphogen profiles to predict ones that maximize PI, thereby satisfying property 5 (Sokolowski and Tkačik, 2015; Tkačik and Walczak, 2011; Tkačik et al., 2009).

Within this theoretical framework, PI summarizes the fidelity by which position is encoded in any number of morphogen gradients of arbitrary shapes, independent of the system and biological mechanisms. While such a formalism employing a single statistic is undeniably attractive, its benefits come at a price (see also Box 5): a single number might measure the overall limits of patterning, but it cannot explain how and where these limits arise. Specifically, PI cannot answer local questions or make testable predictions about limits to patterning at individual positions within an embryo. To this end, the PI framework must be appropriately extended (see Box 6).

Decoding PI

An undifferentiated cell in a field of morphogen concentrations needs to determine its location by 'reading out' the available PI. It thus needs to perform local concentration measurements and estimate, or infer, its position. Early demonstrations of quantitative limits to this process (Gregor et al., 2007) were followed by the development of a rigorous mathematical framework for optimal decoding (Hironaka and Morishita, 2012; Morishita and Iwasa, 2009, 2011), which has since been applied to data and connected to information-theoretic concepts (Dubuis et al., 2013b; Petkova et al., 2019; Tkačik et al., 2015; Zagorski et al., 2017), as summarized in Box 6.

Suppose that the distribution of morphogen concentrations given position, $P(\boldsymbol{g}|x)$, is known. For example, an image collected in an experiment provides absolute knowledge about position, and multiple images can then deliver the probability of finding a

Development (2021) 148, dev176065. doi:10.1242/dev.176065

particular concentration at that position across a set of samples. If the cell measures one set of local morphogen concentrations, $\boldsymbol{g}$, to estimate its location, what would that estimate be and how precise could it be? Here, the true location of the cell, x (unknown to the cell, but known to the experimenter), needs to be clearly distinguished from the best estimate of the location that the cell might be able to extract from $\boldsymbol{g}$, denoted here as implied position, x^*.

Cells can extract x^* from morphogen concentration measurements by means of a decoding mechanism. Although many such mechanisms and their biological implementations are possible, there is a single decoding algorithm that is statistically optimal, leading to the best positional estimate, given by Bayes' law:

$$P(x^*|\boldsymbol{g}) = \frac{1}{Z} P(\boldsymbol{g}|x^*) P_x(x^*). \quad \text{(Eqn 2)}$$

On the right-hand side, we have the *a priori* distribution of locations (e.g. cell positions) to be decoded, $P_x(x^*)$, which for spatially uniformly distributed cells is a uniform distribution; $P(\boldsymbol{g}|x^*)$ is the measured distribution of concentrations introduced earlier; and a normalization factor Z enforces that the resulting posterior distribution $P(x^*|\boldsymbol{g})$ is correctly normalized.

The posterior distribution summarizes all knowledge about x^* that can possibly be extracted by measuring morphogen concentrations, $\boldsymbol{g}$. It is a distribution over implied locations, and there are multiple qualitative shapes that this distribution may take (Fig. 5). In scenario A, for a particular $\boldsymbol{g}$, the posterior may be sharply localized around a single peak $X^*(\boldsymbol{g})$, typically at the mean of the posterior distribution,

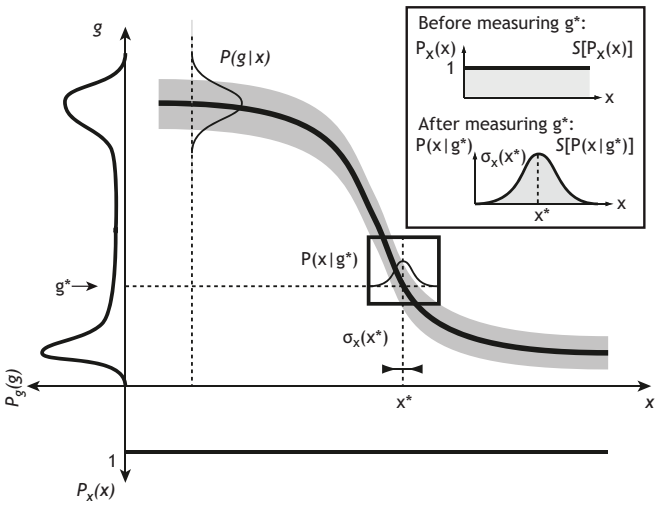


Fig. 2. A graphical illustration of the ingredients for PI. An example gene, g, makes a mean profile in coordinate x (thick sigmoidal black line), with the intrinsic variability denoted by gray shading. For each location x, gene expression levels are described by a distribution $P(g|x)$, depicted by a Gaussian centered on the mean profile with width $\sigma_g(x)$. Nuclei are spaced uniformly across x, as shown by the uniform distribution, $P_x(x)$, at the bottom; averaged across all these nuclei, the probability of observing a gene expression level g is given by $P_g(g)$ (distribution on left). Knowing a particular value of gene expression, g*, implies limited knowledge about position: very likely, the position is x*, but fluctuations in gene expression will give rise to positional error, σ_x(x*), around this position, as indicated. Inset: before making a gene expression measurement, our knowledge about position is zero and the distribution over possible locations is uniform; after an observation, the distribution over possible locations is much more localized and the uncertainty about position is smaller. PI measures the average reduction in uncertainty (mathematically quantified by the entropy, S, of a distribution) about position due to morphogen gradient observation.

Box 2. Measuring positional information

$P(\boldsymbol{g}|x)$ can be estimated experimentally: samples with simultaneously recorded concentrations $\boldsymbol{g}$ can be collected at every position x from many identical specimens. In biological systems, it is most common to focus on the mean or the 'mean spatial profile' in the case of the embryo. Thus, implicitly, the joint distribution is reduced (i.e. marginalized) to averages, $\overline{g_i}(x) = \int d\boldsymbol{g}\, g_i\, P(\boldsymbol{g}|x)$. Yet there is no fundamental reason to focus solely on averages. Crucially, retaining the variability in the profiles [mathematically given by $\sigma_i^2(x) = \int d\boldsymbol{g}\, (g_i - \overline{g_i})^2\, P(\boldsymbol{g}|x)$] is in fact necessary for a probabilistic approach. $P(\boldsymbol{g}|x)$ keeps all the information about concentration profiles, their variability and co-variability (for multiple genes), and even their higher-order statistics. Experiments that reliably sample this distribution are significantly more demanding than experiments that solely focus on measuring mean profiles, but this difficulty is technical rather than fundamental, and it can be be surmounted (Dubuis et al., 2013a; Petkova et al., 2019; Tkačik et al., 2015). A full protocol for the experimental procedures and the measurement error treatment to quantify PI in fly embryos can be found elsewhere (Dubuis et al., 2013a,b; Gregor et al., 2014; Tkačik et al., 2015). Here, we stress that, in order to test the theoretical formalism applied to PI, precision measurements are necessary. Such measurements are typical for testing theories in the physical sciences, but are still not the norm for biological systems.

$X^*(\boldsymbol{g}) = \int dx^*\, x^*\, P(x^*|\boldsymbol{g})$. Mathematically, this scenario is equivalent to the statistical inference of a 'parameter' x from noisy data $\boldsymbol{g}$ in the regime where the posterior is nearly Gaussian. In this case, the maximum likelihood estimate [assuming a uniform prior $P_x(x^*)$], the *maximum a posteriori* (MAP) estimate, and the posterior mean all coincide. Concentrations $\boldsymbol{g}$ accurately and unambiguously determine a single location, a hallmark of a good positional code. The decoding error, formally defined as the spread of the posterior around its mean, is low. In scenario B, a single maximum of the posterior exists, but the decoding error is large, implying that the set of morphogen concentrations $\boldsymbol{g}$ provides only weak evidence for a particular location and that, at these morphogen concentrations, the precise localization of morphological features is impossible. In scenario C, $P(x^*|\boldsymbol{g})$ peaks either around the location X^* that is very far from the true location x, or peaks at multiple locations X^*, and is thus not unique. In this case, essential errors or ambiguities in the positional code exist, with the morphogen concentrations $\boldsymbol{g}$ likely 'pointing' to either wrong or multiple locations.

Applied to a realistic biological scenario, the decoding of cellular location along the AP axis of the early *Drosophila* embryo, one can construct $P(\boldsymbol{g}|x)$ from many samples of wild-type morphogen profiles and their biologically relevant variabilities (Petkova et al., 2019). The measured $P(\boldsymbol{g}|x)$ are used in Eqn 2. Mathematically, any set of concentrations $\boldsymbol{g}$ can be inserted to decode the most likely implied position, $X^*(\boldsymbol{g})$. Biologically, however, the focus must be on those concentration combinations that are actually observed. This is a non-trivial point: if multiple morphogens $\boldsymbol{g}$ vary along a single positional axis, many combinations of $\boldsymbol{g}$ are unlikely ever to happen (at least in the wild-type embryo), and thus their decoded locations are irrelevant.

When a particular embryo is selected with a specific realization of morphogen profiles, $\boldsymbol{g}^\alpha(x)$ (not an average over embryos!), then these observed morphogen expression levels inserted into Eqn 2 will generate a decoding map for embryo α:

$$P_\alpha(x^*|x) = P(x^*|\boldsymbol{g}^\alpha(x)). \quad \text{(Eqn 3)}$$

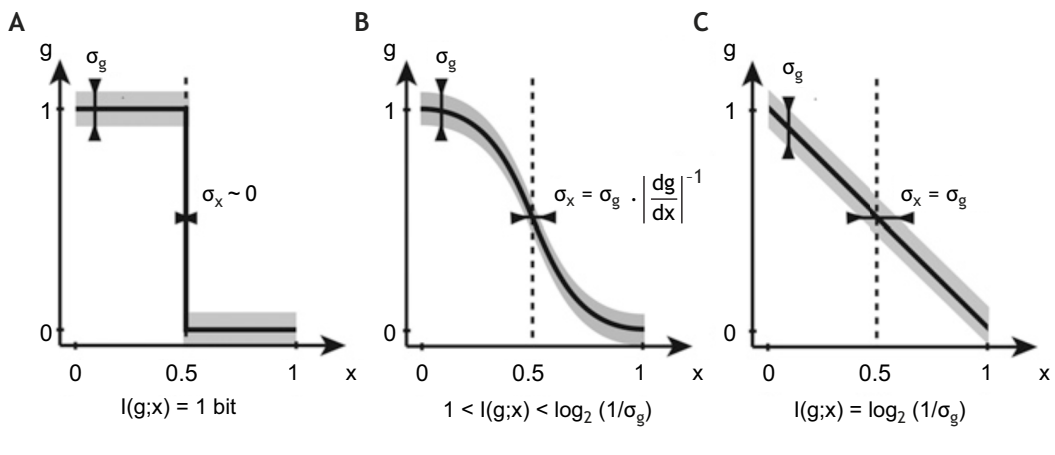
Eqn 3 represents a fundamental relationship between the real locations x in a single specific embryo α, and what is implied about these locations by the morphogen profiles, assuming optimal use

Development (2021) 148, dev176065. doi:10.1242/dev.176065

Box 3. Primer on theoretical formalism of positional information

Positional information (PI) measures any kind of statistical dependence between position *x* and morphogen concentrations ***g***. PI is zero only if there is no systematic variation in morphogen mean profile or any other statistic with position: in this case no mechanism exists to extract knowledge about position from morphogen concentrations (Cover and Thomas, 2006). Otherwise, positional knowledge can be extracted using a properly constructed decoding mechanism (which may, however, be biologically unrealistic). Even though linear gradients are often used as example cases, real gradients are not linear (and sometimes not even monotonic, e.g. for patterning 'stripes'); their variance typically changes with position (known in statistics as 'heteroscedasticity'); and their fluctuations may not be Gaussian, requiring a more powerful alternative to linear correlation.

As an example, the figure shows three gene expression profiles $g(x)$, with variability σ_g (shaded area). A step function (A) carries (at most) one bit of PI by perfectly distinguishing between 'off' (not induced, posterior) and 'on' (induced, anterior) states. A sigmoidal profile (B) has a wider boundary, but PI can be >1 bit because the transition region itself is distinguishable from the on and off domains. A linear gradient (C) has no boundary but increases PI by being equally sensitive at every value of *x*. In the absence of noise, B and C could theoretically reach arbitrarily high PI, as each concentration would correspond to a unique position without ambiguity. In reality, such infinities are avoided because the mapping is noisy and positions are discrete (e.g. columns of nuclei rather than physical coordinates with infinite precision).



Box 4. The meaning of 'bits'

Information is measured in bits, which are meaningful units: 1 bit of PI in the morphogen gradient suffices to make a reliable discrimination between two sets of positions that are, in the absence of morphogen readout, equally likely. For example, 1 bit of PI suffices to reliably discriminate the front half of the embryo from the back; or odd columns of cells from even columns. More generally, *I* bits of information are necessary and sufficient to distinguish 2^I discrete alternatives with zero error. Thus, the patterning of an embryo with *N* columns of nuclei that need to be uniquely distinguished with no possibility of error requires at least $I_0=\log_2 N$ bits of PI. If some error in specification can be tolerated, the required amount of PI is smaller than I_0. More PI can be provided (usually at a higher metabolic or time cost) to compensate for the decoding processes that do not use the information optimally. If the morphogens provide $I<I_0$ bits of PI, a minimal error exists by which cells can determine their positions: they can do worse (perhaps due to biological limitations in their gradient readout) but not better.

PI and the associated bounds to positional error provide a powerful and unbiased tool for asking biologically-relevant questions. How much additional PI is provided by each morphogen gradient in systems with multiple gradients? Are their individual PI contributions additive, redundant or synergistic? How much information is there in non-monotonic profiles (such as stripes) and how much information does each profile 'feature' contribute, especially when the features can be generated *in silico,* or isolated *in vivo* through appropriate genetic modifications? PI can be computed for various morphogen profiles (e.g. a sharp step, or a linear or exponential ramp) and compared with data, to question whether our expectations about 'ideal' shapes align with reality. Ultimately, morphogen profiles can be computationally optimized to find those that maximize PI, thus deriving the best morphogen patterns *ab initio*, and comparing such first-principle theory predictions with data.

('optimal decoding') of PI. The decoding map can be visualized as a matrix of implied versus true locations in the embryo (Fig. 6). A precise positional code, corresponding to scenario A discussed above, will result in $P_\alpha(x^*|x)$, which is tightly localized around the diagonal where $x^*=x$. Here, positions implied by noisy morphogen profiles are almost equal to the true, ideal positions known to the experimenter. Scenario B, with high positional error, corresponds to situations where at some location x, the decoding map has a single but broad, or 'diffuse', range of locations x^* that are consistent with the measured morphogen profiles. Scenario C typically corresponds to the situations where, at multiple locations, at least two separated peaks of implied positions x^* exist, and where cells cannot unambiguously determine whether they reside in one or the other peak (Fig. 7A).

The decoding map is a very powerful construction: it predicts the ability of a cell to determine its position locally and specifically, at a chosen real location x within a particular embryo α. By determining how the best estimate of position, i.e. the peak X^* of the map at every position x, varies between embryos, it predicts how embryo-to-embryo variability maps into uncertainty in specifying position estimates. By averaging individual embryo decoding maps across all embryos α of the same class, one can obtain an average decoding map $P(x^*/x)$ that, for wild-type embryos in scenario A, defines the positional error, $\sigma_x(x)$, as a function of real position x:

$$\sigma_x^2(x) = \int dx^* (x^* - X^*(x))^2 \, P(x^*|x), \qquad \text{(Eqn 4)}$$

where $X^*(x) = \int dx^* \, x^* \, P(x^*|x)$. This positional error quantifies how precisely positional markers can be localized in the embryo

Development (2021) 148, dev176065. doi:10.1242/dev.176065

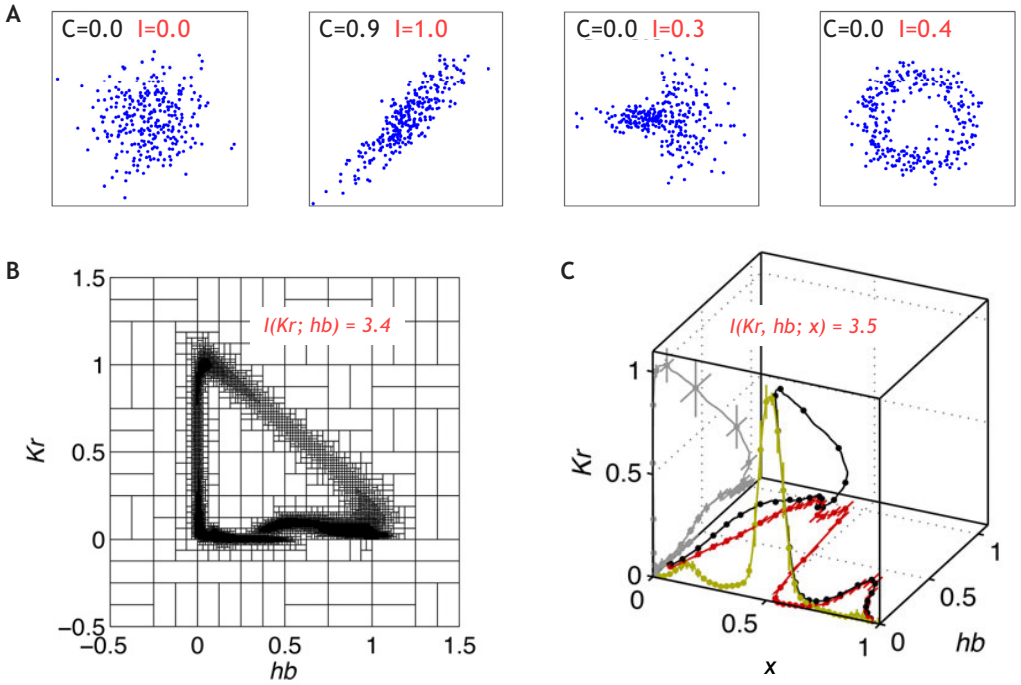


Fig. 3. Information as a measure for statistical dependence. (A) Four examples in which points (*x* and *y*), depicted in the plane as blue dots, were drawn from joint probability distributions, *P(x,y)*, with varying types of statistical dependency between *x* and *y*. C (black) denotes linear (Pearson) correlation coefficient, whereas I (red) denotes mutual information (in bits) between *x* and *y* for each of the cases. In the first panel, *x* and *y* are statistically fully independent. In the second panel, *x* and *y* are linearly correlated. In the third panel, the conditional average of *y* at a given *x* is constant, but for small values of *x*, the variance in *y* is smaller than for large values of *x*. Linear correlation fails to detect any kind of dependence, even if the number of samples is infinite; in contrast, mutual information is non-zero. In the fourth panel, *x* and *y* lie on a circular manifold, with zero linear correlation and non-zero mutual information. (B) Depiction of the joint probability distribution between measured expression levels of *Kruppel* (*Kr*) and *Hunchback* (*hb*) in *Drosophila* embryos; denser tiling represents higher probability weight. Such joint dependence (reminiscent of the fourth panel in A) leads to a small linear correlation, but 3.4 bits of mutual information. (C) As anterior-posterior position in the embryo, *x*, varies along the horizontal axis, two gap genes *hb* and *Kr* trace out a trajectory in the *y*, *z* coordinate space, as indicated in this 3D plot (black line; the yellow and red lines show projections on the sides of the cube that represent the profiles of Kr and Hb, respectively, separately). This strongly nonlinear joint dependence can be quantified by PI, showing that *Kr* and *hb* together encode *I(Kr,hb;x)*=3.5 bits about position; a linear measure such as a correlation coefficient would clearly fail to properly capture all observed statistical dependencies.

(Fig. 7B). For example, if wild-type embryos are known to express a positional marker at some position x based on morphogen readout, this framework states that, except for some residual experimental error, the positional accuracy of a marker across embryos is bounded from below by the positional error, $\sigma_x(x)$, at that position. $\sigma_x(x)$ thus quantifies the minimal uncertainty about the implied cellular location due to the combined variability and intrinsic noise in the morphogen profiles (Morishita and Iwasa, 2011; Tkačik et al., 2015).

Optimal decoding is particularly relevant in the context of mutations that affect a patterning system. Here, the decoding map $P(x^*|x)$ becomes a mathematical and quantitative formalization of the classical concept of a fate map (Conklin, 1905; Gilbert, 2000; Schüpbach and Wieschaus, 1986). Often, a mutation has consequences for the entire morphogen system, causing a global shift in the decoding map $P(x^*|x)$. In this case, the decoding map predicts how physical locations in the mutant (x) map to cell fates that are characteristic of the location in the wild type (x^*). But within a probabilistic framework there are other possible outcomes, implying that the decoding map can accommodate a richer set of possibilities than a traditional fate map. For example, there could be multiple peaks in x^* for some fixed position x in the mutants, predicting large mutant-to-mutant variability, where the same wild-type positional marker is placed at different, random positions x^* that correspond to the multiple peaks in the mutant.

The decoding map can thus make parameter-free predictions derived solely from wild-type embryos about how patterning mutants behave. Its only assumption is that a very good approximation to optimal decoding of Eqn 2 has evolved in the biological 'hardware'. This is an information-rich, quantitative and falsifiable prediction that can be viewed as the test of the optimality assumption, which, to date, has been experimentally verified with high fidelity in the *Drosophila* AP patterning system (Petkova et al., 2019) and for the mammalian neural tube (Zagorski et al., 2017).

Lessons for biology

By combining our mathematical framework for PI with applicable quantitative measurements, we can gain novel biological insights into patterning events, as summarized below.

Optimal patterning without sharp boundaries

Within the original paradigm for PI, morphogen profiles are 'read out' by downstream genes to guide cell fate decisions. Is there a notion of a best profile shape that supports reliable fate determination? Theoretical work typically considers linear profiles; in contrast, maternal morphogens often exhibit exponentially decaying profiles that span a significant fraction of the length of an embryo. Yet other patterning genes may show very sharp gene expression boundaries (Fig. 5). The theory of PI can guide us on what the best profile shape is for encoding a maximum amount of PI. Perhaps surprisingly, the answer depends on how variability (i.e. noise) changes with position. If variability is independent of position

Development (2021) 148, dev176065. doi:10.1242/dev.176065

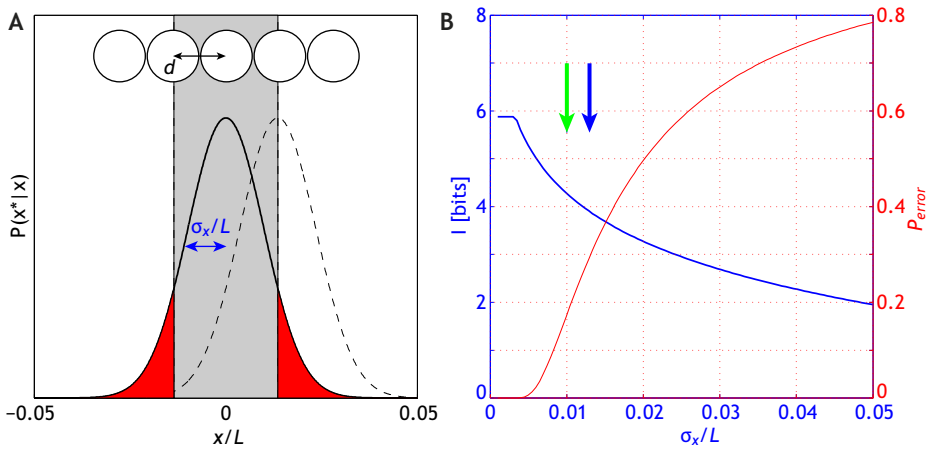


Fig. 4. Positional error and PI. (A) Schematic representation of a row of nuclei (top), equally spaced *d* µm apart, that tile the length *L* of the embryo axis. A positional error of σ_x at the location of a focal nucleus at $x = 0$ implies that, given the limited information contained in morphogen profiles, the focal (and its neighboring) nucleus can only estimate its position as depicted by a solid (dashed) Gaussian curve. Whenever that estimate falls outside the gray band, the identity of the focal nuclei will likely be mistaken for its neighbors; the probability of this happening, P_{error}, is represented in red and can be easily computed from the Gaussian distribution. (B) Relationship between positional error (the width of the Gaussian curve in A), PI (left axis) and the probability of error (right axis). In the limit when positional error vanishes, the information saturates at $log_2(N)$ bits, where *N* is the number of nuclei tiling the embryo axis and the probability of error becomes zero: this is the error-free positional code. As positional error increases, PI must decrease and the probability of mistaken nuclear identity rises. Blue arrow corresponds to the example depicted in A; green arrow to the estimated 1% positional error (and 4.3 bits of PI) reported by Dubuis et al. (2013b) for the gap gene system in *Drosophila*.

and is low compared with the maximum gene expression magnitude, then the optimal profile is linear. In this case, a single profile can encode more than one bit of PI.

In biochemical networks, however, the noise magnitude typically changes with position. Intrinsic noise, e.g. fluctuations in morphogen levels, depend on the mean morphogen concentration, and thus on position. This is true empirically and is expected on biophysical grounds, because, when morphogen concentrations are low, noise at these concentrations is ultimately Poissonian and its variance scales linearly with the mean. In this case, the optimal shape can be computed from the noise profile, and is typically not a linear one. Last, when noise is large, PI drops to below one bit, where even a trivial discrimination of location, such as between the front and back of the positional axis, can no longer be error free. Generally, with noise being low enough, most of PI is encoded in the smooth slopes of a (monotonic) profile; with high noise, slopes cannot be read out precisely and PI is reduced to the binary discrimination of being below or above an expression threshold (Tkačik et al., 2008b,c, 2015). This insight parallels the discussion in neuroscience on the optimal shape of tuning curves of sensory neurons (Butts and Goldman, 2006).

Box 5. Limitations of the framework

Patterning dynamics

Although it is possible to mathematically extend the PI framework to cases where PI is encoded in temporal trajectories of morphogen concentrations, this has not been tried in practice. In the *Drosophila* example considered here, information is stored in a single static snapshot of gene expression patterns, which greatly simplifies the technical analyses and their interpretation.

Positional coordinate

The theory is agnostic about how 'position' *x* should be represented to compute PI, $I(\mathbf{g};x)$. In the *Drosophila* example considered here, *x* is a relative coordinate along the anterior-posterior axis of the embryo. This choice relies on the finding that demonstrated spatial scaling of the morphogenetic patterns in this system (Gregor et al., 2005; Houchmandzadeh et al., 2002). An absolute coordinate *x* would thus be less appropriate. Nevertheless, a relative coordinate is not the only possible choice: *x* could also be a discrete nuclear column index. In contrast, it is much less is clear how to choose a representation for position in a growing or deforming tissue: should position be taken at a particular temporal snapshot or perhaps relative to a constantly co-moving and growing reference frame? Although the theory can be applied in either case, it does not provide us with an answer about the positional coordinate system.

How much of the information is biologically relevant?

Information-theoretic definition for PI has many attractive mathematical properties, but it does not tell us how many bits can actually be extracted from single morphogen snapshots with biologically plausible mechanisms. One can imagine gene expression patterns that formally carry a lot of PI, but the interpretation of which would likely require unrealistic computations.

Patterning genes are more than binary ON/OFF switches

Hunchback (Hb), a gap gene involved in *Drosophila* AP patterning, primarily responds to a gradient of maternal Bicoid, resulting in an expression profile that makes a seemingly sharp transition between high expression ('ON' domain) in the anterior half of the embryo and low expression ('OFF' domain) in the posterior half (Albert and Othmer, 2003; Alberts et al., 2002; Meinhardt, 1986; Spirov and Holloway, 2003). Hb has been the paradigm of a switch-like gene whose threshold is positioned precisely and reproducibly across embryos, roughly at the half-way point of the axis of the embryo (Crauk and Dostatni, 2005; Gregor et al., 2007; Holloway et al., 2006; Houchmandzadeh et al., 2002). Switches are expected to encode, at most, one bit. Surprisingly, our model-free estimates of PI reveal empirically that Hb encodes almost 2.2 bits of PI, indicating that the switch-like approximation would miss more than half of the available information, vastly underestimating the capacity of this patterning system (Dubuis et al., 2013b; Tkačik et al., 2015). The extra bit comes from the fact that Hb expression, although steep, is not a step function; indeed, about one third of the nuclei experience intermediate levels of expression, clearly distinguishable from the ON or OFF states.

Similar values have been reported for other gap genes in the early *Drosophila* embryo. Together, the four trunk gap genes provide ∼4.2 bits of PI, enough to specify every nucleus in the

Development (2021) 148, dev176065. doi:10.1242/dev.176065

Box 6. Positional information, positional error and decoding maps

We have introduced the concepts of PI (Eqn 1), positional error (Eqn 4) and the decoding map (Eqn 3) (Fig. 7). PI is entirely agnostic to encoding and decoding mechanisms, and is a single number expressed in bits that characterizes the global performance of the patterning system. Positional error and the decoding map are local constructs that characterize the performance of the patterning system location by location, but assume statistically optimal readout of the morphogen profiles. The positional error can be derived from the average decoding map in Eqn 4 and, under the assumptions of scenario A (Fig. 5A), has a clear biological interpretation.

The precise relationship between PI and the two decoding-related quantities is technically involved, but two generic statements hold universally. First, from the fundamental theorem of information theory known as the Data Processing Inequality (DPI) (Cover and Thomas, 2006), we can assert that, regardless of the chosen decoding algorithm (e.g. Eqn 2), PI is always greater or equal to the mutual information between the true locations and the best estimates of position (Brunel and Nadal, 1998). In other words, PI is an upper bound to the information between true and implied positions.

Second, when every encountered combination of morphogen concentrations $\boldsymbol{g}$ at true location x decodes to a single peak in the posterior X^*, the width of which is given by positional error, $\sigma_x(x) \ll L$, the approximation holds:

$$I(x;x^*) \approx -\langle \log_2 \sqrt{2\pi e\, \sigma_x^2(x)/L^2} \rangle_x, \qquad \text{(Eqn 5)}$$

where L is the range of x over which the patterned cells are uniformly distributed. As the DPI has general validity, Eqn 5 will always bound PI from below; but as the positional error shrinks and the posterior approaches a Gaussian distribution (as in scenario A), Eqn 5 will also be a good approximation for PI $I(\boldsymbol{g};x)$. Indeed, for the case of *Drosophila* anterior-posterior patterning, the direct estimate of the PI, $I(\boldsymbol{g};x)$, and the decoding estimate from positional error, $I(x;x^*)$, differ by only ∼0.1 bit out of 4.3 bits, a discrepancy of ∼2% (Dubuis et al., 2013b). This agreement is a quantitative consistency check that the gap gene system of wild-type *Drosophila* embryos indeed forms a precise, unambiguous positional code in which positional error is small and nearly Gaussian almost everywhere.

central 80% of the AP axis of the embryo with only ∼1% positional error. This precision is completely inaccessible if each gap gene provides at most one bit of PI (Fig. 8). Distinguishing between the binary or analog character of these gene expression profiles thus clearly necessitates a quantitative analysis framework.

The role of spatiotemporal averaging during patterning

What can regulatory circuits do to mediate the impact of noise intrinsic to chemical reactions taking place at low molecule copy numbers? Cells can reduce the impact of such noise by performing many noisy concentration measurements of morphogen molecules and then averaging across them. This averaging can happen either over time or across space. But although these mechanisms are thought to play an essential role, they are subject to biophysical limits. Temporal averaging is in tradeoff with dynamics: regulatory circuits with long timescales that can average their inputs imply a slowdown in response dynamics (which may be undesirable) and require temporally stable morphogen inputs. Spatial averaging is in tradeoff with sharp spatial gradients: noise can be reduced if morphogen inputs are nearly constant over the spatial averaging window, but if the averaging window is larger, it will 'flatten out' information-carrying morphogen gradients.

In *Drosophila*, PI carried by the Bicoid gradient [$I(Bcd;x) \sim 1.6$ bits] is roughly equal to the mutual information between Bicoid and Hunchback [$I(Bcd;Hb) \sim 1.5$ bits], yet considerably lower than the PI carried by the spatial profile of Hunchback alone [$I(Hb;x) \sim 2.3$ bits], even though Hunchback is downstream of Bicoid (Dubuis, 2012). However, according to a naïve application of the Data Processing Inequality (DPI; Box 6), if concentration levels $\boldsymbol{c}$ serve to locally regulate the expression of downstream genes $\boldsymbol{g}$, the PI in $\boldsymbol{g}$ should be less than in $\boldsymbol{c}$, $I(\boldsymbol{g};x) < I(\boldsymbol{c};x)$. How then can empirical observations for Bicoid and Hunchback be reconciled with the DPI?

One possibility is that Hunchback receives additional PI from inputs other than Bicoid, although a strong and precise Hunchback boundary is observed in mutants deficient in AP morphogens aside from Bicoid (Petkova et al., 2019). Another possibility is that PI carried by Hunchback is higher because of the spatiotemporal averaging performed over Bicoid concentration by the Hunchback readout mechanism (Gregor et al., 2007; Little et al., 2013; Zoller et al., 2018). Hence, a local, instantaneous measurement of

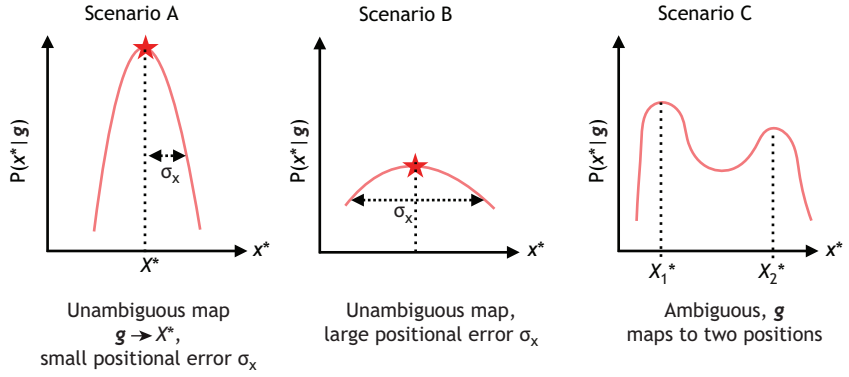


Fig. 5. Three possible decoding scenarios. Given the observed gene expression profiles $\boldsymbol{g}$, the posterior over likely position in the embryo, x^*, peaks sharply around the particular value X^* (scenario A). Sharp localization implies a small positional error, and it is possible to decode using a 'dictionary' or a 'lookup table', $\boldsymbol{g} \rightarrow X^*$. This is a hallmark of a good positional code if it can be performed for all values of gene expression $\boldsymbol{g}$ typically observed in the wild-type embryo. A defect in precise positional code will be observed (scenario B) when the posterior over likely positions in the embryo does not sharply peak but is 'diffuse'. Although it is formally possible to identify a single location as the peak of the posterior, the large spread around the peak implies a high positional error. Another type of defect in the code happens when the posterior over position does not even have a single peak given the gene expression profiles (scenario C). In this case, essential ambiguity exists in the positional code, and the gene expression levels would map to two distinct locations, X_1^* or X_2^*. Experimentally, this could predict a bimodal population of embryos.

Development (2021) 148, dev176065. doi:10.1242/dev.176065

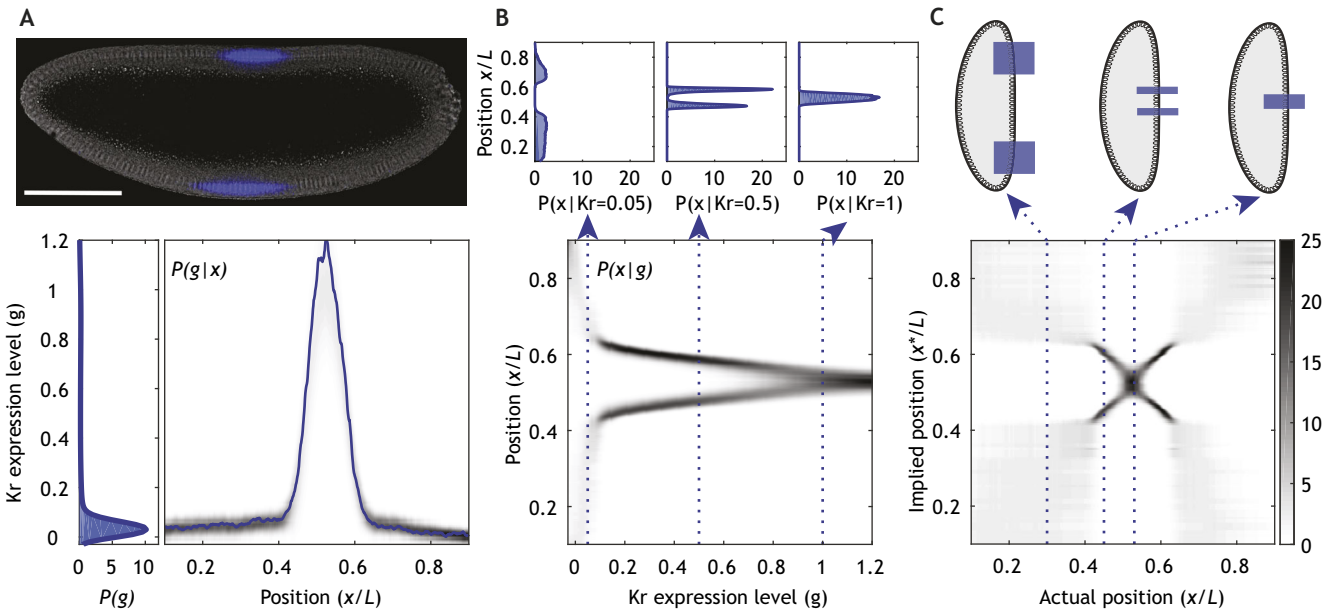


Fig. 6. Step-by-step construction of a decoding map. (A) Top: fluorescence image of a fixed *Drosophila* embryo at roughly 2.5 h of development. Fluorescence represents gene expression levels of the morphogen Kruppel (Kr), revealed through antibody staining (blue). Scale bar: 100 µm. Bottom: fluorescence intensity profile extracted from the top embryo (blue) and its probability of occurrence (gray shading) summarized as a distribution *P(Kr|x)*, constructed from a set of 37 similar embryos. The shaded probability weight intuitively accounts for the standard deviation of Kr expression at each position, *x*. The distribution of Kr expression, integrated over the whole embryo, is shown on the far left. (B) Bottom: the posterior of distributions over position at given levels of Kruppel expression, following Eqn 2. Posterior positions at three *Kr* expression levels are shown at the top: at low Kr, the posterior over positions is both ambiguous and broad – low Kr expression levels carry essentially no information about position (except that the likely position is not in the middle of the embryo); at medium Kr, the posterior is localized into two sharp peaks, but still ambiguous, as Kr alone does not specify whether the encoded location is at the right or left flank of the Kr peak; at high Kr, the posterior is sharply localized to the middle of the embryo and positional encoding here is good. (C) A decoding map, constructed by inserting Kr expression levels from a specific single embryo is shown. For each true position *x* that the experimentalist can measure (horizontal axis), the decoding map shows a posterior or a distribution over implied positions, *x**, that would be consistent with the Kr expression observed at *x*. The decoding map is unambiguous at the center (peak of Kr), ambiguous but precise at the flanks, and diffuse and ambiguous in the far anterior or posterior.

Hunchback is in fact a function of the temporal and spatial history of Bicoid. DPI applies when $\boldsymbol{c}$ and $\boldsymbol{g}$ correspond to complete spatiotemporal patterns of Bicoid and Hunchback, but not necessarily when they are local instantaneous values. Thus, the biophysical mechanisms of spatial and temporal averaging increase the local instantaneous PI in the Hunchback profile. Temporal averaging is achieved through gene expression dynamics (Tkačik et al., 2008a) and spatial averaging through diffusion of the regulated gene product (Erdmann et al., 2009; Gregor et al., 2007; Little et al., 2013; Sokolowski and Tkačik, 2015).

PI quantitatively predicts number of unique cell fates

The values of PI are not only comparative (i.e. between morphogen profiles) but also have absolute meaning (Box 3). If unique identities for N cells have to be conferred without error, $\log_2(N)$ bits of PI are required. Typically, biological systems can tolerate some positional error (e.g. cell width sets a limit to positional accuracy), and thus a smaller number of bits of PI is required. For example, during nuclear cycle 14, *Drosophila* embryos have about 60 columns of nuclei in the central 80% of the AP axis, implying that at least $\log_2(60) \sim 5.9$ bits of PI would be needed for error-free unique specification of each column. However, if the tolerated positional error is ∼1%, then ∼4.3 bits are sufficient (Fig. 4), which corresponds precisely to the physical distance expressed in terms of embryo length between two adjacent cells (Dubuis et al., 2013b). Thus, interpreting absolute values of PI is a simple, yet powerful concept, free from the arbitrariness of normalization procedures, null-model formulations and aesthetic or philosophical decisions about what constitutes ‘precise’ or ‘imprecise’ patterning. The cost of extracting an absolute value, however, comes with the requirement that the measurements themselves are precise, are systematically unbiased and are in a regime in which intrinsic biological noise – and not experimental or statistical noise – is largely the dominant source of the measured variance (Box 5).

Threshold-free positional cues from multiple combined patterning systems

Although conceptually simple, a threshold-dependent concentration readout process is problematic (Houchmandzadeh et al., 2002; Jaeger et al., 2004): the concentration of the signaling molecules is often very low, resulting in very high concentration noise levels (Gregor et al., 2007). These concentration fluctuations propagate to downstream genes, and a reliable outcome of the morphogenic process would be questionable if it is implemented via sharp thresholding (Lacalli and Harrison, 1991).

Several hypotheses exist to explain how cells can integrate PI from single morphogen gradients without thresholding or from multiple morphogen gradients. For example, cells could sense the difference or ratio of two opposing morphogen gradients (Houchmandzadeh et al., 2005; McHale et al., 2006), compare concentration values at two nearby spatial locations and thus estimate the local gradient (Mugler et al., 2016), or respond to temporal dynamics of the morphogen (Bergmann et al., 2007; Cepeda-Humerez et al., 2019). However, given the shapes and variabilities of gradients, the only statistically optimal possibility is the maximum *a posteriori* (MAP) decoding rule (Eqn 2). Recent analyses of the *Drosophila* embryo (Petkova et al., 2019) and the

Development (2021) 148, dev176065. doi:10.1242/dev.176065

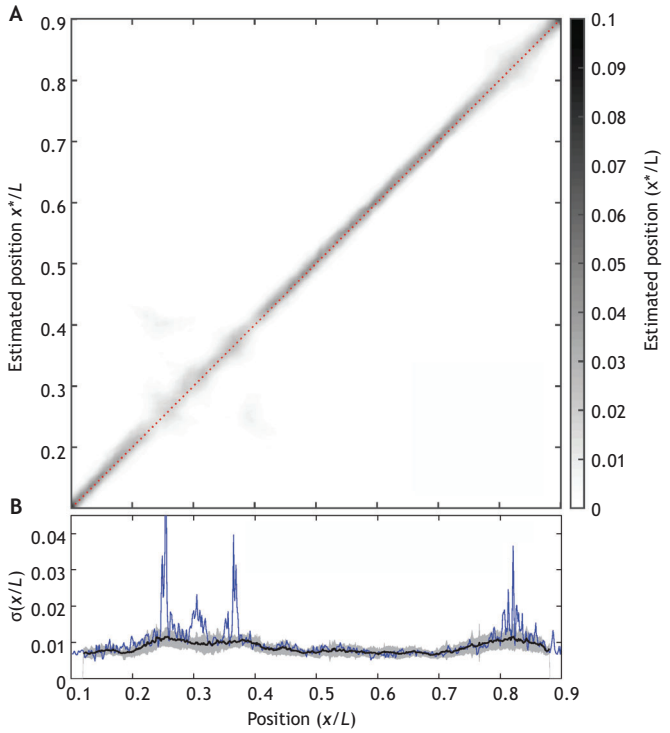


Fig. 7. Precise decoding from four *Drosophila* morphogens. (A) A generalized decoding map for four morphogens called 'gap genes' (Hb, Gt, Kni and Kr) in the early *Drosophila* embryo (Petkova et al., 2019), containing 4.3 bits of PI (Dubuis et al., 2013b). Gray shading represents the probability that a cell located at a given *x*-axis value decodes its position at a *y*-axis value along the AP axis of the embryo. (B) Comparison of positional error σ_x defined in Fig. 4 (blue) with Gaussian fitted width of gray diagonal band in A, showing equivalence between the concept of a decoding map and that of positional error.

vertebrate neural tube (Zagorski et al., 2017) have shown that biological results are indeed consistent with a statistically optimal readout of PI provided by multiple patterning cues. Thus, the mathematical framework for PI generalizes naturally to patterning by multiple gradients.

While Wolpert's prescription of applying thresholds to a single gradient is intuitive, it is unclear how to extend it to multiple gradients. In particular, an obvious generalization whereby cells apply independent thresholds to each gradient suffers from three fundamental problems. First, after applying multiple thresholds, how can a readout decision be computed from the resulting set of thresholded (binary) values (Fig. 8); second, how could such computation be implemented in biophysical circuitry; and third, why would thresholding of multiple gradients be optimal? Statistically optimal decoding of Eqn 2 is free of such *a priori* constraints, and suggests that independent thresholds typically are not optimal. The exact connection between the mathematical structure of optimal decoding and the mechanisms that fulfill it remain to be determined.

Evolution can drive patterning systems towards theoretically optimal performance

How far did evolution drive a real patterning system towards the mathematically optimal patterns that maximize PI? The same framework that allows us to estimate PI from real data also allows us to formulate an optimization theory to search for optimal patterns and to predict empirically observable signatures of optimal patterning. One of the salient predictions of such a theory has been the constancy of the positional error $\sigma_x(x)$ across position for uniformly distributed cells/nuclei (Dubuis et al., 2013b; Tkačik et al., 2015). In the *Drosophila* embryo, this prediction is remarkably well matched by the data (Fig. 7B). Similarly, the theory of optimal information transmission quantitatively predicts the distribution of Hb expression levels from Hb expression noise (Tkačik et al., 2008b), which has been confirmed experimentally (Tkačik et al., 2008a). Optimal decoding – but not other schemes that map developmental gene expression levels into estimates of position – correctly predicts developmental consequences in *Drosophila* mutants for maternal morphogens, based solely on wild-type data (Petkova et al., 2019).

Such evidence suggests that evolution can drive patterning systems towards theoretically optimal performance. The question of whether the biological systems are 'at' or 'near' optimality is an interesting empirical question about the strength of evolutionary pressure to use limited resources in an efficient manner in a given population. Far away from optimality, where PI is small, biological function can simply not be supported irrespective of the resource availability, leading to malformation or death, as in some patterning mutants. What remains to be seen is whether such an optimization principle is powerful enough to quantitatively predict the entire set of spatial patterning gene expression profiles *ab initio*, thus leading to a potential design principle for the observed wild-type system (Tkačik and Walczak, 2011). The success of this approach depends precisely on how close to optimality evolution has driven a particular patterning system, and whether near-optimal solutions can also be explored mathematically.

Open conceptual puzzles

Abstractions and graphical visualizations of regulatory networks have been fundamental in allowing cross-species studies (Davidson, 2002; Gerstein et al., 2012). In a similar manner, a mathematical framework for PI should allow us to analyze and quantitatively compare different patterning systems to find evolutionary convergence or divergence in their function. How many bits are provided by different patterning systems and how is their decoding precision distributed across space? Does PI depend on the number of specified cell types, on the number of system components (e.g. genes or gradients), or perhaps some (more qualitative) notion of patterning complexity? In parallel to these direct applications, the framework also enables us to revisit several fundamental questions that we highlight below.

Is PI encoded by temporal dynamics of developmental genes?

During AP patterning of the *Drosophila* embryo, PI can be encoded in a single temporal snapshot of gene expression patterns: it has been empirically shown that a single snapshot of gap gene expression is sufficient to provide the PI required to quantitatively decode the positions of striped patterns of pair-rule genes with the precision that matches natural reproducibility (Petkova et al., 2019). Nevertheless, the temporal dynamics leading up to this snapshot are essential for bringing about this instantaneous state. Moreover, information could be directly encoded in these transient dynamics (Granados et al., 2018), e.g. in temporal trajectories of morphogen concentrations, ***g(t)***, at different spatial locations (Heemskerk et al., 2019; Rushlow and Shvartsman, 2012; Villoutreix et al., 2017). A rise and subsequent fall in morphogen concentration with time could designate a different position than a fall followed by a rise, even though the average initial and final morphogen concentrations were identical. The mathematical framework is readily generalizable for such a case: optimal decoding would be carried out using full temporal trajectories of morphogen concentrations, ***g(t)***. Extension of the framework to intrinsically

Development (2021) 148, dev176065. doi:10.1242/dev.176065

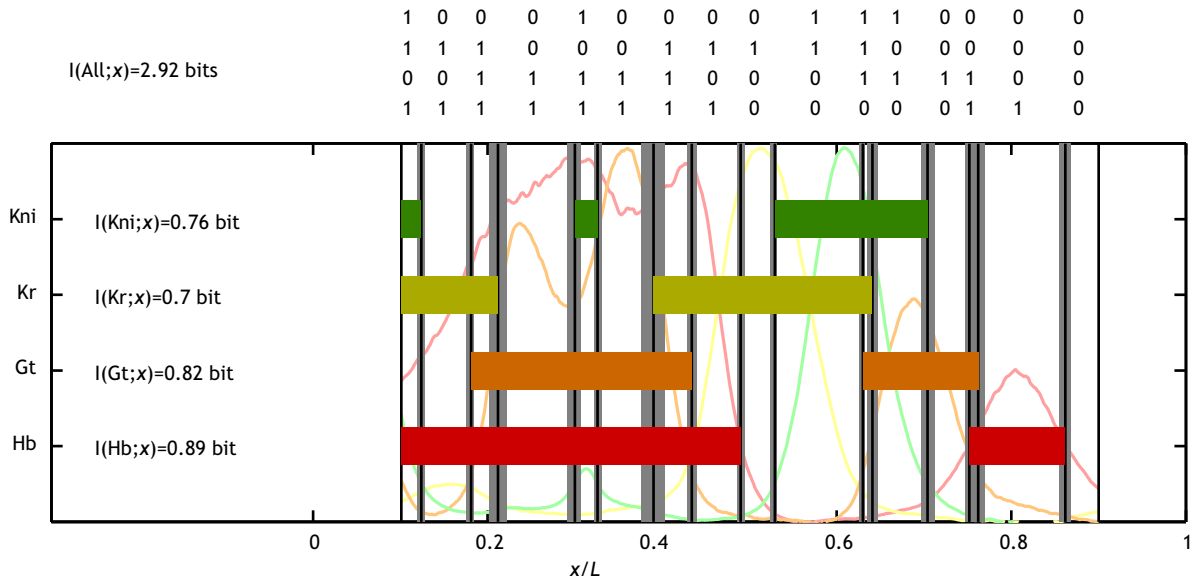


Fig. 8. Binary view of the gap gene system. For each gap gene *i* (indexing the four trunk gap genes of the anterior-posterior patterning system in *Drosophila*: Knirps, Kruppel, Giant and Hunchback), the expression level is quantized such that the gene is ON (1) if g_i is greater than a threshold θ_i and OFF (0) otherwise. The resulting domains of gene expression are shown with colored bars. The fluctuations of the domain borders are shown in gray for a set of thresholds that maximizes the total information carried by the binary variables (θ_{Kni}=0.125, θ_{Kr}=0.05, θ_{Gt}=0.1, θ_{Hb}=0.2). For reference, the mean expression profiles of the gap genes are plotted with lines in the background. Information carried by the quantized profiles of the individual genes is shown on the left. The joint pattern of gap gene activity at each position is represented by a four-digit binary code (shown at the top, with the bits representing Kni, Kr, Gt and Hb from top to bottom). The total information encoded jointly by the ON/OFF model of gap gene expression is, at most, 2.92 bits, significantly below the 4.3 bits carried by full expression values before thresholding.

dynamic processes could be particularly relevant to vertebrate somitogenesis, where system growth and patterning are dynamically highly intertwined (Oates et al., 2012).

Dynamics open up new operating regimes for patterning circuits: while in a static picture reading out more than one bit of PI would imply the ability to precisely respond to graded morphogen concentrations, in a dynamic picture the same amount of information could be extracted by temporally varying morphogens driving a simple binary switch through a sequence of ON/OFF transitions. This picture is attractive as any single temporal snapshot would only carry, at most, a single bit of information per patterning gene, whereas temporal dynamics could encode significantly more. An advantage of such a strategy is that achieving gene expression precision corresponding to a single bit is metabolically cheaper than scaling the information to two or more bits in the static case (Tkačik et al., 2008a). Another possible advantage would be for patterning in growing tissues, as binary expression states can be made persistent and robust against external perturbations using simple bistable genetic circuitry. On the other hand, it is unclear how biological circuits would implement the computations necessary to decode such temporal profiles.

Alternatively, PI could depend not only on the local concentration, but also on some other relevant variable that is set in the history of a cell (or its lineage). Mathematically, this could be implemented by increasing the dimensionality of $\boldsymbol{g}$ to incorporate recent history. In practice, however, estimating PI from high-dimensional trajectories or internal cell states is challenging (Cepeda-Humerez et al., 2019), and the number of possibilities of what constitutes an unknown 'internal state of the cell' is vast. What constitutes a relevant internal state is also unclear. As such, we are far from fully understanding the range of patterning possibilities that can emerge when cells not only read out local morphogen values but also have memory and can act and interpret morphogens based on their internal state.

Is PI 'produced' during development?

As discussed above, spatiotemporal averaging can increase the amount of PI available from a single snapshot of downstream gene expression patterns relative to a single snapshot of input morphogen profiles, without violating the DPI. But how is the inequality consistent with the establishment of the primary morphogen gradient? Is PI created from nothing during this process? More generally, how should we think about Turing patterning and mechanisms of lateral inhibition (Afek et al., 2011), which establish spatial patterns *de novo?* For all of these cases, PI seemingly emerges. But how?

Turing patterning can be reconciled with the PI framework (Green and Sharpe, 2015). In essence, information about initial and boundary conditions is transformed into PI in the bulk of the organism (Hillenbrand et al., 2016). A key insight here is that establishing a sharp pattern with clear boundaries is insufficient. Such patterns need to be generated reproducibly from specimen to specimen. In the Turning mechanism, which is deterministic, the locations of boundaries depend on the exact geometry and on the initial 'noise' in the system that breaks the symmetry. For the same pattern to emerge reproducibly, the initial noise and the geometry need to be controlled precisely. Thus, PI in the final pattern of the dynamic process arises from the bits that carefully specify the geometry and the initial conditions. Nevertheless, much is yet to be understood, both conceptually as well as mathematically, even in simple toy models of gradient establishment, or models where cells are seen as proceeding algorithmically through sequences of switch-like decisions to set up a spatial pattern. These questions are especially pertinent when self-organized patterning systems based on reaction-diffusion mechanisms interact with global PI (Green and Sharpe, 2015).

What sources of variability constitute 'biologically relevant' variability?

In its information-theoretic formulation, PI fundamentally depends on fluctuations in morphogen patterns and on the variability of

Development (2021) 148, dev176065. doi:10.1242/dev.176065

morphogen profiles. Although experimental noise must clearly be accounted for before PI can be computed, the other sources of variability that should be considered are less clear. We stress that this is not a mathematical or a technical issue, but a matter that depends on the system and the biological objective or question. Different choices of variability imply different interpretations for the resulting PI. The fundamental question here is what constitutes biologically relevant variability?

Is it simply single-embryo variability due to intrinsic stochasticity of molecular biochemical reactions? This would be appropriate if one assumes that molecular decoding mechanisms within individual embryos can compensate for systematic embryo-to-embryo variability, e.g. due to variation in the finite amount of deposited maternal morphogen molecules. If this is unlikely, then we must include such 'extrinsic noise' into the relevant variability in addition to intrinsic noise, which in turn must decrease PI. It is even less clear whether environmental noise should be included in biologically relevant variability. For example, exposure of embryos to temperature or chemical variations certainly occurs under natural conditions (Kuntz and Eisen, 2014); but should such variability be removed under laboratory conditions? The choice would again depend on assumptions about potential compensation mechanisms for such variability. For example, when we subtract variability due to developmental timing, we assume that the system determines its PI according to an internal timing mechanism. Thus, when the internal timing is slowed due to, e.g. lower temperature conditions, PI readout must be slowed accordingly.

How is PI related to robustness?

The advantage of a quantitative framework is that it circumvents *a priori* choices about the relevance of biological variability. In fact, it also allows concepts such as robustness of developmental networks and canalization to be interrogated. PI can be measured in differently conditioned ensembles of specimens, and its dependence on various sources of variability can be determined. Such an exercise provides a productive way to understand and mathematically formalize the notion of robustness (Barkai and Leibler, 1997; Goldman et al., 2001) under the hypothesis that a patterning system is robust when PI is maintained under parameter variation, both environmental (temperature, genetics and food) or internal (embryo size) (Cheung et al., 2014; Gregor et al., 2005; Houchmandzadeh et al., 2002; Miles et al., 2011). Selection for robustness thus implies that we should observe small differences in PI between wild-type embryo populations that are perfectly environmentally controlled, and between the ones that also include environmental variability. Making this link precise, putting robustness on a firm mathematical footing that is inherited from PI (Hillenbrand et al., 2016) and testing the above hypothesis empirically are exciting future prospects.

Why is PI transformed and how are the different representations related to developmental networks?

PI present in primary morphogen gradients is transformed, or recoded, in a series of steps before cells commit to discrete fates. Understanding the rationale for the emergence of these transformations is still an unresolved issue. In part, recoding can effectively implement spatiotemporal averaging, as explained above, thereby increasing the amount of PI available at a single point in space and time. This is the case for the transformation of primary morphogens into gap gene expression profiles in *Drosophila*. Alternatively, network interactions among gap genes could increase robustness (Hillenbrand et al., 2016), i.e. stabilize the representation of PI against external sources of variability, or ensure that the representation of position is equally precise along the whole body axis, a hallmark of optimality. In growing tissues, information could also be read out from a primary morphogen gradient at an early developmental timepoint and recoded stably into a new representation with a time delay (Zagorski et al., 2017).

From an information-theoretic perspective, however, the necessity for long developmental cascades is still largely unresolved. The positional code of the gap genes, for example, contains sufficient PI already at a local level (Petkova et al., 2019); why then recode it into expression patterns of pair-rule and segment-polarity genes (Lawrence, 1992)? One hypothesis is that these subsequent transformations, while retaining PI, make it more explicit, allowing cells to ultimately turn on or off individual fate-specifying genes in a switch-like fashion to resolve and then permanently memorize a particular cell fate. PI would thus be transformed from graded, combinatorial representations carried by a small number of genes, into more binary, and possibly less-combinatorial, representations distributed over more genes (McGinnis and Krumlauf, 1992). Such an architecture has analogies to signal processing in natural and artificial neural networks, where inputs are transformed layer by layer into robust, invariant and easily learnable representations, before being acted on by a discrete 'decision-making' circuit that minimizes the classification error (Kriegeskorte, 2015; Yamins and DiCarlo, 2016).

Can PI be related to cell fate and canalization?

The information-theoretical framework for PI describes how information about position is represented biochemically, while decoding prescribes how to extract that information optimally. Cells, however, do not need to estimate a positional coordinate in the embryo, but instead need to decide on a discrete cell fate. Although similar, these two problems are mathematically not identical. First, a coordinate is a continuous variable (making its decoding a regression problem), whereas cell fate is typically thought of as discrete (making its decoding a classification problem). Although the positional coordinate in an organism can typically be discretized by cell diameters, the issue remains of whether the task of the patterning system is indeed to permit cells to learn about their absolute positions. Second, even in a discrete cellular lattice, there is no one-to-one mapping between different cell types and different cell positions; a region of one type can, for example, stretch over more than one position. Third, when making fate decisions, different 'errors' that cells can make might not be equally deleterious; some errors, such as mis-specifying one cell in a homogenous island of other cells, could perhaps be locally corrected.

Yet the biggest challenge may be in the definition of 'cell fate' itself. What precisely constitutes cell fate or identity? In the French Flag problem, fate is the unambiguous red/white/blue 'color' of the cell denoting its discrete type, and this choice is concomitant with applying a threshold on the primary morphogen gradient. But what is the equivalent representation of fate in real cells? In *Drosophila*, local combinations of four genes at 2-3 h of development suffice to identify a specific position for a cell along the AP axis of the embryo. However, specifying the position of a cell or its fate are very different processes. In fact, it is unclear what exactly specifies fate molecularly. Even though there is enough PI to establish a fate, the actual molecular committal might only happen in subsequent layers of the regulatory network.

To tackle this problem, PI theory needs to be extended to describe how discrete fate decisions are taken optimally. It should be based on the PI encoded in the morphogen profiles, and on minimization of deleterious patterning errors. Bayesian decision making or rate-

distortion theory could potentially address this issue formally (Bowsher and Swain, 2014). Recent advances in single-cell sequencing, in particular in conjunction with machine learning and large dataset analyses (Van Der Maaten and Hinton, 2008), allow for connections between developmental patterns and fates, and the systems biology of gene expression (Baron and van Oudenaarden, 2019). Theoretical frameworks for PI and (putatively) cell-fate determination should thus incorporate single-cell gene expression data, but how to achieve that in a way that is theoretically coherent and computationally tractable remains an unresolved issue.

The powerful concept of canalization put forward by Waddington in the 1940s (Waddington, 1942) provides an intuitive explanation of how cells are reliably guided to their final fates through a series of decision events on a 'genetic landscape' that resembles a potential energy surface. A major outstanding issue is therefore whether we can elevate canalization (analogously to PI) from a biological concept to a mathematical object within the information, rate-distortion and/or decision-making theory (Cover and Thomas, 2006).

Conclusions

This Review is a biased historical appraisal of the PI paradigm, written from our perspective on how the concepts of information theory can be incorporated into developmental biology. Time will tell whether this fusion of ideas will be productive and/or whether it will lead to novel insights with predictions that would not be possible without this rigorous formalization. In our view, the act of applying an exact mathematical framework to a biological concept and actual data has already helped sharpen ideas and concepts, and has led to the next generation of precision experiments focusing on testing a theory. One might wonder whether it has been worth the effort. In this context, it is interesting to look back at Shannon's opinion piece 'The Bandwagon', which appeared 8 years after he published his seminal work on information theory. Shannon warned of hype and blind over-application of information-theoretical concepts and words across the spectrum of natural and social sciences, calling for restraint and meticulous work (Shannon, 1956). Nevertheless, his vision is optimistic:

> …many of the concepts of information theory will prove useful in these other fields but the establishing of such applications is not a trivial matter of translating words to a new domain, but rather the slow tedious process of hypothesis and experimental verification.

Fifty years on from Wolpert's seminal idea of PI, and 70 years since Shannon's work on information theory, we are truly beginning to make a connection between these two ideas and encourage more work to strengthen this connection in the future.

Acknowledgements

We thank J. Briscoe, T. R. Sokolowski and B. Zoller for helpful comments and discussion.

Competing interests

The authors declare no competing or financial interests.

Funding

This work was supported in part by the National Science Foundation, through the Center for the Physics of Biological Function (PHY-1734030), by the National Institutes of Health (R01GM097275) and by the Fonds zur Förderung der wissenschaftlichen Forschung (FWF P28844). Deposited in PMC for release after 12 months.

References

Afek, Y., Alon, N., Barad, O., Hornstein, E., Barkai, N. and Bar-Joseph, Z. (2011). A biological solution to a fundamental distributed computing problem. *Science (80-.)* **331**, 183-185. doi:10.1126/science.1193210

Akam, M. (1989). Hox and HOM: Homologous gene clusters in insects and vertebrates. *Cell* **57**, 347-349. doi:10.1016/0092-8674(89)90909-4

Albert, R. and Othmer, H. G. (2003). The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in Drosophila melanogaster. *J. Theor. Biol.* **223**, 1-18. doi:10.1016/S0022-5193(03)00035-3

Arias, A. M. and Hayward, P. (2006). Filtering transcriptional noise during development: concepts and mechanisms. *Nat. Rev. Genet.* **7**, 34-44. doi:10.1038/nrg1750

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter, P. (2002). Molecular biology of the cell. In *Molecular Biology of the Cell*, p. chapter 22. New York: Garland Science.

Barkai, N. and Leibler, S. (1997). Robustness in simple biochemical networks. *Nature* **387**, 913-917. doi:10.1038/43199

Baron, C. S. and van Oudenaarden, A. (2019). Unravelling cellular relationships during development and regeneration using genetic lineage tracing. *Nat. Rev. Mol. Cell Biol.* **20**, 753-765. doi:10.1038/s41580-019-0186-3

Bentovim, L., Harden, T. T. and DePace, A. H. (2017). Transcriptional precision and accuracy in development: from measurements to models and mechanisms. *Development* **144**, 3855-3866. doi:10.1242/dev.146563

Bergmann, S., Sandler, O., Sberro, H., Shnider, S., Schejter, E., Shilo, B.-Z. and Barkai, N. (2007). Pre-steady-state decoding of the bicoid morphogen gradient. *PLoS Biol.* **5**, 0232-0242. doi:10.1371/journal.pbio.0050046

Blake, W. J., Kærn, M., Cantor, C. R. and Collins, J. J. (2003). Noise in eukaryotic gene expression. *Nature* **422**, 633-637. doi:10.1038/nature01546

Bollenbach, T., Pantazis, P., Kicheva, A., Bökel, C., González-Gaitán, M. and Jülicher, F. (2008). Precision of the Dpp gradient. *Development* **135**, 1137-1146. doi:10.1242/dev.012062

Borst, A. and Theunissen, F. E. (1999). Information theory and neural coding. *Nat. Neurosci.* **2**, 947-957. doi:10.1038/14731

Boveri, T. (1901a). Die Polarität von Ovocyte, Ei, und Larve des Strongylocentrotus lividus. *Zool. Jahrb. Abt. Anat. Ontog. Tiere* **14**, 630-653.

Boveri, T. (1901b). Über die Polarität des Seeigels. *Verh. dt. phys. med. Ges.* **34**, 145-175.

Bowsher, C. G. and Swain, P. S. (2014). Environmental sensing, information transfer, and cellular decision-making. *Curr. Opin. Biotechnol.* **28**, 149-155. doi:10.1016/j.copbio.2014.04.010

Briscoe, J. and Small, S. (2015). Morphogen rules: design principles of gradient-mediated embryo patterning. *Development* **142**, 3996-4009. doi:10.1242/dev.129452

Brunel, N. and Nadal, J.-P. (1998). Mutual information, fisher information, and population coding. *Neural Comput.* **10**, 1731-1757. doi:10.1162/089976698300017115

Butts, D. A. and Goldman, M. S. (2006). Tuning curves, neuronal variability, and sensory coding. *PLoS Biol.* **4**, e92. doi:10.1371/journal.pbio.0040092

Capovilla, M., Eldon, E. D. and Pirrotta, V. (1992). The giant gene of Drosophila encodes a b-ZIP DNA-binding protein that regulates the expression of other segmentation gap genes. *Development* **114**, 99-112.

Cepeda-Humerez, S. A., Ruess, J. and Tkačik, G. (2019). Estimating information in time-varying signals. *PLoS Comput. Biol.* **15**, e1007290. doi:10.1371/journal.pcbi.1007290

Chen, Y. and Schier, A. F. (2001). The zebrafish Nodal signal Squint functions as a morphogen. *Nature* **411**, 607-610. doi:10.1038/35079121

Cheung, D., Miles, C., Kreitman, M. and Ma, J. (2014). Adaptation of the length scale and amplitude of the Bicoid gradient profile to achieve robust patterning in abnormally large Drosophila melanogaster embryos. *Development* **141**, 124-135. doi:10.1242/dev.098640

Conklin, E. (1905). Organ-forming substances in the eggs of ascidians. *Biol. Bull.* **8**, 205-230. doi:10.2307/1535879

Cover, T. M. and Thomas, J. A. (2006). *Elements of Information Theory*. Wiley.

Crauk, O. and Dostatni, N. (2005). Bicoid determines sharp and precise target gene expression in the Drosophila embryo. *Curr. Biol.* **15**, 1888-1898. doi:10.1016/j.cub.2005.09.046

Crick, F. (1970). Diffusion in embryogenesis. *Nature* **225**, 420-422. doi:10.1038/225420a0

Davidson, E. H. (2002). A genomic regulatory network for development. *Science* **295**, 1669-1678. doi:10.1126/science.1069883

de Polavieja, G. G. (2004). Reliable biological communication with realistic constraints. *Phys. Rev. E Stat. Physics Plasmas Fluids Relat. Interdiscip. Top.* 70, 061910. doi:10.1103/PhysRevE.70.061910

de Ronde, W., Tostevin, F. and ten Wolde, P. R. (2011). Multiplexing biochemical signals. *Phys. Rev. Lett.* **107**, 048101. doi:10.1103/PhysRevLett.107.048101

Desponds, J., Tran, H., Ferraro, T., Lucas, T., Perez Romero, C., Guillou, A., Fradin, C., Coppey, M., Dostatni, N. and Walczak, A. M. (2016). Precision of readout at the hunchback gene: analyzing short transcription time traces in living fly embryos. *PLoS Comput. Biol.* **12**, e1005256. doi:10.1371/journal.pcbi.1005256

Driever, W. and Nüsslein-Volhard, C. (1988a). A gradient of bicoid protein in Drosophila embryos. *Cell* **54**, 83-93. doi:10.1016/0092-8674(88)90182-1

Driever, W. and Nüsslein-Volhard, C. (1988b). The bicoid protein determines position in the Drosophila embryo in a concentration-dependent manner. *Cell* **54**, 138-143. doi:10.1038/337138a0
Dubuis, J. O. (2012). Quantifying positional information during early embryonic development. PhD Thesis, Princeton University, NJ, USA.
Dubuis, J. O., Samanta, R. and Gregor, T. (2013a). Accurate measurements of dynamics and reproducibility in small genetic networks. *Mol. Syst. Biol.* **9**, 639. doi:10.1038/msb.2012.72
Dubuis, J. O., Tkacik, G., Wieschaus, E. F., Gregor, T. and Bialek, W. (2013b). Positional information, in bits. *Proc. Natl. Acad. Sci. USA* **110**, 16301-16308. doi:10.1073/pnas.1315642110
Elowitz, M. B., Levine, A. J., Siggia, E. D. and Swain, P. S. (2002). Stochastic gene expression in a single cell. *Science (80-.)* **297**, 1183-1186. doi:10.1126/science.1070919
England, J. L. and Cardy, J. (2005). Morphogen gradient from a noisy source. *Phys. Rev. Lett.* **94**, 078101. doi:10.1103/PhysRevLett.94.078101
Ephrussi, A. and St. Johnston, D. (2004). Seeing is believing: the bicoid morphogen gradient matures. *Cell* **116**, 143-152. doi:10.1016/S0092-8674(04)00037-6
Erdmann, T., Howard, M. and Ten Wolde, P. R. (2009). Role of spatial averaging in the precision of gene expression patterns. *Phys. Rev. Lett.* **103**, 258101. doi:10.1103/PhysRevLett.103.258101
Fasano, L. and Kerridge, S. (1988). Monitoring positional information during oogenesis in adult Drosophila. *Development* **104**, 245-253.
Fujioka, M., Jaynes, J. B. and Goto, T. (1995). Early even-skipped stripes act as morphogenetic gradients at the single cell level to establish engrailed expression. *Development* **121**, 4371-4382.
Gerstein, M. B., Kundaje, A., Hariharan, M., Landt, S. G., Yan, K.-K., Cheng, C., Mu, X. J., Khurana, E., Rozowsky, J., Alexander, R. et al. (2012). Architecture of the human regulatory network derived from ENCODE data. *Nature* **489**, 91-100. doi:10.1038/nature11245
Gilbert, S. F. (2000). Developmental biology. Sunderland, MA: Sinauer Associates.
Golding, I., Paulsson, J., Zawilski, S. M. and Cox, E. C. (2005). Real-time kinetics of gene activity in individual bacteria. *Cell* **123**, 1025-1036. doi:10.1016/j.cell.2005.09.031
Goldman, M. S., Golowasch, J., Marder, E. and Abbott, L. F. (2001). Global structure, robustness, and modulation of neuronal models. *J. Neurosci.* **21**, 5229-5238. doi:10.1523/JNEUROSCI.21-14-05229.2001
Granados, A. A., Pietsch, J. M. J., Cepeda-Humerez, S. A., Farquhar, I. L., Tkačik, G. and Swain, P. S. (2018). Distributed and dynamic intracellular organization of extracellular information. *Proc. Natl. Acad. Sci. USA* **115**, 6088-6093. doi:10.1073/pnas.1716659115
Green, J. B. A. and Sharpe, J. (2015). Positional information and reaction-diffusion: two big ideas in developmental biology combine. *Development* **142**, 1203. doi:10.1242/dev.114991
Green, J. B. A. and Smith, J. C. (1990). Graded changes in dose of a Xenopus activin A homologue elicit stepwise transitions in embryonic cell fate. *Nature* **347**, 391-394. doi:10.1038/347391a0
Green, J. B. A. and Smith, J. C. (1991). Growth factors as morphogens: do gradients and thresholds establish body plan? *Trends Genet.* **7**, 245-250. doi:10.1016/0168-9525(91)90159-N
Green, J. B. A., Howes, G., Symes, K., Cooke, J. and Smith, J. C. (1990). The biological effects of XTC-MIF: quantitative comparison with Xenopus bFGF. *Development* **108**, 173-183.
Gregor, T., Bialek, W., van Steveninck, R. R. d. R., Tank, D. W. and Wieschaus, E. F. (2005). Diffusion and scaling during early embryonic pattern formation. *Proc. Natl. Acad. Sci. USA* **102**, 18403-18407. doi:10.1073/pnas.0509483102
Gregor, T., Tank, D. W., Wieschaus, E. F. and Bialek, W. (2007). Probing the limits to positional information. *Cell* **130**, 153-164. doi:10.1016/j.cell.2007.05.025
Gregor, T., Garcia, H. G. and Little, S. C. (2014). The embryo as a laboratory: quantifying transcription in Drosophila. *Trends Genet.* **30**, 364-375. doi:10.1016/j.tig.2014.06.002
He, F., Saunders, T. E., Wen, Y., Cheung, D., Jiao, R., ten Wolde, P. R., Howard, M. and Ma, J. (2010). Shaping a morphogen gradient for positional precision. *Biophys. J.* **99**, 697-707. doi:10.1016/j.bpj.2010.04.073
Heemskerk, I., Burt, K., Miller, M., Chhabra, S., Guerra, M. C., Liu, L. and Warmflash, A. (2019). Rapid changes in morphogen concentration control self-organized patterning in human embryonic stem cells. *Elife* **8**, e40526. doi:10.7554/eLife.40526
Hillenbrand, P., Gerland, U. and Tkačik, G. (2016). Beyond the French flag model: exploiting spatial and gene regulatory interactions for positional information. *PLoS ONE* **11**, e0163628. doi:10.1371/journal.pone.0163628
Hironaka, K. and Morishita, Y. (2012). Encoding and decoding of positional information in morphogen-dependent patterning. *Curr. Opin. Genet. Dev.* **22**, 553-561. doi:10.1016/j.gde.2012.10.002
Holloway, D. M., Harrison, L. G., Kosman, D., Vanario-Alonso, C. E. and Spirov, A. V. (2006). Analysis of pattern precision shows that Drosophila segmentation develops substantial independence from gradients of maternal gene products. *Dev. Dyn.* **235**, 2949-2960. doi:10.1002/dvdy.20940
Houchmandzadeh, B., Wieschaus, E. and Leibler, S. (2002). Establishment of developmental precision and proportions in the early Drosophila embryo. *Nature* **415**, 798-802. doi:10.1038/415798a
Houchmandzadeh, B., Wieschaus, E. and Leibler, S. (2005). Precise domain specification in the developing Drosophila embryo. *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.* **72**. doi:10.1103/PhysRevE.72.061920
Howard, M. (2012). How to build a robust intracellular concentration gradient. *Trends Cell Biol.* **22**, 311-317. doi:10.1016/j.tcb.2012.03.002
Hu, B., Chen, W., Rappel, W. J. and Levine, H. (2010). Physical limits on cellular sensing of spatial gradients. *Phys. Rev. Lett.* **105**, 048104. doi:10.1103/PhysRevLett.105.048104
Illmensee, K. and Mahowald, A. P. (1974). Transplantation of posterior polar plasm in Drosophila. Induction of germ cells at the anterior pole of the egg. *Proc. Natl. Acad. Sci. USA* **71**, 1016-1020. doi:10.1073/pnas.71.4.1016
Jaeger, J. and Reinitz, J. (2006). On the dynamic nature of positional information. *BioEssays* 28, 1102-1111. doi:10.1002/bies.20494
Jaeger, J., Surkova, S., Blagov, M., Janssens, H., Kosman, D., Kozlov, K. H., Manu, K. N., Myasnikova, E., Vanario-Alonso, C. E., Samsonova, M. et al. (2004). Dynamic control of positional information in the early Drosophila embryo. *Nature* **430**, 368-371. doi:10.1038/nature02678
Jaeger, J., Irons, D. and Monk, N. (2008). Regulative feedback in pattern formation: towards a general relativistic theory of positional information. *Development* **135**, 3175-3183. doi:10.1002/bies.20494
Kirschner, M. and Gerhart, J. (1997). *Cells, Embryos and Evolution*. Blackwell.
Kraut, R. and Levine, M. (1991). Spatial regulation of the gap giant during Drosophila development. *Development* **111**, 601-609.
Kriegeskorte, N. (2015). Deep neural networks: a new framework for modeling biological vision and brain information processing. *Annu. Rev. Vis. Sci.* **1**, 417-446. doi:10.1146/annurev-vision-082114-035447
Kuntz, S. G. and Eisen, M. B. (2014). Drosophila embryogenesis scales uniformly across temperature in developmentally diverse species. *PLoS Genet.* **10**, e1004293. doi:10.1371/journal.pgen.1004293
Lacalli, T. C. and Harrison, L. G. (1991). From gradient to segments: models for pattern formation in early Drosophila. *Semin. Dev. Biol.* **2**, 107-117.
Lawrence, P. A. (1970). How do cells know where they are? *Adv. Sci.* **132**, 121-128.
Lawrence, P. A. (1988). Background to bicoid. *Cell* **54**, 1-2. doi:10.1016/0092-8674(88)90172-9
Lawrence, P. A. (1992). *The Making of a Fly: The Genetics of Animal Design*. Oxford, U.K: Blackwell Scientific Publications.
Lawrence, P. A. (2001). Morphogens: how big is the big picture? *Nat. Cell Biol.* **3**, E151-E154. doi:10.1038/35083096
Lecuit, T., Brook, W. J., Ng, M., Calleja, M., Sun, H. and Cohen, S. M. (1996). Two distinct mechanisms for long-range patterning by Decapentaplegic in the Drosophila wing. *Nature* **381**, 387-393. doi:10.1038/381387a0
Lewis, J., Slack, J. M. W. and Wolpert, L. (1977). Thresholds in development. *J. Theor. Biol.* **65**, 579-590. doi:10.1016/0022-5193(77)90216-8
Little, S. C., Tikhonov, M. and Gregor, T. (2013). Precise developmental gene expression arises from globally stochastic transcriptional activity. *Cell* **154**, 789-800. doi:10.1016/j.cell.2013.07.025
McGinnis, W. and Krumlauf, R. (1992). Homeobox genes and axial patterning. *Cell* **68**, 283-302. doi:10.1016/0092-8674(92)90471-N
McHale, P., Rappel, W.-J. and Levine, H. (2006). Embryonic pattern scaling achieved by oppositely directed morphogen gradients. *Phys. Biol.* **3**, 107-120. doi:10.1088/1478-3975/3/2/003
McMahon, A. P., Ingham, P. W. and Tabin, C. J. (2003). Developmental roles and clinical significance of hedgehog signaling. *Curr. Top. Dev. Biol.* **53**, 1-114. doi:10.1016/S0070-2153(03)53002-2
Meinhardt, H. (1986). Hierarchical inductions of cell states: a model for segmentation in Drosophila. *J. Cell Sci. Suppl.* **4**, 357-381. doi:10.1242/jcs.1986.Supplement_4.20
Meinhardt, H. and Gierer, A. (1980). Generation and regeneration of sequence of structures during morphogenesis. *J. Theor. Biol.* **85**, 429-450. doi:10.1016/0022-5193(80)90318-5
Miles, C. M., Lott, S. E., Luengo Hendriks, C. L., Ludwig, M. Z., Manu, Williams, C. L. and Kreitman, M. (2011). Artificial selection on egg size perturbs early pattern formation in Drosophila melanogaster. *Evolution (N. Y).* **65**, 33-42. doi:10.1111/j.1558-5646.2010.01088.x
Morgan, T. H. (1904). An attempt to analyse the phenomena of polarity in tubularia. *J. exp. Zool.* **1**, 587-591. doi:10.1002/jez.1400010404
Morgan, T. H. (1905). Polarity considered as a phenomenon of gradation of materials. *J. exp. Zool.* **2**, 495-506. doi:10.1002/jez.1400020404
Morishita, Y. and Iwasa, Y. (2009). Accuracy of positional information provided by multiple morphogen gradients with correlated noise. *Phys. Rev. E Stat. Nonlinear Soft. Matter Phys.* **79**, 061905. doi:10.1103/PhysRevE.79.061905
Morishita, Y. and Iwasa, Y. (2011). Coding design of positional information for robust morphogenesis. *Biophys. J.* **101**, 2324-2335. doi:10.1016/j.bpj.2011.09.048
Moses, K. and Rubin, G. M. (1991). glass encodes a site-specific DNA-binding protein that is regulated in response to positional signals in the developing Drosophila eye. *Genes Dev.* **5**, 583-593. doi:10.1101/gad.5.4.583

Mugler, A., Walczak, A. M. and Wiggins, C. H. (2010). Information-optimal transcriptional response to oscillatory driving. *Phys. Rev. Lett.* **105**, 058101. doi:10.1103/PhysRevLett.105.058101
Mugler, A., Levchenko, A. and Nemenman, I. (2016). Limits to the precision of gradient sensing with spatial communication and temporal integration. *Proc. Natl. Acad. Sci. USA* **113**, E689-E695. doi:10.1073/pnas.1509597112
Neumann, C. and Cohen, S. (1997). Morphogens and pattern formation. *BioEssays* **19**, 721-729. doi:10.1002/bies.950190813
Oates, A. C., Morelli, L. G. and Ares, S. (2012). Patterning embryos with oscillations: Structure, function and dynamics of the vertebrate segmentation clock. *Development* 139, 625-639. doi:10.1242/dev.063735
Ozbudak, E. M., Thattai, M., Kurtser, I., Grossman, A. D. and Van Oudenaarden, A. (2002). Regulation of noise in the expression of a single gene. *Nat. Genet.* **31**, 69-73. doi:10.1038/ng869
Patel, N. H. and Lall, S. (2002). Precision patterning. *Nature* **415**, 748-749. doi:10.1038/415748a
Petkova, M. D., Little, S. C., Liu, F. and Gregor, T. (2014). Maternal origins of developmental reproducibility. *Curr. Biol.* **24**, 1283-1288. doi:10.1016/j.cub.2014.04.028
Petkova, M. D., Tkačik, G., Bialek, W., Wieschaus, E. F. and Gregor, T. (2019). Optimal decoding of cellular identities in a genetic network. *Cell* **176**, 844-855.e15. doi:10.1016/j.cell.2019.01.007
Postlethwait, J. H. and Schneiderman, H. A. (1971). Pattern formation and determination in the antenna of the homoeotic mutant Antennapedia of Drosophila melanogaster. *Dev. Biol.* 25, 606-640. doi:10.1016/0012-1606(71)90008-x
Raser, J. M. and O'Shea, E. K. (2004). Control of stochasticity in eukaryotic gene expression. *Science* **304**, 1811-1814. doi:10.1126/science.1098641
Reeves, G. T., Trisnadi, N., Truong, T. V., Nahmad, M., Katz, S. and Stathopoulos, A. (2012). Dorsal-ventral gene expression in the Drosophila embryo reflects the dynamics and precision of the dorsal nuclear gradient. *Dev. Cell*. 22, 544-557. doi:10.1016/j.devcel.2011.12.007
Reinitz, J., Mjolsness, E. and Sharp, D. H. (1995). Model for cooperative control of positional information inDrosophila by bicoid and maternal hunchback. *J. Exp. Zool.* **271**, 47-56. doi:10.1002/jez.1402710106
Rivera-Pomar, R., Lu, X., Perrimon, N., Taubert, H. and Jäckle, H. (1995). Activation of posterior gap gene expression in the Drosophila blastoderm. *Nature* **376**, 253-256. doi:10.1038/376253a0
Rosenfeld, N., Young, J. W., Alon, U., Swain, P. S. and Elowitz, M. B. (2005). Gene regulation at the single-cell level. *Science (80-.).* **307**, 1962-1965. doi:10.1126/science.1106914
Rushlow, C. A. and Shvartsman, S. Y. (2012). Temporal dynamics, spatial range, and transcriptional interpretation of the Dorsal morphogen gradient. *Curr. Opin. Genet. Dev.* **22**, 542-546. doi:10.1016/j.gde.2012.08.005
Saunders, J. W. and Gasseling, M. T. (1968). Ectodermal-mesenchymal interactions in the origin of limb symmetry. In *Epithelial-Mesenchymal Interactions* (ed. R. Fleischmajer and R. E. Billingham), pp. 78-97. Williams and Wilkins.
Schüpbach, T. and Wieschaus, E. (1986). Maternal-effect mutations altering the anterior-posterior pattern of the Drosophila embryo. *Roux's Arch. Dev. Biol.* **195**, 302-317. doi:10.1007/BF00376063
Shannon, C. E. (1948). A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 623-656. doi:10.1002/j.1538-7305.1948.tb00917.x
Shannon, C. (1956). The bandwagon. *IRE Trans. Inf. Theory* **2**, 3. doi:10.1109/TIT.1956.1056774
Sharpe, J. (2019). Wolpert's French flag: what's the problem? *Development* **146**, dev185967. doi:10.1242/dev.185967
Sokolowski, T. R. and Tkačik, G. (2015). Optimizing information flow in small genetic networks. IV. Spatial coupling. *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.* **91**, 062710. doi:10.1103/PhysRevE.91.062710
Sokolowski, T. R., Erdmann, T. and ten Wolde, P. R. (2012). Mutual repression enhances the steepness and precision of gene expression boundaries. *PLoS Comput. Biol.* **8**, e1002654. doi:10.1371/journal.pcbi.1002654
Spemann, H. and Schotté, O. (1932). Über xenoplastische Transplantation als Mittel zur Analyse der embryonalen Induktion. *Naturwissenschaften* **20**, 463-467. doi:10.1007/BF01504985
Spirov, A. V. and Holloway, D. M. (2003). Making the body plan: precision in the genetic hierarchy of Drosophila embryo segmentation. *In Silico Biol.* **3**, 89-100.
Strong, S. P., Koberle, R., de Ruyter van Steveninck, R. R. and Bialek, W. (1998). Entropy and information in neural spike trains. *Phys. Rev. Lett.* **80**, 197-200. doi:10.1103/PhysRevLett.80.197
Swain, P. S., Elowitz, M. B. and Siggia, E. D. (2002). Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc. Natl. Acad. Sci. USA* **99**, 12795-12800. doi:10.1073/pnas.162041399
Tkačik, G. and Bialek, W. (2016). Information processing in living systems. *Annu. Rev. Condens. Matter Phys.* **7**, 89-117. https://doi.org/10.1146/annurev-conmatphys-031214-014803
Tkačik, G. and Walczak, A. M. (2011). Information transmission in genetic regulatory networks: a review. *J. Phys. Condens. Matter* **23**. doi:10.1088/0953-8984/23/15/153102
Tkačik, G., Gregor, T. and Bialek, W. (2008a). The role of input noise in transcriptional regulation. *PLoS One* **3**. doi:10.1371/journal.pone.0002774
Tkačik, G., Callan , Jr, C. G. and Bialek, W. (2008b). Information flow and optimization in transcriptional regulation. *Proc. Natl. Acad. Sci. USA* **105**, 12265-12270. doi:10.1073/pnas.0806077105
Tkačik, G., Callan, C. G. and Bialek, W. (2008c). Information capacity of genetic regulatory elements. *Phys. Rev. E* **78**, 011910. doi:10.1103/PhysRevE.78.011910
Tkačik, G., Walczak, A. M. and Bialek, W. (2009). Optimizing information flow in small genetic networks. *Phys. Rev. E Stat. Nonlinear, Soft Matter Phys.* **80**. doi:10.1103/PhysRevE.80.031920
Tkačik, G., Dubuis, J. O., Petkova, M. D. and Gregor, T. (2015). Positional information, Positional error, and readout precision in morphogenesis: a mathematical framework. *Genetics* **199**, 39-59. doi:10.1534/genetics.114.171850
Tomlinson, A., Bowtell, D. D. L., Hafen, E. and Rubin, G. M. (1987). Localization of the sevenless protein, a putative receptor for positional information, in the eye imaginal disc of Drosophila. *Cell* **51**, 143-150. doi:10.1016/0092-8674(87)90019-5
Tostevin, F. and Ten Wolde, P. R. (2009). Mutual information between input and output trajectories of biochemical networks. *Phys. Rev. Lett.* 102, 218101. doi:10.1103/PhysRevLett.102.218101
Tostevin, F., Ten Wolde, P. R. and Howard, M. (2007). Fundamental limits to position determination by concentration gradients. *PLoS Comput. Biol.* 3, e78. doi:10.1371/journal.pcbi.0030078
Tsimring, L. S. (2014). Noise in biology. *Reports Prog. Phys.* **77**. doi:10.1088/0034-4885/77/2/026601
Turing, A. M. (1952). The chemical basis of morphogenesis. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **237**, 37-71. doi:10.1098/rstb.1952.0012
Van Der Maaten, L. and Hinton, G. E. (2008). Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579-2605.
van Kampen, N. G. (2007). *Stochastic Processes in Physics and Chemistry*. Elsevier.
Villoutreix, P., Andén, J., Lim, B., Lu, H., Kevrekidis, I. G., Singer, A. and Shvartsman, S. Y. (2017). Synthesizing developmental trajectories. *PLoS Comput. Biol.* 13, e1005742. doi:10.1371/journal.pcbi.1005742
Waddington, C. H. (1942). Canalization of Development and the Inheritance of Acquired Characters. *Nature* **150**, 563-565. doi:10.1038/150563a0
Wilson, E. B. (1904). Experimental studies on germinal localization. I The germ regions in the egg of Dentalium. II Experiements of the cleavage mosaic in Patella and Dentalium. *J. Exp. Zool.* **1**, 1-72. doi:10.1002/jez.1400010102
Wolpert, L. (1969). Positional information and the spatial pattern of cellular differentiation. *J. Theor. Biol.* **25**, 1-47. doi:10.1016/S0022-5193(69)80016-0
Wolpert, L. (1971). Positional information and pattern formation. *Curr. Top. Dev. Biol.* 117, 597-608. doi: 10.1016/bs.ctdb.2015.11.008
Wolpert, L. (1989). Positional information revisited. *Development* **107**, 3-12.
Wolpert, L. (1996). One hundred years of positional information. *Trends Genet.* 12, 359-364. doi:10.1016/s0168-9525(96)80019-9
Yamins, D. L. K. and DiCarlo, J. J. (2016). Using goal-driven deep learning models to understand sensory cortex. *Nat. Neurosci.* 19, 356-365. doi:10.1038/nn.4244
Zagorski, M., Tabata, Y., Brandenberg, N., Lutolf, M. P., Tkacik, G., Bollenbach, T., Briscoe, J. and Kicheva, A. (2017). Decoding of position in the developing neural tube from antiparallel morphogen gradients. *Science* **356**, 1379-1383. doi:10.1126/science.aam5887
Ziv, E., Nemenman, I. and Wiggins, C. H. (2007). Optimal signal processing in small stochastic biochemical networks. *PLoS ONE* **2**. https://doi.org/10.1371/journal.pone.0001077
Zoller, B., Little, S. C. and Gregor, T. (2018). Diverse spatial expression patterns emerge from unified kinetics of transcriptional bursting. *Cell* **175**, 835-847.e25. doi:10.1016/j.cell.2018.09.056