Positional information, in bits

Julien O. Dubuis,[1] William Bialek,[1,2] Eric F. Wieschaus,[2,3,4] and Thomas Gregor[1,2]

***Short Abstract* — In early embryonic development, spatial variations in the concentration of transcription factors generate more complex patterns of gene expression, thus transmitting "positional information." Using the *Drosophila* embryo as a model system, we measure the amount of positional information carried by a group of four 'gap genes' that respond to the primary maternal morphogen gradients. The information carried by individual genes is much larger than one bit, and preliminary data indicate that all four genes provide enough information to specify the location of every cell along the anterior-posterior axis, close to the physical limits on information transmission set by noise.**

***Keywords* — Information theory, embryonic development, gene regulatory networks**

DURING the development of multicellular organisms such as the fruit fly *Drosophila melanogaster*, cells acquire information about their position in the embryo in response to "morphogens" whose concentrations vary along the major axes of the embryo [1,2]. In the case of *Drosophila*, the maternal morphogen gradient of *Bicoid (Bcd)* is established along the anterior-posterior axis of the embryo within 2 hours of development, precisely and reproducibly [3]. This gradient controls the expression of four "gap genes" (*Hb, Kr, Gt* and *Kni*), which in turn a second generation of genes, the "pair-rule genes," with more complex spatial patterns. Although the concept of positional information has been discussed for nearly a century, there have been few attempts to quantify just how much positional information is available in these patterns of gene expression. Are genes restricted to being 'on' and 'off,' so that each gene conveys at most one bit? Can just a handful of genes at early stages of the network convey enough information to specify the location of every cell in the embryo? What limits the transmission of information in these systems [4]? Are there strategies that allow the system to maximize information transmission in the presence of these limits [5]? Here we try to answer these questions, using the gap genes as a model system.

The positional information carried by a gene quantifies how well a nucleus can estimate its position by reading out the concentration of the corresponding protein. With I bits of information, cells can reliably distinguish $\sim 2^I$ different states. Given that there are roughly 100 rows of nuclei along the anterior-posterior axis of the fly embryo, cells would need a total positional information of $\log_2(100) \approx 6.7$ bits in order for each nucleus to have its own identity [4]. If the patterns of the gap genes were absolutely reproducible from one embryo to another, and from cell to cell at the same anterior-posterior location, then the information that they would carry would be infinite. The variability of the patterns across embryos means that the information that these genes carry is finite, and quantifying the reproducibility vs. variability of the expression patterns should lead us to an estimate of the positional information.

Using immunofluorescence techniques we separately measured the profiles of the gap genes along the dorsal part of hundreds of embryos at different times. We then used these profiles to estimate the information carried by each gap gene. We found that each gap gene carries positional information in the 1.5-2 bits range. This is much more than one bit, as would have been expected if the only reliable signal were the boundary between high and low expression.

Estimating the information carried by all four gap genes together is more challenging. Looking at pairs, we find that the information carried by two genes is close to the sum of the information carried by the individual genes, although this depends on a balance between redundancy in the overall expression profiles and synergy created by correlations between variability in different genes at the same position. We suggest approximate strategies for estimating the information in larger groups of genes, and preliminary results suggest that the total information is in the range of 6-7 bits. Thus, it is possible that the gap genes carry enough information to specify the location of every cell along the anterior-posterior axis. Also, we will argue that this is close to the physical limits on how much information can be extracted from a morphogen profile given the maximal concentration and the minimum noise levels due to molecule counting.

REFERENCES

[1] L Wolpert, Positional information and the spatial pattern of cellular differentiation. *J Theor Biol* **25**, 1-47 (1969)

[2] PA Lawrence, *The Making of a Fly: The genetics of Animal Design* (Blackwell, Oxford, 1992)

[3] T Gregor, DW Tank, EF Wieschaus & W Bialek, Probing the limits of positional information. *Cell* **130**, 153-164 (2007).

[4] G Tkacik, CG Callan & W Bialek, Information flow and optimization in transcriptional control. *Proc Natl Acad Sci (USA)* **105**, 12265-12270 (2005)

[5] G Tkačik, AM Walczak & W Bialek, Optimizing information flow in small genetic networks. *arXiv:0903.4491* (2009).

This work was supported in part by NSF Grant PHY-0650617, and by NIH Grants P50 GM071508 and R01 GM077599

[1]Joseph Henry Laboratories of Physics, [2]Lewis Sigler Institute for Integrative Genomics, [3]Department of Molecular Biology, and [4]Howard Hughes Medical Institute, Princeton University 08544 Princeton, NJ, USA

[*]Correspondence: jdubuis@princeton.edu