Spring 2026 – Systems Biology of Reproduction
Discussion Outline (Systems Biology)
Eric Nilsson – Biol 475/575
Weeks 1 and 2 (January 22, 2026)

Systems Biology

Primary Papers

1. Westerhoff & Palsson (2004) Nat Biotech 22:1249-1252
2. Joyner (2011) J Appl Physiol 111:335-342
3. Stanoev, et al. (2021) Development 148:dev197608
4. Zhang, et al. (2023) Human Reproduction Update 00(0), 1-17

Discussion

Student 1 - Ref #1 above
-How does this support evolutionary systems biology?
-Give an example that supports this perspective.

Ref #2 above
-What is the problem with reductionism?
-What is the void?
-What is the solution?

Student 2 - Ref #3 above
-What is robustness?
-What is the role of inhomogeneous steady state?
-What insights in the systems biology of cellular differentiation are described?

Student 4 - Ref #4 above
-What are spatial transcriptomes?
-What is the experimental design?
-What reproductive tissues used and insights obtained?

nature biotechnology

The evolution of molecular biology into systems biology

Hans V Westerhoff[1] & Bernhard O Palsson[2]

Systems analysis has historically been performed in many areas of biology, including ecology, developmental biology and immunology. More recently, the genomics revolution has catapulted molecular biology into the realm of systems biology. In unicellular organisms and well-defined cell lines of higher organisms, systems approaches are making definitive strides toward scientific understanding and biotechnological applications. We argue here that two distinct lines of inquiry in molecular biology have converged to form contemporary systems biology.

Whereas the foundations of systems biology-at-large are generally recognized as being as far apart as 19th century whole-organism embryology and network mathematics, there is a school of thought that systems biology of the living cell has its origin in the expansion of molecular biology to genome-wide analyses. From this perspective, the emergence of this 'new' field constitutes a 'paradigm shift' for molecular biology, which ironically has often focused on reductionist thinking. Systems thinking in molecular biology will likely be dominated by formal integrative analysis going forward rather than solely being driven by high-throughput technologies.

It is, however, incorrect to state that integrative thinking is new to molecular biology. The first molecular regulatory circuits were mapped out over 40 years ago. The feedback inhibition of amino acid biosynthetic pathways was discovered in 1957 (refs. 1,2), and the transcriptional regulation associated with the glucose-lactose diauxic shift led to the definition of the *lac* operon and the elucidation of its regulation[3]. With the study of these regulatory mechanisms, admittedly on a small scale, molecular biologists began to apply systems approaches to unravel the molecular components and logic that underlie cellular processes, often in parallel with the characterization of individual macromolecules. High-throughput technologies have made the scale of such inquiries much larger, enabling us to view the genome as the 'system' to study. Thus, the popular contemporary view of systems biology may be synonymous with 'genomic' biology.

This article discusses two historical roots of systems biology in molecular biology (**Fig. 1**). Although we briefly outline the more familiar first root—which stemmed from fundamental discoveries about the nature of genetic material, structural characterization of macromolecules and later developments in recombinant and high-throughput technologies—more emphasis is placed on the second root, which sprung from nonequilibrium thermodynamics theory in the 1940s, the elucidation of biochemical pathways and feedback controls in unicellular organisms and the emerging recognition of networks in biology. We conclude by discussing how these two lines of work are now merging in contemporary systems biology.

Scaling-up molecular biology

In the decades following its foundational discoveries of the structure and information coding of DNA and protein, molecular biology blossomed as a field, with a series of breathtaking discoveries (**Fig. 1**). The description of restriction enzymes and cloning were major breakthroughs in the 1970s, ushering in the era of genetic engineering and biotechnology. In the 1980s, we began to see the scale-up of some of the fundamental experimental approaches of molecular biology. Automated DNA sequencers began to appear and reached genome-scale sequencing in the mid-1990s[4,5]. Automation, miniaturization and multiplexing of various assays led to the generation of additional 'omics' data types[6,7].

The large volumes of data generated by these approaches led to rapid growth in the field of bioinformatics, again largely emanating from the reductionist perspective. Although this effort was mostly focused on statistical models and object classification approaches in the late 1990s, it was recognized that a more formal and mechanistic framework was needed to analyze multiple high-throughput data types systematically[8,9]. This need led to efforts toward genome-scale model building to analyze the systems properties of cellular function.

Molecular self-organization

Even before the first key events in the history of molecular biology, several lines of reasoning revealed that integration of multiple molecular processes is fundamental to the living cell. Biochemical processes necessitate the production of entropy (chaos in the thermodynamic sense) as driving force. The paradox felt by many, but expressed by Schrödinger in his war-time lectures[10], was how one could explain the progressive ordering that occurs in developmental biology (that is, the 'self-organization,' decrease in chaos) when entropy ('chaos') must be increased.

The answer was that one process could produce order (negative entropy or negentropy) provided it was coupled to a second process that produced more chaos (entropy): coupling, another word for integration of processes, is therefore essential for life. Onsager[11] provided the basis for this concept by stressing the significance of the coupling of dissimilar processes. He is also relevant because he discovered a law for such systems of coupled processes: close to equilibrium the dependence of the one process rate on the driving force of the other process should equal the dependence of the other process rate on the

[1]Departments of Molecular Cell Physiology and Mathematical Biochemistry, BioCentrum Amsterdam, De Boelelaan 1085, NL-108, HV Amsterdam, the Netherlands. [2]Department of Bioengineering, University of California-San Diego, 9500 Gilman Drive, La Jolla, California 92093-0412, USA. Correspondence should be addressed to H.V.W. (hw@bio.vu.nl) or B.O.P. (palsson@ucsd.edu).

Published online 6 October 2004; doi:10.1038/nbt1020

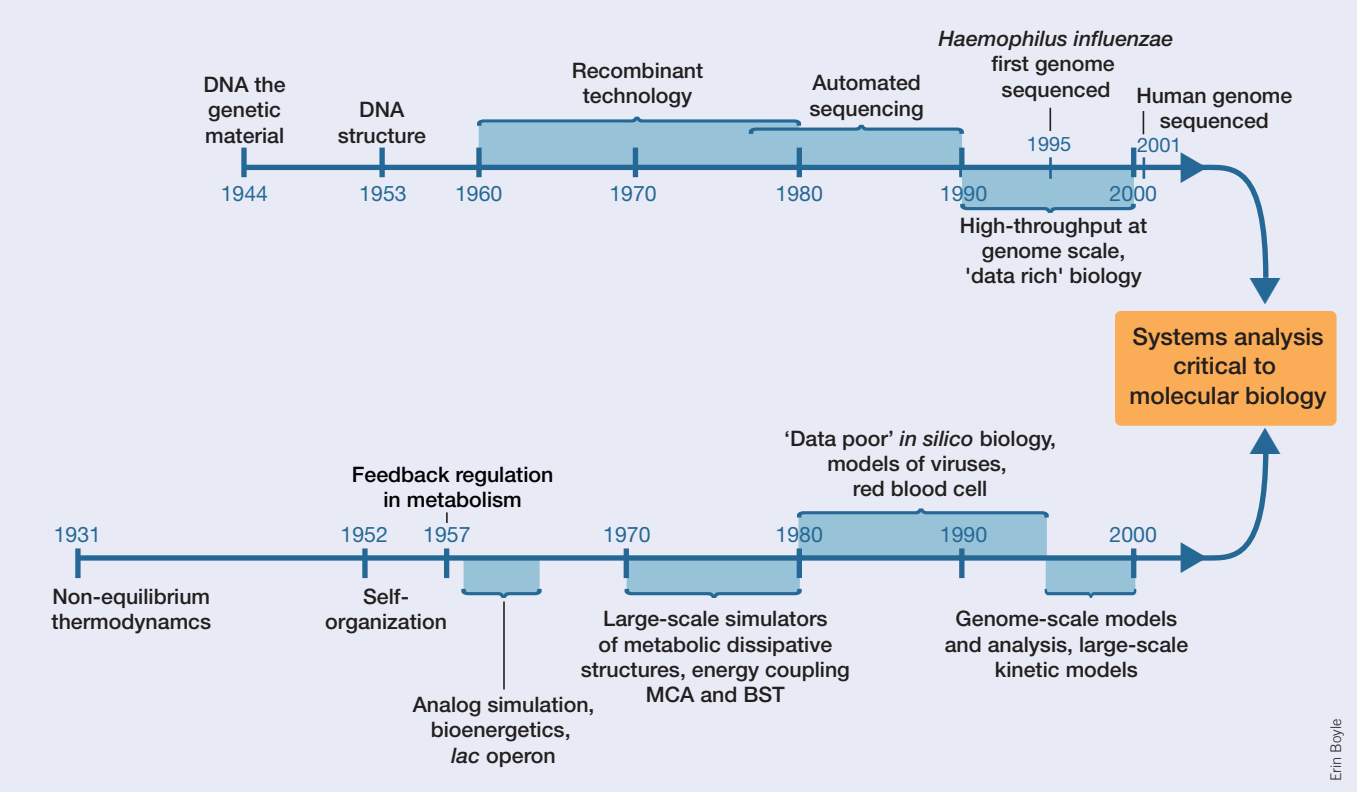


Figure 1 Two lines of inquiry led from the approximate onset of molecular biological thinking to present-day systems biology. The top timeline represents the root of systems biology in mainstream molecular biology, with its emphasis on individual macromolecules. Scaled-up versions of this effort then induced systems biology as a way to look at all those molecules simultaneously, and consider their interactions. The lower timeline represents the lesser-known effort that constantly focused on the formal analysis of new functional states that arise when multiple molecules interact simultaneously.

former driving force. Caplan, Essig and Rottenberg[12] later defined a coupling coefficient, which quantifies the extent to which two processes are coupled in a system and showed that this coefficient must range between 0 and 1.

These approaches were called nonequilibrium thermodynamics and constituted a prelude to systems biology at the cell and molecular levels in that they (i) dealt with integration quantitatively and (ii) aimed to discover general principles rather than just being descriptive. An improved procedure for describing ion movement and energy transduction in biological membranes, termed mosaic nonequilibrium thermodynamics, further progressed towards systems thinking in that it (iii) established a connection to molecular mechanisms and (iv) enabled the determination of the stoichiometry of membrane energy transduction from system data[13]. Peter Mitchell's[14] chemiosmotic coupling principle was another early case of systems analysis in cell and molecular biology. It stated that ATP synthesis was coupled in quite an indirect way to respiration, involving an entire intracellular system, including a volume surrounded by an ion-impermeable membrane and proton movement across it. Indeed, for eukaryotes, this provided much of the rationale for the organization of the mitochondrion. In his calculations verifying that that the proposed chemiosmotic mechanisms transferred sufficient free energy to empower ATP synthesis, Mitchell demonstrated the sort of quantitative thinking that would eventually prove crucial to the study of biochemical systems[14].

The problem of biological self-organization was to understand how structures, oscillations or waves arise in a steady and homogenous environment, a phenomenon called symmetry breaking. Turing[16] led the way, but the Prigogine school[17] and others developed the topic from the perspective of nonequilibrium thermodynamics in molecular contexts such as biochemical reactions involved in sugar metabolism (glycolysis). They demonstrated how having a sufficient number of nonlinearly interacting chemical processes in a single system such as the Zhabotinski reaction, a developing tissue, or glycolysis, could lead to symmetry-breaking as a result of self-amplification of random fluctuations. Of course, more recent molecular developmental biology studies have shown that reality is even more complicated; prespecification by external (maternally specified) gradients of morphogens may substitute for the random fluctuations, increasing the robustness of development[18]. Perhaps more importantly, Prigogine searched for and found a law (on minimum entropy production). Although it is strictly valid only in Onsager's near-equilibrium domain, it testified to the systems scientists' quest for the principles underlying systems, rather than just for their appearances.

Early on, oscillations in yeast glycolysis were the experimental systems of choice. Although intact cells were studied[19], more often measurements were made using cell extracts[20]. Reductionist biochemical thinking proclaimed that a single pacemaker enzyme should be responsible for the oscillations. Only relatively recently has systems-based analysis in one of our laboratories (H.V.W.) been used to reveal that the oscillations are simultaneously controlled by many steps in the intracellular network[21] and how the oscillations in the individual cells synchronize actively[22]. Of course, with the more recent experimental capability to inspect single cells dynamically, more and more cells are

seen to exhibit asynchronous oscillations of all sorts and some of these cases are up for systems biology analysis. Slime mold aggregation was another early case where a network of reactions was shown to be essential for systems biology reaching one step beyond cell biology, again by combining mathematical modeling with experimental molecular information[23].

Building large-scale models

Following the events of the late 1950s and early 1960s, researchers undertook efforts that were not well publicized and formulated mathematical models to simulate the functions of newly discovered regulatory circuits in cells. Even before digital computers became available, simulations of integrated molecular functions were performed on analog computers[24]. These efforts grew in scale to dynamic simulation of large metabolic networks in the 1970s[25–27]. Following the pathway-centered kinetic models in the seventies[28], cell-scale flux models of the human red cell were published by the late 1980s (ref. 29), and by the early 1990s genome-scale models of viruses and large-scale models of mitosis were formulated[30]. With the advent of genome-scale sequencing, the first genome-scale, constraint-based metabolic models for bacteria were constructed[31]. These models describe reconstructed networks and their possible functional states (phenotypes) and are now available at the genome-scale for a growing number of organisms. They treat the 'genome' as the 'system.'

Progress toward the development of detailed kinetic models at a large scale has proven to be slower. Some of these models approach computer replicas of pathways of metabolism, signal transduction and gene expression, and are active on the web, ready for experimentation and integration (compare http://www.siliconcell.net/). Obtaining *in vivo* numerical values for kinetic constants remains a key challenge.

Metabolic control analysis

We have agreed that contemporary systems biology has an historical root outside mainstream molecular biology, ranging from basic principles of self-organization in nonequilibrium thermodynamics, through large-scale flux and kinetic models to 'genetic circuit' thinking in molecular biology. 'Systems thinking' differs from 'component thinking' and requires the development of new conceptual frameworks.

Metabolic control analysis (MCA), developed in the early seventies[28,32], presented a key example of approaches to characterize properties of networks of interacting chemical reactions. At this time, thinking in biochemistry was dominated by the concept that there had to be a single 'rate-limiting' step at the beginning of all metabolic pathways. Criteria used to establish whether a given enzyme was rate-limiting referred to it as being far from equilibrium, strongly regulated by various metabolic factors or causing pathway flux to decrease when inhibited.

However, the application of these criteria to some metabolic pathways suggested that they contained more than a single rate-limiting step. Network thinking through MCA helped to resolve this paradox. First, mathematical models of metabolic pathways were developed both for inspiration and discovery, and subsequently used to check numerically the principles they conjectured[28,32]. Second, quantitative definitions were developed to describe the extent to which a step limited the flux through a pathway. This 'flux-control coefficient' of a particular step corresponded to the sensitivity coefficient of the pathway flux with respect to the activity of the particular enzyme. Third, these investigators looked for proof of the concept that there should be a single rate-limiting enzyme in a pathway that should have a flux-control coefficient of unity, with all others having flux control coefficients of zero. Instead, they found a theorem stating that all the flux-control coefficients must sum to unity[28,32]. This result then suggested that there need not be a single rate-limiting step to a pathway and that instead many enzymes can contribute simultaneously to the control of the network. Thus, control was not a component property but a network property. The network nature of regulation was shown experimentally to be the case for mitochondrial ATP generation, where control was indeed distributed over more than three steps, and quite notably not particularly strong, neither for the first nor for the irreversible step of the pathway[33].

An important aspect of systems biology is to relate the system properties to the molecular properties of components that comprise a network. The kinetics-based sensitivity analysis by MCA, and its close relative, biochemical systems theory proposed by H.V.W and Chen[34], showed that by focusing on the properties of an individual component, one cannot properly decipher its role in the context of a whole network. The connectivity laws proven by MCA[28,34] (see other references in ref. 35) pinpointed how that distribution of control relates to network structure and the kinetic properties of all network components simultaneously. Similarly, the topological analyses of network structure by our groups[31,36] have revealed the existence of network-based definitions of pathways that can be used mathematically to represent all possible functional states of reconstructed networks[37]. Thus, a growing number of methods now exist to analyze the properties mathematically of the large-scale networks that we are now able to reconstruct based on high-throughput data.

Convergence

Figure 1 presents our interpretation of the history of systems analysis in cell and molecular biology. Events in the upper timeline have been much more to the fore of scientific thinking than those in the lower timeline. In one sense, the dazzling stream of discoveries and exciting technologies (most recently with genome-wide data) provides the 'biology' root to contemporary systems biology. In contrast, scientific progress in the lower timeline has never gained much notoriety, although work in this area was much more prominent in European science throughout this period. This latter branch might be thought of as the 'systems' root of systems biology.

Systems modeling and simulation in molecular biology was once seen as purely theoretical and not particularly relevant to understanding 'real' biology. However, now that molecular biology has become such a data-rich field, the need for theory, model building and simulation has emerged. The systems-directed root always had the ambition of discovering fundamental principles and laws, such as those of nonequilibrium thermodynamics and MCA. This ambition should now extend to systems biology.

All too often, the field has been perceived as just pattern recognition and phenomenological modeling. Systems biology is a thorough science with its own quest for scientific principles at the interface of physics, chemistry and biology, with its remarkable mixture of functionality, hysteresis, optimization and physical chemical limitations. *In silico* analysis of complex cellular processes (whether for data description, genetic engineering or scientific discovery), with its focus on elucidating system mechanisms, has in fact become critical for progress in biology.

The historical dichotomy in approaches to molecular biology must now be reconciled with the need to corral resources and expertise in systems approaches. Although the reductionist molecular biological root has been the focus of a plethora of investigations, literature sources and curricula, the same is not true for the systems molecular biology root. There is now a need for development of theoretical and

analytical approaches, curricula and educational materials to advance understanding of the systems in cell and molecular biology. Unknown to many, the 'pre-online PDF' era contains answers to many of the current challenges and pitfalls facing the field. So although systems biology has an intellectually exciting future ahead of it, the leaders in the field should try to minimize rediscovery and focus on the newer challenges facing us, particularly those that come with the application of existing concepts to genome-scale problems and identification of the new issues that arise from the study of cellular functions on this scale.

Where has this history brought us? We now have the growing and general recognition that systems analysis is important to the future evolution of cell and molecular biology. Some reeducation of workers in the field may be in order (http://www.systembiology.net/). Over the near term, it is likely that successes with practical applications of systems biology will be confined to unicellular systems. We are now seeing successful applications of systems biology to microbes, including pathway engineering (e.g., see our recent publications[37,38]), network-based drug design (e.g., H.V.W. and colleagues[39]), and prediction of the outcome of complex biological processes, such as adaptive evolution (B.O.P and colleagues[40]). Although the mathematical modeling of whole-body human systems cannot yet be linked to genome-wide data and models, data analysis and modeling are likely to contribute to the success of realizing the goal of individualized medicine. Even if we have to rely on less precise models than the currently available genome-scale models of microorganisms, systems biology may soon lead to better diagnosis and dynamic therapies of human disease than the qualitative methodology presently in use.

ACKNOWLEDGMENTS

We thank Adam Arkin for comments and Timothy Allen for editing. B.O.P. serves on the scientific advisory board of Genomatica, Inc.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Published online at http://www.nature.com/naturebiotechnology/

1. Umbarger, H.E. & Brown, B. Threonine deamination in *Escherichia coli*. II. Evidence for two L-threonine deaminases. *J. Bacteriol.* **73**, 105–12 (1957).
2. Yates, R.A. & Pardee, A.B. Control by uracil of formation of enzymes required for orotate synthesis. *J. Biol. Chem.* **227**, 677–692 (1957).
3. Beckwith, J.R. Regulation of the *lac* operon. Recent studies on the regulation of lactose metabolism in *Escherichia coli* support the operon model. *Science* **156**, 597–604 (1967).
4. Hunkapiller, T. *et al.* Large-scale and automated DNA sequence determination. *Science* **254**, 59–67 (1991).
5. Rowen, L., Magharias, G. & Hood, L. Sequencing the human genome. *Science* **278**, 605–607 (1997).
6. Scherf, M., Klingenhoff, A. & Werner, T. Highly specific localization of promoter regions in large genomic sequences by PromoterInspector: a novel context analysis approach. *J. Mol. Biol.* **297**, 599–606 (2000).
7. Uetz, P. *et al.* A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. *Nature* **403**, 623–627 (2000).
8. Ge, H., Walhout, A.J. & Vidal, M. Integrating 'omic' information: a bridge between genomics and systems biology. *Trends Genet.* **19**, 551–560 (2003).
9. Palsson, B.O. *In silico* biology through 'omics'. *Nat. Biotechnol.* **20**, 649–650 (2002).
10. Schrödinger, E. *What is life? The physical aspects of the living cell*. Based on Lectures Delivered under the Auspices of the Dublin Institute for Advanced Studies at Trinity College, Dublin, in February 1943. (Cambridge University Press, Cambridge, UK, 1944). http://home.att.net/~p.caimi/oremia.html
11. Onsager, L. Reciprocal relations in irreversible processes. *Phys. Rev.* **37**, 405–426 (1931).
12. Rottenberg, H., Caplan, S.R. & Essig, A. Stoichiometry and coupling: theories of oxidative phosphorylation. *Nature* **216**, 610–611 (1967).
13. Westerhoff, H.V. & Van Dam, K. Thermodynamics and Control of Biological Free-Energy Transduction (Elsevier, Amsterdam, 1987).
14. Mitchell, P. Coupling of phosphorylation to electron and hydrogen transfer by a chemiosmotic type of mechanism. *Nature* **191**, 144–148 (1961).
15. Mitchell, P. Chemiosmotic Coupling in Oxidative and Photosynthetic Phosphorylation. (Glynn Research Ltd., Bodmin, UK, 1966).
16. Turing, A. The chemical basis of morphogenesis. *Phil. Trans. Roy. Soc. London, Ser. B* **237**, 37–72 (1952).
17. Glansdorff, P. & Prigogine, I. *Structure, Stabilité et Fluctuations* (Masson, Paris, 1971).
18. Lawrence, P.A. *The Making of a Fly* (Blackwell, London, 1992).
19. Chance, B., Estabrook, R.W. & Ghosh, A. Damped sinusoidal oscillations of cytoplasmic reduced pyridine nucleotide in yeast cells. *Proc. Natl. Acad. Sci. USA* **51**, 1244–1251 (1964).
20. Hess, B. & Boiteux, A. Oscillatory phenomena in biochemistry. *Annu. Rev. Biochem.* **40**, 237–258 (1971).
21. Teusink, B., Bakker, B.M. & Westerhoff, H.V. Control of frequency and amplitudes is shared by all enzymes in three models for yeast glycolytic oscillations. *Biochim. Biophys. Acta.* **1275**, 204–212 (1996).
22. Wolf, J. *et al.* Transduction of intracellular and intercellular dynamics in yeast glycolytic oscillations. *Biophys. J.* **78**, 1145–1153 (2000).
23. Tyson, J.J. & Murray, J.D. Cyclic AMP waves during aggregation of *Dictyostelium* amoebae. *Development* **106**, 421–426 (1989).
24. Goodwin, B.C. Oscillatory Organization in Cells, a Dynamic Theory of Cellular Control Processes (Academic Press, New York, 1963).
25. Garfinkel, D. *et al.* Computer applications to biochemical kinetics. *Annu. Rev. Biochem.* **39**, 473–498 (1970).
26. Loomis, W. & Thomas, S. Kinetic analysis of biochemical differentiation in *Dictyostelium discoideum*. *J. Biol. Chem.* **251**, 6252–6258 (1976).
27. Wright, B.E. The use of kinetic models to analyze differentiation. *Behavioral Sci.* **15**, 37–45 (1970).
28. Heinrich, R., Rapoport, S.M. & Rapoport, T.A. *Progr. Biophys. Mol. Biol.* **32**, 1–83 (1977).
29. Joshi, A. & Palsson, B.O. Metabolic dynamics in the human red cell. Part I—A comprehensive kinetic model. *J. Theor. Biol.* **141**, 515–528 (1989).
30. Novak, B. & Tyson, J.J. Quantitative analysis of a molecular model of mitotic control in fission yeast. *J. Theor. Biol.* **173**, 283–305 (1995).
31. Edwards, J.S. & Palsson, B.O. Systems properties of the *Haemophilus influenzae* Rd metabolic genotype. *J. Biol. Chem.* **274**, 17410–17416 (1999).
32. Kacser, H. & Burns, J.A. *In Rate Control of Biological Processes* (ed., Davies, D.D.) 65–104 (Cambridge University Press, Cambridge, 1973).
33. Groen. A.K., Wanders, R.J.A., Van Roermund, C., Westerhoff, H.V. & Tager, J.M. Quantification of the contribution of various steps to the control of mitochondrial respiration. *J. Biol. Chem.* **257**, 2754–2757 (1982).
34. Savageau, M.A. *Biochemical Systems Analysis* (Addison-Wesley, Reading, MA, 1976).
35. Westerhoff, H.V. & Chen, Y. How do enzyme activities control metabolite concentrations? An additional theorem in the theory of metabolic control. *Eur. J. Biochem.* **142**, 425–430 (1984).
36. Westerhoff, H.V., Hofmeyr, J.H. & Kholodenko, B.N. Getting to the inside of cells using metabolic control analysis. *Biophys. Chem.* **50**, 273–283 (1994).
37. Papin, J.A., Price, N.D., Wiback, S.J., Fell, D.A. & Palsson, B.O. Metabolic pathways in the post-genome era. *Trends Biochem. Sci.* **28,** 250–258 (2003).
38. Kholodenko, B.N. & Westerhoff, H.V. (eds.) *Metabolic Engineering in the Post Genomics Era* (Horizon Bioscience, UK, 2004).
39. Bakker, B.M. *et al.* Network-based selectivity of antiparasitic inhibitors. *Mol. Biol. Rep.* **29**, 1–5 (2002).
40. Ibarra, R.U., Edwards, J.S. & Palsson, B.O. *Escherichia coli* K-12 undergoes adaptive evolution to achieve *in silico* predicted optimal growth. *Nature* **420,** 186–189 (2002).

J Appl Physiol 111: 335–342, 2011.
First published June 2, 2011; doi:10.1152/japplphysiol.00565.2011.

Edward F. Adolph Distinguished Lectureship

Giant sucking sound: can physiology fill the intellectual void left by the reductionists?

Michael J. Joyner

Department of Anesthesiology, Mayo Clinic, Rochester, Minnesota

Submitted 3 May 2011; accepted in final form 31 May 2011

Joyner MJ. Giant sucking sound: can physiology fill the intellectual void left by the reductionists?. *J Appl Physiol* 111: 335–342, 2011. First published June 2, 2011; doi:10.1152/japplphysiol.00565.2011.—Molecular reductionism has so far failed to deliver the broad-based therapeutic insights that were initially hoped for. This form of reductionism is now being replaced by so-called "systems biology." This is a nebulously defined approach and/or discipline, with some versions of it relying excessively on hypothesis-neutral approaches and only minimally informed by key physiological concepts such as homeostasis and regulation. In this context, physiology is uniquely positioned to continue to provide impressive levels of both biological and therapeutic insight by using hypothesis-driven "classical" approaches and concepts to help frame what might be described as the "pieces of the puzzle" that emerge from molecular reductionism. The strength of physiology as a "bridge" between reductionism and epidemiology, along with its unparalleled ability to generate therapeutic insights and opportunities justifies increased attention and emphasis on our discipline into the future. Arguments relevant to this set of assertions are advanced and this paper, which was based on the 2011 Adolph Lecture, represents an effort to fill the intellectual void left by reductionism and improve scientific progress.

homeostasis; regulation; integrative

THIS PAPER REFLECTS IDEAS that were presented as part of the 2011 Adolph Lecture at the Experimental Biology meeting that was held in Washington, DC. The goal of the talk was to share a physiologist's perspective on what reductionism in general and the "omic" revolution in particular has or has not done for biomedical research and associated therapeutic insights or advances. The main ideas highlighted in the lecture were the following.

1) Reductionism via various flavors of molecular biology and "omics" has so far failed to deliver its self-promoted revolution in clinical medicine.

2) Systems biology has a cell-centric focus that is marked by a limited understanding of and application to biology beyond the cell.

3) The failure of systems biology to recognize and use key concepts from physiology about homeostasis, regulation, redundancy, feedback control, and acclimation/adaptation are major limitations to this poorly defined approach.

4) While all the attention has been focused on reductionism and more recently systems biology, physiology continues to provide important biomedical insights that lead to therapeutic advances.

Address for reprint requests and other correspondence: M. J. Joyner, 200 First St. SW, Dept. of Anesthesiology, Mayo Clinic, Rochester, MN 55905 (e-mail: joyner.michael@mayo.edu).

As the title demonstrates, my goal in the Adolph Lecture and in this paper was and is to be intentionally provocative and hopefully generate a dialogue with the reductionists. In this context, and because I am "taking sides", I have adopted what might be called a conversational approach to this paper.

BIOLOGICAL ORTHOPEDIC SURGERY

A key idea or theme that seems to underpin the impetus for reductionism and various flavors of "omics" as applied to biomedical problems might be described as biological orthopedic surgery: "the gene is broken → fix the broken gene → cure the patient." This thinking clearly seems to explain the enthusiasm about gene therapy that emerged after the discovery of the genetic defect responsible for the most common form of cystic fibrosis and more recently ideas about a limited number of common gene variants explaining the risk for common conditions like atherosclerosis and diabetes (10–12, 43, 51). The line of thinking described above flows from what Denis Noble has critically termed "Neo-Darwinian" thinking about the relationship between genes and phenotype (45, 46). It is exemplified by two quotes, the first from 1989 and second from Francis Collins (the current director of NIH), one of the people involved in the cystic fibrosis gene discovery.

> *The implications of this research are profound; there will be large spin offs in basic biology, especially cell physiology, but the largest impact will be biomedical (51).*

Here we are in 1997, eight years later, and the management of her disease has not changed. . . . .But I will predict that in the course of the next 10 years management of CF will change. . . . .The healthy form of the gene itself may even be used in so-called gene therapy (12).

What is interesting to note is that while gene therapy for cystic fibrosis has failed to materialize in the 20+ years since the gene defect was identified, there are traditional ion channel-based drugs that target the CFTR protein in clinical trials that show promise in cystic fibrosis (18, 66). At one level, the development of these drugs was likely facilitated by the genetic discoveries because they permitted the development of models that advanced the understanding of the biophysics and ultimately pharmacology of the defective channel. However, one is tempted to speculate, for cystic fibrosis and perhaps other diseases, that much faster therapeutic progress might have been made if traditional physiological and pharmacological approaches had been a bigger area of focus. Perhaps the optimism and drive for gene therapy was an example of what might be termed "silver bullet" thinking that I will discuss below.

REDUCTIONISM IS SEDUCTIVE

The type of reductionism that I have termed "biological orthopedic surgery" has a number of attractive features and is at some level very seductive. It is easy to understand, and when it delivers it is associated with a heroic narrative by a lone scientist or team of scientists making a fundamental discovery that solves a problem. This is the sort of silver bullet thinking mentioned above. However, it has been known for some time that both the easy to understand elements and heroic narratives associated with reductionism are mirages. In this context, when the factors that contribute to biomedical breakthroughs were subjected to analysis by Comroe and Drips (13) in the late 1960s and early 1970s via the "retrospectoscope," biomedical breakthroughs are in fact more nuanced, incremental, and associated with a more serendipitous view of progress vs. the heroic narrative of reductionism.

HEMOGLOBIN IS A SHIFTY MOLECULE

Homeostasis—the ability to regulate key bodily functions within a narrow range in response to either internal (e.g., exercise) or external (e.g., harsh environmental conditions)—is one of the fundamental (perhaps the fundamental) concept in physiology (7). Homeostasis is also subserved by ideas about regulated systems, feedback control, redundant control mechanisms, and adaptation and acclimation over time. These physiological concepts and mechanisms contribute to what might be described as emergent properties, so that the behavior of the system is far more complex and (and likely more robust) than might be predicted on the basis of a single reductionist property (35).

A good, and early, example of this concept comes from the textbook description about the right shift in the oxygen-hemoglobin dissociation curve that occurs at high altitude or during other forms of hypoxia. The standard teaching is that under these conditions there is a rise in 2–3 DPG that allosterically modifies oxygen-hemoglobin dissociation curve and creates a right shift that facilitates the unloading of oxygen at the tissues. However, when measurements of the oxygen-hemoglobin dissociation curve are made in humans who have traveled to high altitude (Fig. 1), under many circumstances there is in fact a net left shift in the oxygen hemoglobin dissociation curve. This left shift is facilitated by the rise in pH and fall in CO_2 caused by the hyperventilation driven by systemic hypoxia. Additionally, under some circumstances, it is driven further leftward by a fall in body temperature (68). Furthermore, it is of interest to note that all genetically adapted high altitude animals and the human fetus in the hypoxic intrauterine environment also have left shifted oxygen-hemoglobin dissociation curves, some with P50 values in the teens.

These observations make it seem likely that the main adaptive strategy is to shift the oxygen-hemoglobin dissociation curve to the left to facilitate the "loading" of oxygen at the lung in conditions (altitude) where oxygen availability is limited. This strategy also takes advantage of the fact that the mitochondria in the tissues can work efficiently at very low P_{O_2} values (and that under specific needs such as muscular exercise in hypoxia local increases in $[H^+]$ and temperature will reduce the leftward shift in muscle capillaries so that "unloading" of oxygen and tissue O_2 levels can be facilitated). It is also of note that the left shift in the oxygen-hemoglobin dissociation curve has been "known" since at least the 1920s. Along these lines, the sequencing of hemoglobin and the understanding of its biophysical properties was one of the earliest triumphs of what has come to be described as molecular biology (55). However, when the interpretation of such discoveries is too narrow, key physiological insights can be missed. The 2–3 DPG story is also an excellent and early example of how physiology trumps reductionist molecular biology as multiple systems and regulatory strategies interact to regulate homeostasis for the whole organism.

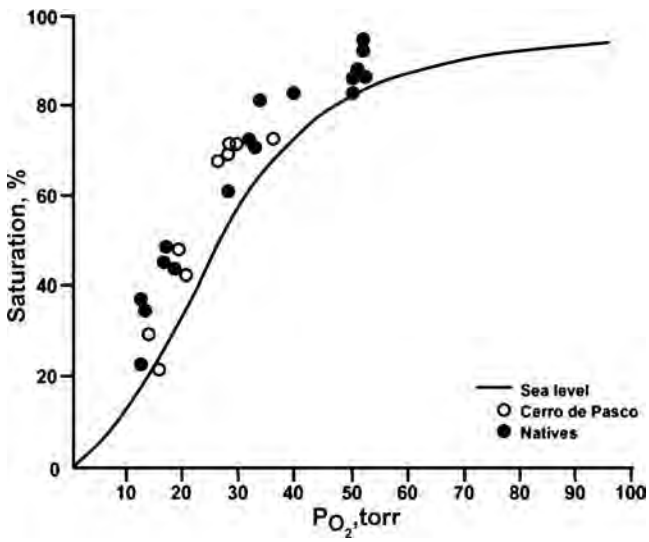


Fig. 1. Oxygen-hemoglobin dissociation curve demonstrating a left shift among sojourners (○) to high altitude and natives. The left shift in the oxygen-hemoglobin dissociation curve under these circumstances demonstrates that the combined effects of hypocapnia, increased pH, and cold override the simple effects of 2–3 DPG on the oxygen-hemoglobin dissociation curve. These data are an outstanding example of the limits of single mechanism reductionism. They are also consistent with the left shift seen in many genetically adapted animals that are native to high altitude. [Reprinted from Ref. 68, with permission from Elsevier.]

PREDICTIVE POWERS OF GENES?

In addition to gene therapy and other molecular treatments for rare diseases, reductionism also made promises about its ability to provide insight about who gets what complex disease like atherosclerosis, diabetes, hypertension, etc. As the quote below demonstrates, this idea became extremely popular after the sequencing of the human genome, and scientific funding agencies like the National Institutes of Health have invested huge sums of money in so-called "genome wide association studies" (GWAS) and other efforts to determine if a few genetic variants are harbingers of future disease in the population as a whole (10, 12, 43).

> *. . . because it been known all along that virtually every disease tends to track in families. What has changed is that. . . . .we are now beginning to see possible therapeutic approaches based on gene discoveries that will change the way medicine is practiced (12).*

One attractive element of this paradigm was that if a few common variants explained much of the risk for disease like diabetes, then it should be possible to identify those at risk and target them for early intervention. So far, the data from many, if not most or even all of these studies, have been underwhelming (43). First, a large number of variants seem to cause a significant increase in risk, but this increase is small compared with behavioral and environmental factors. An increased risk of several percent seems also likely to fall below what might be described as a phenotypic signal-to-noise ratio. Second, when the gene variants (single nucleotide polymorphisms, SNPs) that have been identified via GWAS or other experimental approaches are tested in large populations, the distribution of risk SNPs is typically strikingly similar in populations with and without disease (50, 63; Fig. 2). Third, when so-called genetic risk scores for disease are compared with predictive algorithms based on traditional risk factors (family history, lifestyle, age, etc.), the genetic risk scores are far less predictive than traditional phenotype-based risk scores. Furthermore, addition of genetic risk elements to phenotypically based scores adds little or no additional predictive power (50, 63). Finally, the idea that identifying prospective genetic risks for complex diseases that include a number of lifestyle and environmental factors (and increasingly even prenatal factors) is fundamentally wishful thinking, because behavioral health issues and culture play such a dominant role in determining who gets what disease when, and it is unclear if people will change their behavior in a positive way if they know prospectively they are at increased risk (24). Paradoxically, perhaps those at reduced genetic risk would pay less attention to behavioral risks.

SUCCESS IN PHARMACOGENOMICS AND ANTHROPOLOGY

So far, this paper has offered a sharp critique of the reductionists and taken the position that they over-sold what their technology had to offer on both the individual (gene therapy) basis and also in terms of population risk and intervention. However, there have been some notable successes stemming from application of this technology and two that seem especially worthy of comment. For example, there has been success in so-called pharmacogenomics. It has been well-known for some time that there are "responders" and "non-responders" to many forms of drug therapy. In many cases, this is related to how rapidly drugs are metabolized. In the case of tamoxifen, which had a dramatic effect on the recurrence of breast cancer, individuals with decreased drug metabolism appear to be at increased risk for recurrence. This is especially important for drugs like tamoxifen, which are ingested as pro-drugs with one or more metabolites that are active (56).

Another field where "omic" approaches have yielded dividends is anthropology. Two good examples include discoveries related to the independent development of lactase persistence into adulthood in areas of the world that were early adopters of herding (23, 34). In this context, one can imagine that the ability to digest lactose into adulthood provided the affected individuals a significant survival advantage and thus became the dominant genotype in only a few generations. Another good example that is perhaps counterintuitive relates to the individuals who migrated to the Tibetan plateau. These individuals do not develop chronic mountain sickness even with lifelong living at 3–4,000 m of elevation. These responses contrast to the high altitude natives in the Andes Mountains, who do develop chronic mountain sickness (58, 61, 70). Along these lines, those who migrated to the Tibetan plateau appear to have had selection pressure that favored a less functional variant of the hypoxia-inducible factor that, among other things, prevents them from developing excessive polycythemia, which plays a critical role in chronic mountain sickness.

INTERIM SUMMARY

So far, I have provided a general critique of what might broadly be termed "molecular reductionism". I have presented evidence that its failure to live up to its self-generated hype is in reality a failure to recognize larger ideas about homeostasis

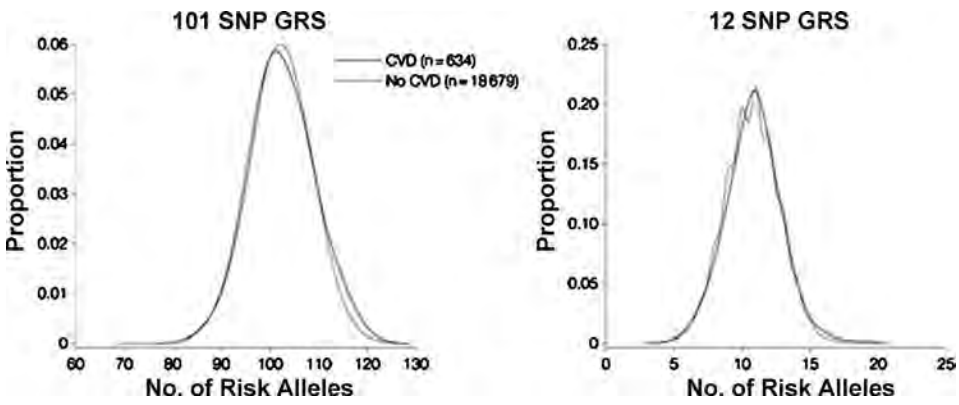


Fig. 2. Distribution of so-called high risk genes for cardiovascular disease in women with and without known coronary artery disease. The distribution of risk genes is similar, and construction of a genetic risk score for cardiovascular disease is thus problematic. This is just one example of the limited predictive power of "genomics" as it relates to the ability of relatively common gene variants to predict common diseases. [Borrowed with permission from Ref. 50. Copyright © 2010 American Medical Association. All rights reserved.]

and regulation that are central to physiology. This includes the specific example of the idea of gene therapy for relatively common genetic disorders like cystic fibrosis and also the limited predictive power of gene variants for common diseases. The question now is whether there is some way out of this problem and a better way to use potentially powerful technologies championed by the reductionists in a biomedical context.

IS SYSTEMS BIOLOGY THE ANSWER?

One idea to address the "failure" of molecular reductionism described above is to use a new approach called systems biology. The idea is that if powerful modeling tools and other data analysis techniques could be applied to the data generated via high throughput molecular reductionism, then somehow more meaningful insights would be generated and ultimately exploited for predictive or therapeutic purposes. The rationale for systems biology comes from a sampling of the comments on www.systemsbiology.org web site (34a).

> *Systems biology is the study of an organism, viewed as an integrated and interacting network of genes, proteins and biochemical reactions which give rise to life. Instead of analyzing individual components or aspects of the organism, such as sugar metabolism or a cell nucleus, systems biologists focus on all the components and the interactions among them, all as part of one system. These interactions are ultimately responsible for an organism's form and functions.*
>
> *Traditional biology—the kind most of us studied in high school and college, and that many generations of scientists before us have pursued—has focused on identifying individual genes, proteins and cells, and studying their specific functions. But that kind of biology can yield relatively limited insights about the human body.*
>
> *Biologists, geneticists, and doctors have had limited success in curing complex diseases such as . . . . diabetes because traditional biology generally looks at only a few aspects of an organism at a time.*

To a physiologist, there are obvious problems with systems biology. The problems start with the fact that physiology has been attempting for hundreds of years to understand the integrated function of organs and whole organisms that culminated in unifying big ideas about homeostasis and regulation discussed earlier. It is also clear that the type of biology that physiologists have been interested in starting with Harvey and the circulation has been about systems and has used modeling and computational techniques (1, 32, 57). Additionally, at this time the concept of systems biology and how it is defined remains very nebulous (52). Is systems biology a new discipline, an approach, a collection of tools, or merely a new name for integrative physiology generated by individuals who are generally unaware that our field exists (2, 28, 34a, 36, 40, 41, 45, 57)? Clearly physiology has provided and continues to provide insight about human disease, including insight that has led to vast therapeutic advances in recent years (37). Perhaps, the obvious question for the advocates of the cell-centric view of systems biology is did they skip physiology as part of their course work as students?

The concerns about systems biology outlined above at some level are about definitions and perhaps intellectual ownership. However, it also seems fair to ask what the long-term outlook for cell-centric systems biology is as an approach to making sense out of the vast amounts of data that can be generated using modern "omic" technology. In this context, there are key intellectual issues related to how data elements are generated, their spatial and temporal relationships, and how many ways they might interact (Fig. 3) that question the very fundamental assumptions about systems biology and its reliance on "bottom up" or "hypothesis neutral" modeling (2, 6, 15, 27, 35, 36, 38, 48, 67). It seems to me that without a narrative approach that includes hypothesis testing and key concepts like homeostasis, systems biology runs the risk of becoming scientific "Abstract Expressionism". Given the issues discussed earlier with gene therapy and GWAS approaches and the hype that surrounds systems biology, these concerns raise questions about what kind of science and scientific approaches deserve our future attention and funding (2, 24, 35).

REDUCTIONISM STALLS PHYSIOLOGY PROGRESSES

This is not the place for a comprehensive review of the contributions of physiology to biomedical research and therapeutic progress over the last 20–30 years. However, a few highlights that were initially seen as counterintuitive seem warranted. An obvious one is the discovery of EDRF and nitric oxide (25). This observation, which challenged the idea of the endothelium as merely a barrier, led to the discovery of gas-based signaling mechanisms and new therapeutic targets for conditions as diverse as erectile dysfunction and pulmonary hypertension. Would gas-based signaling mechanisms have been discovered by sequencing genes? Physiology has also helped redefine the optimal strategy used during mechanical ventilation in patients with adult respiratory distress syndrome (ARDS; 26). This has led to abandonment of strategies associated with high airway pressures and maintenance of arterial blood gases toward so-called permissive hypercapnia, alternate forms of mechanical ventilation and pressure support. Importantly, these new strategies that emphasize the avoidance of barotrauma have been associated with significant reductions in morbidity and mortality for ARDS. While part of the conventional wisdom now, this strategy was initially seen as counterintuitive.

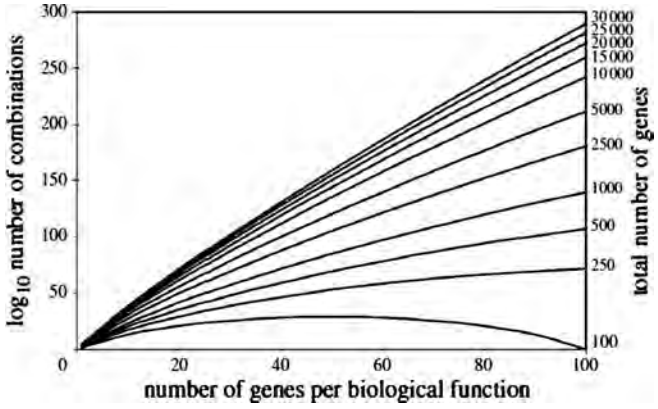


Fig. 3. Simulation of a number of possible combinations of genes gene interactions depending on the number of genes per biological function (*x*-axis) and the total number of genes in the organism. For biological functions with roughly 50 genes, $\sim 10^{150}$ possible combinations exist for most mammals. This figure shows the immense challenge associated with hypothesis-neutral systems biology and "bottom up" modeling. [Borrowed with permission from Ref. 46.]

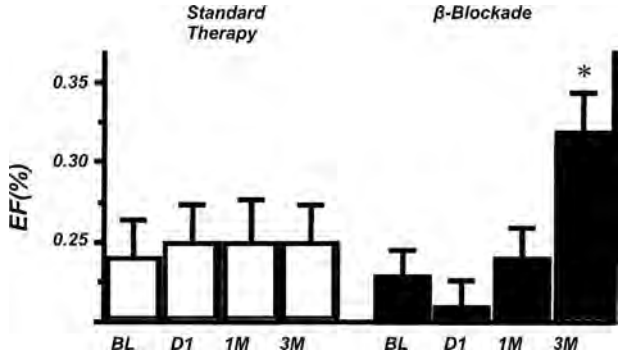


Fig. 4. Demonstration that beta-blockade can improve ventricular function (%EF) in humans with congestive heart failure over time. Standard therapy was associated with stable ventricular ejection fraction over 3 mo. By contrast, metoprolol (β-blockade) increased ventricular ejection fraction by >50% over 3 mo (*<0.05 vs. baseline). This finding, while initially counterintuitive, was based on sound physiological reasoning and along with other therapies has improved outcomes for patients with congestive heart failure. [Adapted from Ref. 20, with permission from Wolters Kluwer Health.]

Another example of a counterintuitive physiologically based clinical strategy was the use of beta-blockers in congestive heart failure. For many years these drugs were contraindicated in congestive heart failure (CHF) because it was felt that high sympathetic drive to the heart was required to maintain an adequate cardiac output in CHF. In reality, high sympathetic activity to the heart over time contributed to the progression of the disease and promoted a downward spiral of cardiac remodeling and reduced function (20). Thus the use of beta-blockers along with vasodilator therapy has been revolutionary and can interrupt or slow the downward spiral noted above in patients with congestive heart failure (Fig. 4). Again, the conventional wisdom was turned on its head and provided new insights that ultimately led to improved therapy. In the case of ARDS and congestive heart failure there has also been a two-way street between observations from clinical research conducted "at the bedside" to more fundamental observations in the laboratory.

Three other examples of more straight forward physiologically based therapeutic successes in recent years include the long story of improved outcomes for premature infants cared for in the neonatal ICU including altered ventilatory strategies, avoidance of oxygen toxicity, and surfactant therapy (9, 60). These improved outcomes, in the littlest ICU survivors, continue to seem miraculous to individuals who care for these patients and practiced medicine or nursing prior to their use. A second example has been oral rehydration solutions that are life saving in infants and children with diarrheal disease, especially in developing countries where it is a primary and frequent cause of death (8). Finally, in the developed world, where obesity and physical inactivity are leading to a pandemic of type 2 diabetes, physical activity (especially walking training in middle-aged people) has been proven to be highly effective in preventing, limiting, and in some cases reversing type 2 diabetes (16, 29). Each of these therapeutic successes is based on a foundation of physiologically based experimental evidence and insights.

REDUNDANCY, FEEDBACK, AND ACCLIMATION/ADAPTATION

Why has physiology continued to contribute in the era of reductionism? Physiologists are well versed in the overall concept of homeostasis, regulation, feedback, redundancy, and acclimation/adaptation. A classic example of redundancy comes from coronary circulation where coronary vasodilation is tightly linked to myocardial oxygen demand. In this context, a number of vasodilator systems likely contribute to this response. However, pharmacological blockade of one system, or in fact multiple systems, fails to alter this fundamental relationship between coronary vasodilation and myocardial oxygen demand in most species (19, 64; Fig. 5) This suggests that multiple redundant pathways contribute to this critical physiological response so that when one is blocked or absent, oxygen supply to the heart is not threatened when demand rises.

The fundamental relationship between coronary vasodilation and myocardial oxygen demand is also an observation that has had vast therapeutic implications and explains in large part why age specific death rates for cardiovascular disease have fallen dramatically over the last 30–40 years. There are drugs the reduce myocardial oxygen demand, mechanical therapy like stents, bypass surgery that improves myocardial oxygen delivery, and other drugs and lifestyle interventions that can affect both elements of the equation over time (30, 44). This physiological narrative and the progress that has flowed from it is in stark contrast to the relative lack of progress against cancer where there does not seem to be a unifying physiologically based story or model that can be exploited to address the general problem of cancer.

One of the classic feedback control mechanisms in physiology is the arterial baroreflex. While barodenervated animals have relatively normal blood pressure over a given 24 h period, their blood pressure becomes much more variable (14). The relative stability of blood pressure in the long run shows the power of redundant control via renal regulation of arterial pressure. However, for short-term adaptations, essential for things like exercise or changes in posture, feedback control

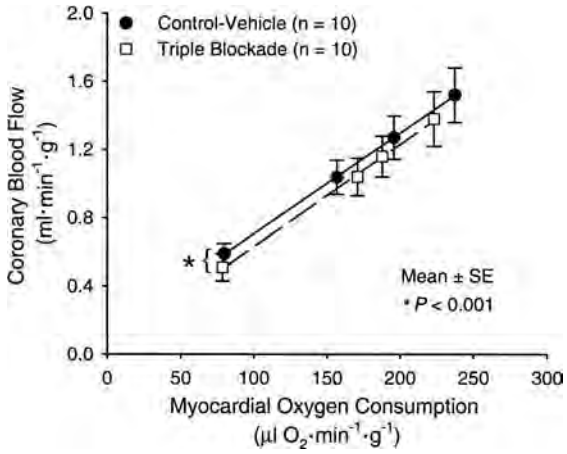


Fig. 5. Myocardial oxygen demand on the x-axis and coronary blood flow on the y-axis. Note that coronary blood flow rises in proportion to myocardial oxygen demand and that this rise is unaffected by triple inhibition of kATP+ channels, nitric oxide synthase, and adenosine receptor s This is a classic example of the concept of physiological redundancy. This well-known phenomenon may also explain why the absence of many so-called critical genes or proteins has limited impact on overall organ or organism function. This is because so-called redundant systems are able to alter their function and "upregulate" when one or more systems is blocked. [Borrowed with permission from Ref. 64.]

from arterial baroreflexes is essential for normal physiological responses.

An outstanding example of how humans acclimatize and adapt to physiological stress comes from studies that demonstrate that the ability of individuals to exercise in the heat can be remarkably improved by a few weeks of training in the heat (54). This improved exercise tolerance in the heat is associated with expanded plasma volume, increased sweating, and altered thermoregulatory skin blood flow. Another outstanding example is what might be called the adaptability of insulin sensitivity and glucose uptake in skeletal muscle. These variables are extremely sensitive to exercise and changes in daily activity and seem especially relevant in the era of the physical inactivity/obesity pandemic (29, 49, 53, 65).

Ideas about redundancy, feedback control, and acclimation/adaptation are also why physiologists are not that surprised by the ability of various gene knockout animals to survive and thrive (33). At some level this approach is conceptually similar to the classic denervation or high dose pharmacological blockade studies used by physiologists for generations and primarily show the power of the regulatory mechanisms highlighted above to preserve both long term phenotype and homeostasis despite the loss of one or more critical pathways or mechanisms (17). In this context, it is not surprising the yeast can survive without 80% of their genes and the function of these genes only becomes apparent when the organism is stressed (33). Is it too cynical to point out that knockout animals are essentially a "can't lose" experimental approach? If the knockout is lethal or leads to significant phenotypic dysfunction it is essential. If it survives then genetic or other compensatory mechanisms were upregulated to overcome the absence of the essential gene.

Physiology or physiologically based tests can also provide insight into the risk of future disease and/or predictive outcomes. For example, the blood pressure responses to common sympathoexcitatory stress can be used to define those at risk for future hypertension in a way that is potentially much more predictive than any current genetic test. Additionally, tests of autonomic function are strong predictors of outcomes in large populations of humans, and cardiorespiratory fitness is an especially good predictor of all-cause mortality.

TOOLS VS. BIG IDEAS

At some level molecular reductionism and systems biology are at existential cross roads. Are they in fact real disciplines informed by big ideas like homeostasis and regulation, or are they essentially tools and approaches that will facilitate the work of disciplines informed by bigger ideas and more importantly bigger questions and more comprehensive strategies? Based on the concepts and examples highlighted in this paper I would argue that until the vast amounts of data generated by modern "omic" techniques are put in a physiological context it will be an exercise in what Sydney Brenner has deemed "low input, high throughput, no output biology" (6). Along these lines, I want to end on an optimistic note with examples of how physiology is making a difference by applying reductionist tools as part of a more comprehensive approach to important questions. Because the Adolph lecture is sponsored by the Exercise and Environmental Physiology section of the American Physiological Society, relevant examples from related areas will be used. In each case there seems to be an overall hypothesis and a strategy that exploits what might be called responders and non-responders to an intervention.

Britton and Koch and colleagues (39, 69) have used selective breeding strategies to develop rats with vastly different inherent aerobic endurance capacities (Fig. 6). These animals have been used in a variety of studies to better understand the gene environment interactions. In many instances the animals selected for low intrinsic aerobic capacity seem to be at increased risk for complex diseases like diabetes, obesity, and heart disease. Additionally, studies using these animals have begun to identify genetic and transcriptional factors and networks that explain in part this increased risk (39).

Another example of how physiologists are using tools from the "new biology" is the HERITAGE study, which broadly seeks to understand the genetic basis for the differing physiological responses to exercise training in a large number of humans exposed to a standard protocol (3–5). This is an excellent example of how what might be called "high resolution" physiologically based phenotyping in conjunction with genetics. This hypothesis-driven approach also includes uses

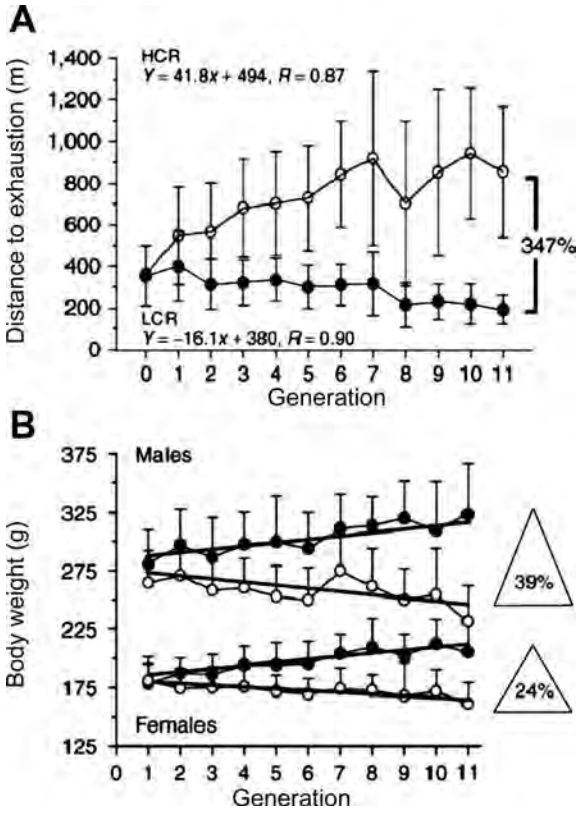


Fig. 6. Selective breeding of rats with divergent aerobic capacities. These data show that animals selected for their running capacity diverge dramatically after a few generations and is sustained for many generations. Importantly, at the same time body weight also began to diverge as did a number of risk factors for cardiometabolic disease. Phenotypic studies conducted on these animals in conjunction with more targeted forms of "omic" approaches and other types of molecular reductionism are providing new insights about gene environment interactions. These findings may also have applicability to physically active and inactive humans. The approach of Britton and Koch is a classic example of using reductionist tools in a physiological context to gain new insights with direct applicability to human health and disease. [Reprinted from Ref. 39 with permission from Macmillan Publishers Ltd. *Obesity Suppl*. copyright 2008.]

various "omic" and systems biology approaches and was initiated by physiologists before the terms genomics or systems biology existed. Additionally, like the examples from pharmacogenomics and anthropology discussed earlier, it takes advantage of the fact that there are responders and non-responders in response to a given intervention or environmental stressor.

Finally, my collaborator John Eisenach and I along with our colleagues have performed carefully controlled studies on how common genetic variants in the β_2-adrenergic receptor influence a number of physiological responses and how any genotype-based differences might be influenced by dietary sodium (21, 22, 31, 59). These studies were initiated because epidemiological evidence suggested that genetic variation in the β_2-adrenergic receptor influenced blood pressure in large populations. In our studies only homozygotes for the genetic variant of interest were recruited in an effort to see the maximum potential physiological effect of the variants. Using this approach, it appears that there are genotype-specific patterns associated with increased cardiac output responses to exercise that may interact with NO-mediated β_2-adrenergic receptor peripheral vasodilation. These responses clearly link and mechanistically define how a common gene variant in a key regulatory system can influence a physiological response in humans. They may also provide physiological explanations relevant to the original epidemiological observations on blood pressure and other outcomes, including those in patients with the acute coronary syndrome (42).

SUMMARY

In this paper and in the Adolph Lecture I have highlighted some of the claims associated with molecular reductionism and more recently systems biology. In both cases I have argued that the apparent inability and/or unwillingness of the advocates of these approaches to use key concepts from physiology and ultimately use their tools in a physiological context has limited the contribution of the approaches they advocate. By contrast physiology has continued to use new tools in the service of its big ideas and also continued to provide biomedical insight and therapeutic advances. As the final examples show, it is possible to incorporate reductionist tools in a physiological context to gain broader biomedical insights. Hopefully these insights will fuel the next wave of physiologically inspired therapeutic advances.

ACKNOWLEDGMENTS

The author thanks Drs. Nisha Charkoudian, Doug Seals, and Jerry Dempsey for critical reviews of the manuscript.

GRANTS

My laboratory has been funded continuously by National Institutes of Health since the early 1990s (HL-46493, HL-83947), and our work has also been facilitated by the former Mayo GCRC grant and more recently the CTSA grant (RR-024150). I have also been supported by the American Heart Association, the Mayo Foundation, and the Frank and Shari Caywood Professorship. I would also like to thank the fellows and collaborators who have worked with me, my outstanding technical support staff, and the many subjects who have volunteered for our studies.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author.

REFERENCES

1. **Auffray C, Noble D.** Origins of systems biology in William Harvey's masterpiece on the movement of the heart and the blood in animals. *Int J Mol Sci* 10: 1658–1669, 2009.
2. **Beard DA, Kushmerick MJ.** Strong inference for systems biology. *PLoS Comput Biol* 5: 1–10, 2009.
3. **Bouchard C, An P, Rice T, Skinner JS, Wilmore JH, Gagnon J, Perusse L, Leon AS, Rao D.** Familial aggregation of $\dot{V}O_{2max}$ response to exercise training: results from the HERITAGE Family Study. *J Appl Physiol* 87: 1003–1008, 1999.
4. **Bouchard C, Leon AS, Rao DC, Skinner JS, Wilmore JH, Gagnon J.** Aims, design, and measurement protocol. *Med Sci Sports Exerc* 27: 721–729, 1995.
5. **Bouchard C, Sarzynski MA, Rice TK, Kraus WE, Church TS, Sung YJ, Rao DC, Rankinen T.** Genomic predictors of maximal oxygen uptake response to standardized exercise training programs. *J* Appl Physiol 110: 1160–1170, 2011.
6. **Brenner S.** Sequences and consequences. *Phil Trans R Soc B* 365: 207–212, 2010.
7. **Cannon WB.** Organization for physiological homeostasis. *Physiol Rev* 9: 399–431, 1929.
8. **Cheng AC, McDonald JR, Thielman NM.** Infectious diarrhea in developed and developing countries. *J Clin Gastroenterol* 39: 757–773, 2005.
9. **Clements JA, Avery ME.** Lung surfactant and neonatal respiratory distress syndrome. *Am J Respir Crit Care Med* 157: S59–S66, 1998.
10. **Collins FS.** Contemplating the end of the beginning. *Genome Res* 11: 641–643, 2001.
11. **Collins FS.** Cystic fibrosis: molecular biology and therapeutic implications. *Science* 256: 774–779, 1992.
12. **Collins FS.** The human genome project and the future of medicine. *Ann NY Acad Sci USA* 882: 42–55, 1999.
13. **Comroe JH Jr, Dripps RD.** Ben Franklin and open heart surgery. *Circ Res* 35: 661–669, 1974.
14. **Cowley AW Jr, Liard JF, Guyton AC.** Role of the baroreceptor reflex in daily control of arterial blood pressure and other variables in dogs. *Circ Res* 32: 564–576, 1973.
15. **Csete ME, Doyle JC.** Reverse engineering of biological complexity. *Science* 295: 1664–1669, 2002.
16. Diabetes Prevention Program Research Group. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *N Engl J Med* 346: 393–403, 2002.
17. **Donald DE, Milburn SE, Shepherd JT.** Effect of cardiac denervation on the maximal capacity for exercise in the racing greyhound. *J Appl Physiol* 19: 849–852, 1964.
18. **Donaldson SH, Boucher RC.** Sodium channels and cystic fibrosis. *Chest* 132: 1631–1636, 2007.
19. **Duncker DJ, Bache RJ.** Regulation of coronary blood flow during exercise. *Physiol Rev* 88: 1009–1086, 2008.
20. **Eichhorn EJ, Bristow MR.** Medical therapy can improve the biological properties of the chronically failing heart. *Circulation* 94: 2285–2296, 1996.
21. **Eisenach JH, Barnes SA, Pike TL, Sokolnicki LA, Masuki S, Dietz NM, Rehfeldt KH, Turner ST, Joyner MJ.** Arg 16/Gly beta-2 adrenergic receptor polymorphism alters the cardiac output response to isometric exercise. *J Appl Physiol* 99: 1776–1781, 2005.
22. **Eisenach JH, McGuire AM, Schwingler RM, Turner ST, Joyner MJ.** The Arg16/Gly β-2 adrenergic receptor polymorphism is associated with altered cardiovascular responses to isometric exercise. *Physiol Genomics* 16: 323–328, 2004.
23. **Enattah NS, Jensen TGK, Nielsen M, Lewinski R, Kuokkanen M, Rasinpera H, El-Shanti H, Seo JK, Alifrangis M, Khalil IF, Natah A, Ali A, Natah S, Comas D, Mehdi SQ, Groop L, Vestergaard EM, Imtiaz F, Rashed MS, Meyer B, Troelsen J, Peltonen L.** Independent introduction of two lactase-persistence alleles into human populations reflects different history of adaptation to milk culture. *Am J Hum Genet* 82: 57–72, 2008.
24. **Evans JP, Meslin EM, Marteau TM, Caulfield T.** Genomics. Deflating the genomic bubble. *Science* 331: 861–862, 2011.
25. **Furchgott RF.** The discovery of endothelium-derived relaxing factor and its importance in the identification of nitric oxide. *JAMA* 276: 1186–1188, 1996.

26. **Gillette MA, Hess DR.** Ventilator-induced lung injury and the evolution of lung-protective strategies in acute respiratory distress syndrome. *Respir Care* 46: 130–148, 2001.
27. **Golub TR.** Counterpoint: Data first. Nature 464: 679, 2010.
28. **Greenhaff PL, Hargreaves M.** "Systems biology" in human exercise physiology: is it something different from integrative physiology? *J Physiol* 589: 1031–1036, 2011.
29. **Henriksen EJ.** Exercise effects of muscle insulin signaling and action. Invited review: effects of acute exercise and exercise training on insulin resistance. *J Appl Physiol* 93: 788–796, 2002.
30. **Heron M, Hoyert DL, Murphy SL, Xu J, Kochanek KD, Tejada-Vera B.** Deaths: Final data for 2006. *National Vital Statistics Reports*; vol 57, no 14. Hyattsville, MD: National Center for Health Statistics, 2009, pp 1–117.
31. **Hesse C, Schroeder DR, Nicholson WT, Hart EC, Curry TB, Penheiter AR, Turner ST, Joyner MJ, Eisenach JH.** Beta-2 adrenoceptor gene variation and systemic vasodilatation during ganglionic blockade. *J Physiol* 588: 2669–2678, 2010.
32. **Hester RL, Iliescu R, Summers R, Coleman TG.** Systems biology and integrative physiological modelling. *J Physiol* 589: 1053–1060, 2011.
33. **Hillenmeyer ME, Fung E, Wildenhain J, Pierce SE, Hoon S, Lee W, Proctor M, St. Onge RP, Tyers M, Koller D, Altman RB, Davis RW, Nislow C, Giaever G.** The chemical genomic portrait of yeast: uncovering a phenotype for all genes. *Science* 320: 362–365, 2008.
34. **Ingram CJE, Mulcare CA, Itan Y, Thomas MG, Swallow DM.** Lactose digestion and the evolutionary genetics of lactase persistence. *Hum Genet* 124: 579–591, 2009.
34a. **Institute for Systems Biology.** (Online). http://www.systemsbiology.org [June 22, 2011].
35. **Joyner MJ, Pedersen BK.** Ten questions about systems biology. *J Physiol* 589: 1017–1030, 2011.
36. **Joyner MJ.** Physiology: alone at the bottom, alone at the top. *J Physiol* 589: 1005, 2011.
37. **Joyner MJ.** Why physiology matters in medicine. *Physiology* 26: 72–75, 2011.
38. **Kell DB, Oliver SG.** Here is the evidence, now what is the hypothesis? The complementary roles of inductive and hypothesis-driven science in the post-genomic era. *Bioessays* 26: 99–105, 2004.
39. **Koch LG, Britton SL.** Development of animal models to test the fundamental basis of gene-environmental interactions. *Obesity* 16*, Suppl* 3: S28–S32, 2008.
40. **Kohl P, Crampin EJ, Quinn TA, Noble D.** Systems biology: an approach. *Clin Pharmacol Ther* 88: 25–33, 2010.
41. **Kuster DWD, Merkus D, Van Der Velden J, Verhoeven AJM, Duncker DJ.** Integrative Physiology 2.0: integration of systems biology into physiology and its application to cardiovascular homeostasis. *J Physiol* 589: 1037–1045, 2011.
42. **Lanfear DE, Jones PJ, Marsh S, Cresci S, McLeod HL, Spertus JA.** Beta2-adrenergic receptor genotype and survival among patients receiving beta-blocker therapy after an acute coronary syndrome. *JAMA* 294: 1526–1533, 2005.
43. **Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TFC, McCarroll SA, Visscher PM.** Finding the missing heritability of complex diseases. *Nature* 461: 747–753, 2009.
44. **Nelson RR, Gobel FL, Jorgensen CR, Wang K, Wang Y, Taylor HL.** Hemodynamic predictors of myocardial oxygen consumption during static and dynamic exercise. *Circulation* 50: 1179–1189, 1974.
45. **Noble D.** New-Darwinism, the Modern Synthesis and selfish genes: are they of use in physiology? *J Physiol* 589: 1007–1015, 2011.
46. **Noble D.** Differential and integral views of genetic in computational systems biology. *Interface Focus* 1: 7–15, 2011.
47. **Noble D.** Genes and causation. *Phil Trans R Soc A* 366: 3001–3015, 2008.
48. **Nurse P, Hayles J.** The cell in an era of systems biology. *Cell* 144: 850–854, 2011.
49. **Olsen RH, Krogh-Madsen R, Thomsen C, Booth FW, Pedersen BK.** Metabolic responses to reduced daily steps in healthy nonexercising men. *JAMA* 299: 1261–1263, 2008.
50. **Paynter NP, Chasman DI, Pare G, Buring JE, Cook NR, Miletich JP, Ridker PM.** Association between a literature-based genetic risk score and cardiovascular events in women. *JAMA* 303: 631–637, 2010.
51. **Pearson H.** One gene, twenty years. *Nature* 460: 165–169, 2009.
52. **Powell A, O'Malley MA, Muller-Wille S, Calvert J, Dupre J.** Disciplinary baptisms: a comparison of the naming stores of genetics, molecular biology, genomics, and systems biology. *Hist Phil Life Sci* 29: 5–32, 2007.
53. **Rogers MA, Yamamoto C, King DS, Hagberg JM, Ehsani AA, Holloszy JO.** Improvement in glucose tolerance after 1 wk of exercise in patients with mild NIDDM. *Diabetes Care* 11: 613–618, 1988.
54. **Robinson S, Turrell ES, Belding HS, Horvath SM.** Rapid acclimatization to work in hot climates. *Am J Physiol* 140: 168–176, 1943.
55. **Rossman MG.** Chapter 3: Recollection of the events leading to the discovery of the structure of haemoglobin. *J Mol Biol* 392: 23–32, 2009.
56. **Schroth W, Goetz MP, Hamann U, Fasching PA, Schmidt M, Winter S, Fritz P, Simon W, Suman VJ, Ames MM, Safgren SL, Kuffel MJ, Ulmer HU, Bolander J, Strick R, Beckmann MW, Koelbl H, Weinshilbum RM, Ingle JN, Eichelbaum M, Schwab M, Brauch H.** Association between CYP2D6 polymorphisms and outcomes among women with early stage breast cancer treated with tamoxifen. *JAMA* 302: 1429–1436, 2009.
57. **Secomb TW, Pries AR.** The microcirculation: physiology at the mesoscale. *J Physiol* 589: 1047–1042, 2011.
58. **Simonson TS, Yang Y, Huff CD, Yun H, Qin G, Witherspoon DJ, Bai Z, Lorenzo FR, Xing J, Jorde LB, Prchal JT, Ge R.** Genetic evidence for high-altitude adaptation in Tibet. *Science* 329: 72–75, 2010.
59. **Snyder EM, Beck KC, Dietz NM, Eisenach JH, Joyner MJ, Turner ST, Johnson BD.** Arg16Gly polymorphism of the β_2-adrenergic receptor is associated with differences in cardiovascular function at rest and during exercise in humans. *J Physiol* 571: 121–130, 2006.
60. **Sol RF.** Current trials in the treatment of respiratory failure in preterm infants. *Neonatology* 95: 368–372, 2009.
61. **Storz JF.** Genes for high altitudes. *Science* 329: 40–41, 2010.
63. **Talmud PJ, Hingorani AD, Cooper JA, Marmot MG, Brunner EJ, Kumari M, Kivimaki M, Humphries SE.** Utility of genetic and non-genetic risk factors in prediction of type 2 diabetes: Whitehall II prospective cohort study. *Br Med J* 340: b4838, 2010.
64. **Tune JD, Richmond KN, Gorman MW, Feigl EO.** K(ATP)(+) channels, nitric oxide, and adenosine are not required for local metabolic coronary vasodilation. *Am J Physiol Heart Circ Physiol* 280: H868–H875, 2001.
65. **van Dieren S, Beulens JW, van der Schouw YT, Grobbee DE, Neal B.** The global burden of diabetes and its complications: an emerging pandemic. *Eur J Cardiovasc Prev Rehabil* 17*, Suppl* 1: S3–S8, 2010.
66. **Verkman AS, Galietta LJV.** Chloride channels as drug targets. *Nat Rev* 8: 153–171, 2009.
67. **Weinberg RA.** Point: Hypotheses first. *Nature* 464: 678, 2010.
68. **Winslow RM.** The role of hemoglobin oxygen affinity in oxygen transport at high altitude. *Respir Physiol Neurobiol* 158: 121–127, 2007.
69. **Wisloff U, Najjar SM, Ellingsen O, Haram PM, Swoap S, Al-Share Q, Fernstrom M, Rezaei K, Lee SJ, Koch LG, Britton SL.** Cardiovascular risk factors emerge after artificial selection for low aerobic capacity. *Science* 307: 418–420, 2005.
70. **Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZXP, Pool JE, Xu X, Jiang H, Vinckenbosch N, Korneliussen TS, Zheng H, Liu T, He W, Li K, Luo R, Nie X, Wu H, Zhao M, Cao H, Zou J, Shan Y, Li S, Yang Q, Asan Ni P, Tian G, Xu J, Liu X, Jiang T, Wu R, Zhou G, Tang M, Qin J, Wang T, Feng S, Li G, Huasang Luosang J, Wang W, Chen F, Wang Y, Zheng X, Li Z, Bianba Z, Yang G, Wang X, Tang S, Gao G, Chen Y, Luo Z, Gusang L, Cao Z, Zhang Q, Ouyang W, Ren X, Liang H, Zheng H, Huang Y, Li J, Bolud L, Kristiansen K, Li Y, Zhang Y, Zhang X, Li R, Li S, Yang H, Nielsen R, Wang J, Wang J.** Sequencing of 50 human exomes reveals adaptation to high altitude. *Science* 329: 75–78, 2010.

© 2021. Published by The Company of Biologists Ltd | Development (2021) 148, dev197608. doi:10.1242/dev.197608

RESEARCH ARTICLE

Robustness and timing of cellular differentiation through population-based symmetry breaking

Angel Stanoev*, Christian Schröter and Aneta Koseska‡,§

ABSTRACT

During mammalian development and homeostasis, cells often transition from a multilineage primed state to one of several differentiated cell types that are marked by the expression of mutually exclusive genetic markers. These observations have been classically explained by single-cell multistability as the dynamical basis of differentiation, where robust cell-type proportioning relies on pre-existing cell-to-cell differences. We propose a conceptually different dynamical mechanism in which cell types emerge and are maintained collectively by cell-cell communication as a novel inhomogeneous state of the coupled system. Differentiation can be triggered by cell number increase as the population grows in size, through organisation of the initial homogeneous population before the symmetry-breaking bifurcation point. Robust proportioning and reliable recovery of the differentiated cell types following a perturbation is an inherent feature of the inhomogeneous state that is collectively maintained. This dynamical mechanism is valid for systems with steady-state or oscillatory single-cell dynamics. Therefore, our results suggest that timing and subsequent differentiation in robust cell-type proportions can emerge from the cooperative behaviour of growing cell populations during development.

KEY WORDS: Symmetry breaking, Differentiation, Cell-cell communication, Inhomogeneous steady state

INTRODUCTION

Functional diversification of cell types during mammalian development is characterised by the transition from an initially homogeneous group of multilineage primed cells towards a heterogeneous population of differentiated cell types (Zhang and Hiiragi, 2018; Simon et al., 2018). To ensure robust development, the onset of the differentiation event must be accurately timed, and the number distribution of each cell type must be correctly established.

The experimental observation that the expression of mutually exclusive genetic markers distinguishes the differentiated cell types from each other, and from the multilineage primed state, has promoted the hypothesis that cell types correspond to one of multiple stable gene expression states that arise from intracellular gene regulatory networks. Switching between these distinct states that dynamically represent stable attractors, specifically from the attractor encoding the multilineage primed state to the differentiated states, has been assumed as the dynamical basis of differentiation (Kauffman, 1969; Andrecut et al., 2011; Wang et al., 2011; Enver et al., 2009). The most common motif that accounts for bistability, i.e. the co-existence of two stable gene expression patterns on a single-cell level, is a two-component toggle-switch gene network (Thomas, 1981; Cherry and Adler, 2000; Snoussi, 1998). The addition of self-activating loops to each of the toggle-switch genes gives rise to a third stable state through which the multilineage primed co-expression state has been typically explained (Huang et al., 2007; Bessonnard et al., 2014; Jia et al., 2017). Such single-cell multistable circuits have been used to describe, for example, the Cdx2-Oct4 (also known as Pou5f1) switch in the differentiation of totipotent cells of the early embryo (Niwa et al., 2005), as well as the Gata6-Nanog switch in the differentiation of cells in the inner cell mass (Bessonnard et al., 2014; Chickarmane and Peterson, 2008; Schröter et al., 2015). The differentiation outcomes have thereby been typically analysed from a single-cell perspective, assuming that the intrinsic transitions towards one of the co-existing stable states are either driven stochastically (Gupta et al., 2011), modulated by an external signalling input, or stemming from cell-to-cell heterogeneity (De Mot et al., 2016; De Caluwé et al., 2019). However, the respective cell-type proportions strongly depend not only on the parameters of the system but also the initial conditions, indicating that the underlying mechanism of reliable proportioning cannot be directly inferred from this view.

On the other hand, experimental evidence suggests that paracrine signals have a crucial role for reliable cell-type establishment (Monk, 1997; Nichols et al., 2009; Yamanaka et al., 2010; Youk and Lim, 2014; Hart et al., 2014; Saiz et al., 2020). This has been particularly recognised in lateral inhibition models, such as the Delta-Notch system (Sprinzak et al., 2010; Collier et al., 1996), in which the intercellular communication realised through toggle switches distributed between the cells enables differentiation into distinct cell types with specific proportions and spatial organisation (Collier et al., 1996; Formosa-Jordan and Ibañes, 2014; Teomy et al., 2019 preprint; O'Dea and King, 2012). A similar model relying on cell-cell communication via fibroblast growth factors has been recently proposed for the cell-type specification in the mouse blastocyst (Saiz et al., 2020).

The emergence of specific gene expression profiles due to cell-cell communication is also well exemplified by the quorum-sensing mechanism in bacteria, in which the timing of emergence is regulated through the concentration of secreted signalling molecules (Taga and Bassler, 2003; You et al., 2004; Danino et al., 2010). Thus, these examples indicate that novel attractors can emerge in populations of communicating cells that are not present for isolated cells. This principle of attractor emergence in coupled systems has been extensively investigated in both natural and synthetic genetic networks (McMillen et al., 2002; Taga and Bassler, 2003; Kuznetsov et al., 2004; Garcia-Ojalvo et al., 2004; Ullner et al.,

Department of Systemic Cell Biology, Max Planck Institute of Molecular Physiology, 44227 Dortmund, Germany.
*Present address: Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA 20147, USA. ‡Present address: Lise Meitner Group Cellular Computations and Learning, Center of Advanced European Studies and Research (caesar), Ludwig-Erhard-Allee 2, D-53175 Bonn, Germany.

§Author for correspondence (aneta.koseska@caesar.de)

A.S., 0000-0001-6478-7496; C.S., 0000-0002-8161-7568; A.K., 0000-0003-4263-2340

Handling Editor: Paul François
Received 7 October 2020; Accepted 24 December 2020

Development (2021) 148, dev197608. doi:10.1242/dev.197608

2007). For example, emergent oscillatory solutions (Ullner et al., 2007, 2008; Koseska et al., 2010) have been proposed as dynamical mechanisms of stem cell differentiation with self-renewal (Suzuki et al., 2011). These theoretical and experimental findings suggest that cooperative dynamics could underlie the emergence of differentiated cell types in cell populations. However, the principles that govern self-organised differentiation timing, while at the same time leading to specific cell-type proportions that can reliably adapt upon perturbations, remain unresolved.

We propose a dynamical mechanism that underlies the differentiation of cell populations into cell types with stable proportions driven by the growth of the communicating population itself. Within this model, the expression profiles of both the multilineage primed and the differentiated cell types can be captured without any change in the parameters, and the transition between them can be established at a specific size of the population. The same mechanism also enables robust recovery of the acquired proportions upon perturbation of the system, accounting for autonomous scaling. We show that these properties of a communication-based cellular system arise from a collective symmetry breaking through a subcritical pitchfork bifurcation (PB). Dynamically, this novel collective state of mutually exclusive gene expression states in the population is represented as a single inhomogeneous steady state (IHSS). This renders the individual differentiated cell types dependent on each other, and the respective gene expression profiles different from those that can be attained in isolated cells. We show that the differentiation timing occurs through the growth of the population beyond a critical size, when the parameters of the undifferentiated cell population correspond to organisation before the PB. These findings indicate that the proposed collective inhomogeneous solution is a generic dynamical mechanism that describes how heterogeneous cellular types emerge from a homogeneous population of precursor cells and are maintained via cell-cell communication.

RESULTS

Timing and stable proportions of differentiated cell types are generated in growing populations

Cell differentiation during early mammalian embryogenesis often occurs from a multilineage primed state (*mlp*), which is maintained for several cell division cycles before differentiation into distinct cell types occurs (Fig. 1A) (Saiz et al., 2016; Hatakeyama et al., 2004; Soldatov et al., 2019). This implies that gene expression states transit from an initial homogeneous pattern across the precursor cells towards a heterogeneous expression state representing the differentiated cell types.

To investigate under which conditions such transition occurs, we considered a generic cell-cell communication system, in which the gene regulatory circuits in single cells are coupled to each other by paracrine signalling molecules s (Fig. 1B). The intracellular circuit consists of two genes u and v that inhibit the transcription of each other, and u additionally positively regulates the secretion of s. In turn, the extracellular concentration of the signalling molecules, s_{ext}, regulates the circuit dynamics (inhibiting u production; Fig. 1B). As the communicating signals are secreted by the cells themselves, their concentration is a variable in the system. This effectively

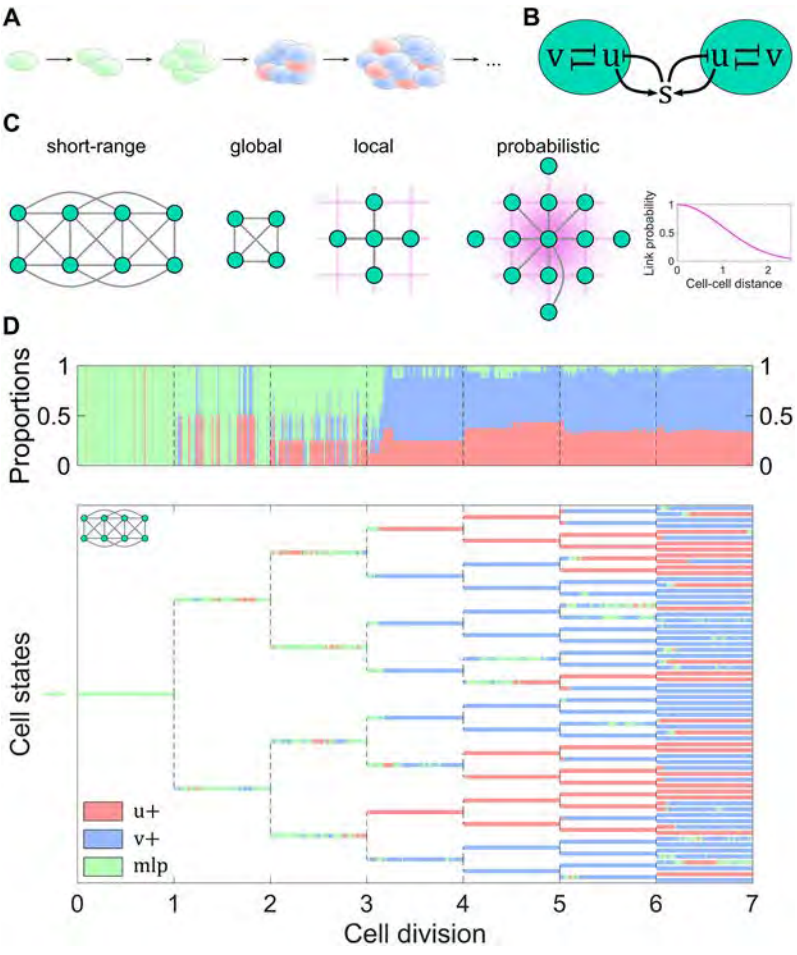


Fig. 1. Emergence of stable proportions of differentiated cell types in growing cell populations. (A) Schematic of the transition from multilineage primed cell types towards differentiated cell types during early embryogenesis. (B) Network topology of the cell-cell communication system (Eqn 1). (C) Different communication scenarios on a grid: short-range, global, local and distance-based probabilistic coupling. (D) Lineage tree depicting the transition from a multilineage primed state (*mlp*) to the *u*+/*v*+ states in a growing population with short-range communication, determined from a stochastic simulation. Green/red/blue: *mlp*/*u*+/*v*+ cells, respectively. Top: respective cell-type proportions. Parameters: α_u=2.3, α_v=3.5, α_s=2, $\alpha_{u,s}$=1, β=γ=δ=η=2 and λ=50.

Development (2021) 148, dev197608. doi:10.1242/dev.197608

establishes a single joint dynamical system that is described by the following dimensionless equations:

$$\begin{aligned}\frac{1}{\lambda}\frac{\mathrm{d}u_i}{\mathrm{d}t} &= \alpha_u \frac{1}{1+v_i^{\beta}} + \alpha_{u,s}\frac{1}{1+s_{i,ext}^{\eta}} - u_i \\ \frac{1}{\lambda}\frac{\mathrm{d}v_i}{\mathrm{d}t} &= \alpha_v \frac{1}{1+u_i^{\gamma}} - v_i \\ \frac{1}{\lambda}\frac{\mathrm{d}s_i}{\mathrm{d}t} &= \alpha_s \frac{u_i^{\delta}}{1+u_i^{\delta}} - s_i,\end{aligned} \tag{1}$$

where $i = \overline{1,N}$ is the single cell index and N is the number of cells. The vector (u_i, v_i, s_i) represents the state of each cell within the $3N$-dimensional state of the coupled system $(u_1, v_1, s_1, \ldots, u_N, v_N, s_N)$. The external amount of signal perceived by cell i from its communicating neighbourhood $N(i)$ (including itself) is defined as $s_{i,ext} = (1/(|N(i)|+1))\sum_{j\in(N(i)\cup i)} s_j$, i.e. the average amount of secreted signalling molecule at the given time. The parameters α_u, α_v, $\alpha_{u,s}$ and α_s represent the respective promoter strengths, whereas β, γ, η and δ are the Hill coefficients. Aiming to conceptualise the problem rather than to provide a quantitative description, the system was scaled to yield a minimal number of parameters: rate constants proportional to the promoter strengths were on the same order of magnitude, and the Hill coefficients were set to 2 (Materials and Methods).

Employing this model, we generated a numerical lineage tree starting from one cell, using stochastic simulations. The population growth was implemented in a simplified manner: after a given time period that mimics cell cycle length, all cells divide and the number of cells is doubled. The initial gene expression states of the daughter cells are inherited from the final state of the mother cell. The cells are placed on a grid with no flux boundary conditions, and the communication between them is short-ranged (within distance 2), i.e. between adjacent and second-adjacent cells on the grid (depicted for N=8 cells in Fig. 1C, left). Although we mainly focus on this communication type, the effect of other coupling types is also explored further below: global all-to-all, local (only between adjacent cell) and distance-based probabilistic coupling, in which interaction links between cells are established with decreasing probability as the distance between them increases (Fig. 1C). Cell divisions occur along the horizontal and vertical axes of the grid alternately, such that the grid doubles in size after each division event, yielding lattices of 1×1, 1×2, 2×2, 2×4, etc. (Fig. S1). The collective system state constituted of u+, v+ or *mlp* cell types was characterised in every time instance by categorising the cell states, and was used to plot the temporal evolution of the lineage tree (Fig. 1D). Furthermore, cellular proportions in the system were estimated from the collective state at each time point, and their temporal evolution is shown in the panel above the lineage tree.

The simulations demonstrated that even in the presence of gene expression noise (σ=0.1, Materials and Methods), for a population size of up to N=4 cells, the dynamical state of the system was homogeneous such that individual cells had equivalent u/v co-expression patterns resembling an *mlp* state (Fig. 1D). However, when the population reached a threshold size of N=8 cells, the expression patterns in single cells collectively transitioned to u- or v-dominated expression (u+ or v+ cells), indicating that the initial symmetry of the system had been broken and cells differentiated. Interestingly, defined proportions of u+ and v+ cells emerged upon differentiation, which were stably maintained over many rounds of cell division (Fig. 1D, top).

That the *mlp* state could be maintained for smaller population sizes, despite the presence of gene expression noise, indicates that stochastic events do not trigger the transition from a homogeneous to a heterogeneous cell population. As the model parameters also do not change, these results rather suggest that the timing of the transition event emerges from the growth of the population, upon which distinct proportions of the generated heterogeneous cell types are established.

Collective state with heterogeneous cell types emerges with precise timing due to subcritical organisation

To uncover the dynamical mechanism that underlies the transition between the *mlp* and the u+/v+ cell types, we performed a bifurcation analysis. The results in Fig. 1D indicated that differentiated cell types emerge with growing population size. However, a direct identification of the transition type using the number of cells as a bifurcation parameter cannot be performed, as it is not an explicit parameter in the model (Eqn 1). As a first step, we therefore chose α_u as a representative bifurcation parameter to identify and characterise the solutions of the system, whereas the impact of population size will be explored further below.

The bifurcation analysis showed that for N=2 coupled cells, the system exhibits qualitatively different dynamics for different α_u values. For lower α_u values, the system exhibits monostable behaviour (Fig. 2A, top, green circle for α_u=2.3). Here, both cells populate the same state (u_1=u_2) as evident from the projection that falls on the diagonal in the u_1–u_2 state space (Fig. 2A, bottom left). This homogeneous steady state (HSS) captures the *mlp* state with a characteristic precursor u/v co-expression pattern. On the level of a single-cell system, this is the only stable regime over the full α_u domain (Fig. S2A), meaning that isolated cells will maintain the *mlp* state.

However, at a critical α_u value, the symmetry of the *mlp* HSS is broken via a subcritical PB. From the point of view of dynamical systems, the HSS (Fig. 2A, black line) loses its stability at this parameter value and a pair of fixed points is generated, giving rise to IHSSs stabilised via saddle-node bifurcations (SN) (Fig. 2A, purple lines). There is a co-existence between the HSS and the IHSS before the PB, rendering it subcritical. The IHSS is a single attractor, here a six-dimensional point (u_1, v_1, s_1, u_2, v_2, s_2), that describes a heterogeneous state. This state consists of a high u-expression level (u+) in one cell (u_2) and a low u-expression level (v+) in the other cell (u_1), as shown by the two-dimensional (u_1, u_2) projection (Fig. 2A, bottom right). This means that the IHSS is a symmetry-broken collective state (Koseska et al., 2013), as it describes simultaneously both cell types with mutually exclusive gene expression patterns (Fig. S2B,C). As there is no preference which cell will acquire the u+ or v+ cell type, both possibilities (u_1<u_2 and u_1>u_2) are present as branches (upper and lower branch in Fig. 2A, top, respectively). Therefore, the same fixed point will be manifested as an upper branch for the u+ cell, but as a lower branch within the equivalent bifurcation diagram for the v+ cell. The u_1–u_2 state space projection also demonstrates that the IHSS solutions are reflections of one another over the diagonal (Fig. 2A, bottom right). This demonstrates that PB provides a unique mechanism for a dynamical transition from a homogeneous (*mlp*, HSS) to a single but heterogeneous (u+/v+, IHSS) state of the population, in which the differentiated cell types always jointly emerge. The described IHSS solution of a coupled system is fundamentally distinct from a bistable system on a single-cell level. There, the u+ and v+ cell types are described by two different steady states. Thus, in the absence of cell-cell communication, each cell in a population can independently transition to one of them.

Development (2021) 148, dev197608. doi:10.1242/dev.197608

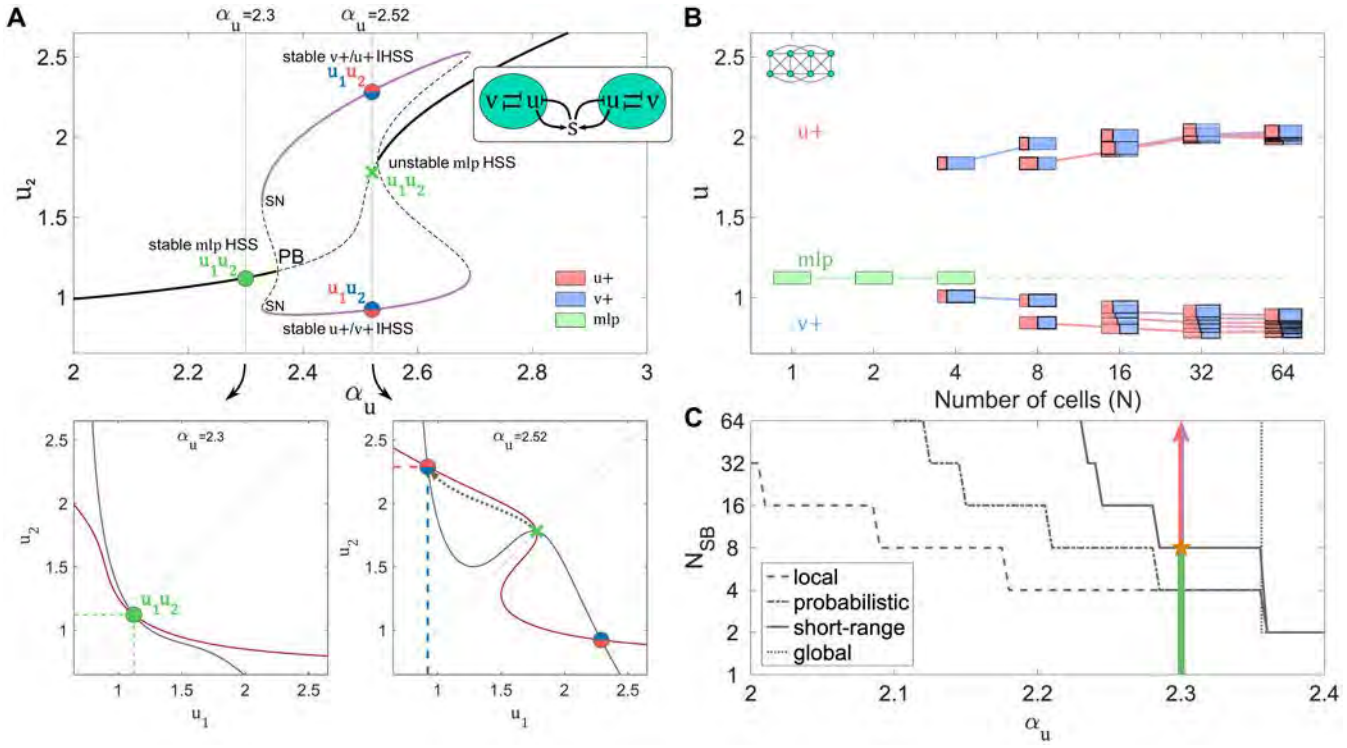


Fig. 2. Subcritical organisation before the PB enables timing of cellular differentiation. (A) Top: bifurcation diagram for N=2 coupled identical cells (inset), using u_2 as a representative variable. Solid lines indicate HSS (black) and IHSS (purple). Dashed lines indicate unstable steady states. Dotted lines indicate organisations in parameter space, with corresponding stable *mlp* HSS (solid green circle, α_u=2.3), *u*+/*v*+ IHSS (solid red/blue circles, α_u=2.52) and unstable HSS (green cross, α_u=2.52). Bottom: corresponding u_1–u_2 state space organisation. Left: stable *mlp* HSS (u_1=u_2) for α_u=2.3. Right: stable IHSS (u_1>u_2 and u_2>u_1) for α_u=2.52. (B) Bifurcation-like diagram depicting the emergence of IHSS distributions with increasing number of cells N. Green boxes indicate *mlp* HSS. Red and blue stack bar markers represent distinct *u*+/*v*+ IHSS proportions. Lines connect distributions with the same proportions. (C) The symmetry-breaking population size threshold (N_{SB}) at which the HSS becomes unstable in relation to α_u. For probabilistic distance-based coupling, the median (n=200) is shown. Green-to-red/blue arrow denotes the symmetry-breaking mlp-to-*u*+/*v*+ transition at N=8 cells (orange star) for α_u=2.3. In all panels, remaining parameters are the same as used in Fig. 1D.

In principle, all cells within a population could acquire *u*+ cell type, without the *v* + cell type ever occurring.

We next investigated which type of network topology could give rise to stable heterogeneous population states. We found that the IHSS emerges for a number of different network topologies that are characterised by inhibitory coupling, i.e. topologies with an effective negative feedback on *u* via the signalling molecule *s* (Fig. S3A to C; Eqn 2). Moreover, IHSS could be also obtained for oscillatory gene expression dynamics in isolated cells (Fig. S3D, Eqn 3). The bifurcation analysis additionally implies that heterogeneous cell types will be reached even when starting from identical initial conditions, representing a true symmetry breaking event.

Interestingly, the value for α_u=2.3 used in Fig. 1D lies before the PB, what we will term herein as a subcritical organisation of the system. The bifurcation diagram for N=2 cells (Fig. 2A) can therefore explain why in the lineage tree in Fig. 1D the cells maintained the *mlp* state and did not switch to the *u*+/*v*+ cell type, but it does not explain how the transition can occur for increasing population size N. We therefore generated a bifurcation-like diagram by varying N as in the lineage tree, while keeping all parameters fixed (with α_u=2.3) and using short-range communication (Fig. 2B). We identified the distinct collective states for each N via their *u*+/*v*+ cell-type proportions. The existence of steady states was estimated with an exhaustive stochastic search, with different initial conditions and noise intensities (Materials and Methods). The identified IHSSs with equivalent proportions were grouped, and the average *u*-values were depicted separately for the *u*+ and *v* + cell types within each distribution group. This yields paired upper and lower branches, analogously to the bifurcation diagram in Fig. 2A. Whereas for N=2 coupled cells only the *mlp* HSS was detected (Fig. 2B green; corresponding to the green circle in Fig. 2A), for N=4 coupled cells stable inhomogeneous solutions with a distinct proportion could be additionally identified (red/blue *u*+/*v*+ horizontal stack bar markers). This resembled the subcritical PB observed in Fig. 2A. The identified inhomogeneous solution corresponds to one cell having high- (*u*+) and three cells having low-expressing *u* state (*v*+), or 1*u*+/3*v* + distribution. However, for larger population sizes, the homogeneous *mlp* state lost its stability and only IHSSs reflecting the two cell types with specific proportioning were identified. In other words, the growth of the population to N=8 triggered a transition from the precursor to the differentiated cell state. This demonstrates that when α_u assumes a sufficiently low value to set the system before the bifurcation point, the growth of the population can trigger a dynamical transition, resembling the one that occurs when α_u is increased (Fig. 2A).

To explore how the population size N_{SB} at which there is an occurrence of symmetry breaking depends on α_u, we performed an equivalent of a two-parameter bifurcation analysis. The diagram shows that the symmetry-breaking transition could be triggered over a distinct α_u parameter region for short-range coupling (Fig. 2C, solid line). Depending on the α_u value before the PB, the symmetry-breaking point will be realised at a distinct size of the population (N_{SB}). Thus, for subcritical organisation, differentiation timings can emerge in a self-organised manner. Similar results were also obtained for local coupling and the probabilistic distance-based coupling

Development (2021) 148, dev197608. doi:10.1242/dev.197608

(Fig. 2C). On the other hand, for global all-to-all coupling, the symmetry-breaking transition could only be triggered stochastically with an increase in N. Therefore, these results suggest that under local and short-range coupling, as the population grows in size, the PB shifts its position towards lower α_u values. This is likely caused by the relative change in the effective communication range of the cells from global to a more local one with the size increase. The PB shift thereby enables transition of the system state from an HSS to an IHSS. Altogether, the analysis renders the number of cells as an effective bifurcation parameter that in conjunction with subcritical organisation drives the timing of cellular differentiation.

Reliable proportioning of differentiated cell types is a dynamical consequence of the sequential ordering of IHSS solutions

To investigate how the u+/v+ cell-type proportions emerge and are stabilised as the size of the population increases, we analysed the IHSS manifestation for N>2 cells in terms of bifurcation structure. For N=4 cells, the short-range and global coupling are equivalent, due to the small system size. The bifurcation analysis showed that at α_u=2.3, although for N=2 only the *mlp* HSS was stable (Fig. 2A), for N=4 coupled cells there was co-existence of the HSS and IHSS (Fig. 3A, equivalent to the result in Fig. 2B). The observed IHSS distribution was 1u+/3v+, with one cell having high- (u+) and three cells having low-expressing u state (v+), which for increasing α_u values was followed by 2u+/2v+, with two cells in each state, and 3u+/1v+, with three cells with high- and one with low-expressing u state (blue, purple and red branches in Fig. 3A, respectively). All of these distributions are associated with stable attractors that emerge from the same PB. In general, for N globally coupled cells, N−1 different distributions of cells between the u+ and v+ cell types are stable ([k]u+/[N−k]v+, for $k = \overline{1, N-1}$) (Koseska et al., 2010). These distributions are always sequentially ordered towards an increasing proportion of u+ cells for increasing α_u values. IHSS branches with adjacent distributions ([k]u+/[N−k]v+ and [k+1]u+/[N−k−1]v+) overlap, whereas the more dissimilar u+/v+ distributions are separated along the bifurcation parameter domain (Fig. 3A). Thus, for a given α_u, either a single IHSS distribution with

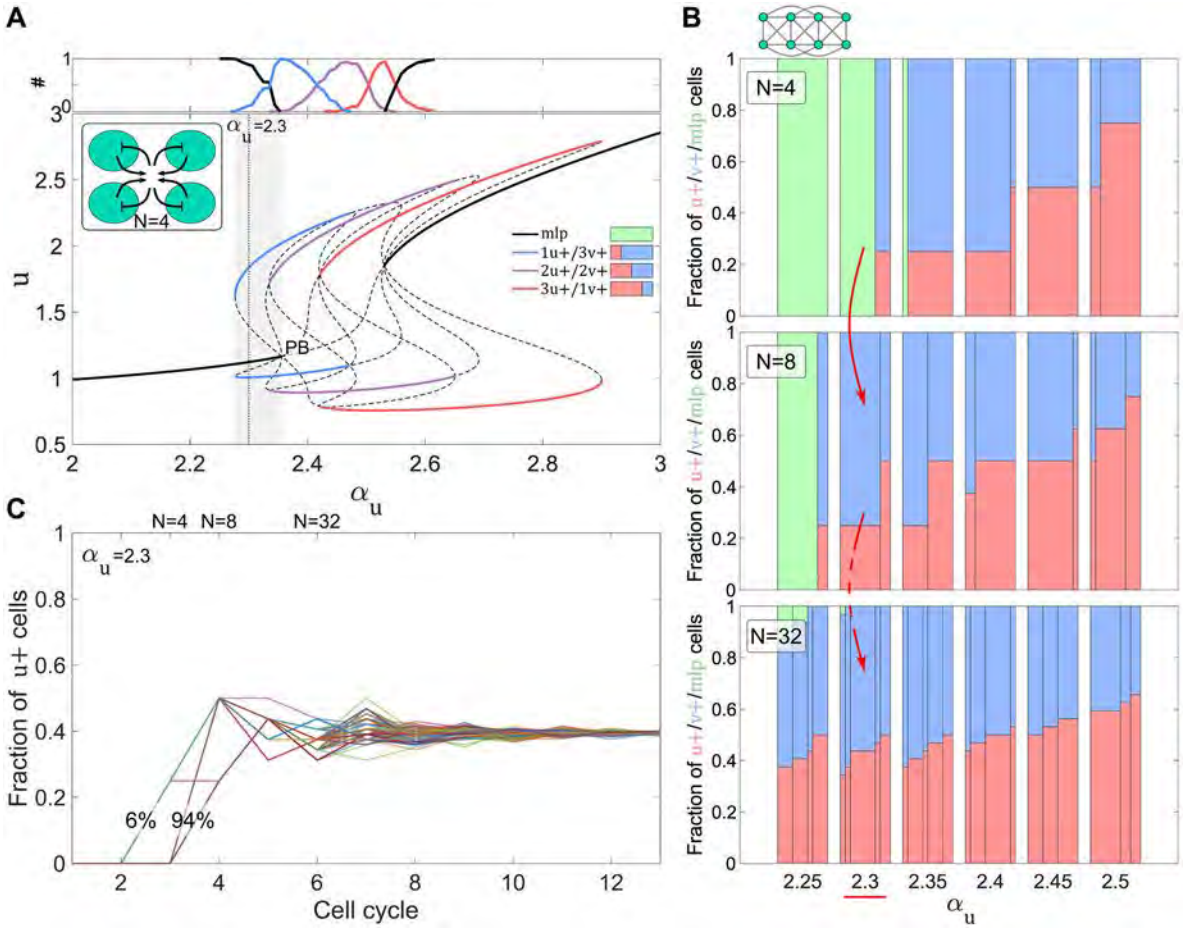


Fig. 3. Reliable cell-type proportions are established through stable IHSS distributions. (A) Bifurcation analysis for N=4 globally/short-range-coupled cells (inset). Solid lines indicate stable HSS (black) and IHSS (blue, purple and red) branches. Three stable IHSS distributions with increasing u+/v + cell-type ratios appear sequentially (key). Dashed lines indicate unstable steady states. Dotted lines indicate organisations in parameter space. The grey shaded area indicates the parameter range of HSS/IHSS co-existence for subcritical organisation. Top: probabilities for visiting IHSS distributions (y-axis) at the respective α_u values, estimated from 200 independent realisations. The initial conditions were randomly drawn from a normal distribution $\mathcal{N}(\mu_{ics}, \sigma^2_{ics})$ around the corresponding α_u-specific *mlp* state (μ_{ics}, σ_{ics}=0.2μ_{ics}). (B) u+/v+/*mlp* cell-type proportions for increasing α_u values, when a short-range coupled population of N=4 (top), N=8 (middle) and N=32 cells (bottom) on a 2×2, 2×4 and 4×8 grid is considered, respectively. The width of each sub-bar within a bar reflects the fraction of occurrence of the respective u+/v+/*mlp* proportion in ten independent realisations. Initial conditions were the same as in A (top). Red arrows indicate the progression of the distribution of stable cell-type proportions with increase of system size (N=4,8,32 cells) for α_u=2.3. (C) Temporal evolution of the fraction of u+ cells at each step of a lineage tree under short-range coupling and α_u=2.3. Results from 100 independent numerical simulations are shown. Other parameters are the same as in Fig. 1D.

Development (2021) 148, dev197608. doi:10.1242/dev.197608

distinct cell-type proportions, or adjacent branches with similar ratios, will be predominantly visited (Fig. 3A, top).

Intrinsic cell-to-cell variability in terms of circuit parameters does not affect the ordering of IHSS distributions and thereby cell-type proportioning. This was demonstrated by a bifurcation analysis on a coupled system of four cells with non-identical α_v values (Fig. S4A). Although in this scenario there are already slightly different *mlp* values in each cell, the parameter range over which the IHSS branches are stable is expanded (Koseska et al., 2009), and the overlapping intervals between adjacent distribution branches are reduced (compare Fig. S4A with Fig. 3A). Thus, the role of cell-to-cell variability is fundamentally different in coupled multicellular systems compared with multistable single-cell systems: it does not cause the symmetry-breaking event, but rather enhances its manifestation. Overall, these results indicate that the reliable cell-type proportions that emerge as the system transits to the differentiated state at a critical population size, are an inherent property of the distribution of IHSS solutions.

To investigate whether for larger population sizes the $u+/v+$ cell type proportions are indeed related to the sequential ordering of IHSS distributions, we compared the proportions for increasing α_u values at N=4, 8 and 32 short-range coupled cells. Multiple realisations of the numerical simulations were performed when starting from initial conditions randomly drawn from a normal distribution $\mathcal{N}(\mu_{ics}, \sigma^2_{ics})$ with a mean equal to the α_u-specific *mlp* state and s.d. $\sigma_{ics}=0.2\mu_{ics}$ (Fig. 3B). For N=4, the results were identical as those obtained from the bifurcation analysis: at α_u=2.3, $1u+/3v+$ proportion was detected, transiting to $2u+/2v+$ and $3u+/1v+$ as α_u was increased (Fig. 3B, top), corresponding to the sequential ordering of the branches and the probabilities for visiting them (Fig. 3A).

For N=8 short-range coupled cells, only the IHSS solutions were stable at α_u=2.3 (Fig. 3B, middle), showing again that the differentiation occurs at the critical transition from N=4 to N=8 cells (red arrow in Fig. 3B; as in Fig. 2B). The PB was shifted to a lower α_u value (α_u=2.281), thus enabling the differentiation timing to be realised in a self-organised manner. Furthermore, the $u+/v+$ cell type proportions were stabilised as the population increased (Fig. 3B, bottom, for N=32 cells). This shows that for a given parameter organisation, defined proportions can be reliably established and reproduced through multiple simulation realisations. We also corroborated the reliability of the timing mechanism by estimating the temporal evolution of the fraction of $u+$ cells across the different stages of the lineage tree (Fig. 3C). This showed that differentiation timing at N=8 cells, following the third cell cycle, was achieved in 94% of the cases, but also, the $u+/v+$ cell type proportions were reliably and stably maintained thereafter. The reliable cell-type proportioning was also manifested for the three additional coupling scenarios (Fig. 1C). In these cases, the increase in the proportion of $u+$ cells with an increase in α_u for N=32 cells is analogous to the progression of branches in the generic bifurcation analysis (N=4, Fig. 3A), although the HSS was destabilised at different α_u values for each of them (Fig. S4B-D). The variability between the cell-type proportions for a fixed α_u value is nevertheless with a $\lesssim$5% s.d. For local coupling, specifically, a 50%−50% ratio was maintained in a large α_u interval, indicating that the probability of visiting a different IHSS manifestation increases only for higher α_u values (Fig. S4C).

Therefore, this analysis showed that although different initial conditions can lead to different IHSS distributions, in most cases these are distributions with similar cell-type proportions. Thus, (1) reliable cell-type proportioning is a result of the sequential ordering of IHSS solutions in parameter space and (2) cell-cell variability enhances the manifestation of the identified symmetry breaking solution.

The IHSS distributions confer robust cell-type proportioning and mediate its recovery from perturbations

Our demonstration that the proposed symmetry-breaking mechanism is population-based suggests that robust cell-type proportions would be generated despite differences in initial conditions, and that they can be dynamically recovered upon perturbation. To probe the robustness property, we investigated the influence of different initial conditions distributions (changing the variance or shifting the mean value) that determine the initial gene expression states in single cells, as well as the effect of gene expression noise intensity. The results were obtained for a population of fixed size, N=32 cells, under the four distinct coupling types (Fig. 1C), and a fixed α_u value before the PB as before (α_u=2.3). Sampling the single-cell initial conditions from a normal distribution with increasing s.d. around the *mlp* value (Fig. 4A, top) produced distributions with reliably conserved $u+/v+$ cell-type proportions under short-range coupling (Fig. 4A, ratios with $\lesssim$8% s.d.). This demonstrates that variance in initial gene expression at the single-cell level does not majorly affect the differentiation outcome. Rather, cell types are established in characteristic proportions within a coupled system, even when starting from broad distributions of initial conditions. However, between the different coupling types, different stable $u+/v+/mlp$ proportions were generated for this fixed α_u, in agreement with previously estimated values (Fig. 3B; Fig. S4B-D): 0.45 for short-range (Fig. 4A), 0.5 for local (Fig. S5A) and 0.4 for probabilistic distance-based coupling (Fig. S5G), whereas the HSS remained stable against moderate perturbations for global coupling (Fig. S5D). Furthermore, stochastic realisations with a stepwise shift in the initial mean value from high v-expression to high u-expression state (Fig. 4B, top) also resulted in reliable cell-type proportions (Fig. 4B; Fig. S5B,E,H). The proportions were also reliable for simulations when multiplicative white noise intensity was increased (Fig. 4C; Fig. S5C,F,I, Materials and Methods). We also observed a manifestation in which, besides populating $u+/v+$ states within the IHSS solution, few cells also populated the *mlp* state, resembling a chimera-like state (Kuramoto and Battogtokh, 2002; Abrams and Strogatz, 2004). This was observed mainly for the probabilistic distance-based coupling (Fig. S5G-I), and for the short-range coupling in rare instances (Fig. 4A). As chimera states have been predominantly characterised for systems of coupled oscillators, an additional study would be required to dynamically classify these solutions.

We next explored whether this population-based symmetry-breaking mechanism could also underlie the ability of the early embryo to re-establish exact cell-type proportions following perturbations (Martinez Arias et al., 2013). To investigate this, we performed a numerical experiment in which cells were separated according to their type after the fourth cycle of the lineage tree (N=8 cells, Fig. 1D), forming two single-type subpopulations of different sizes that could further continue to grow and divide (Fig. 4D, left and right). The initial conditions of the two new lineage trees were therefore not positioned in the vicinity of the stable steady states. The subpopulation of two coupled cells with high u-expression reverted to the only stable attractor for this system size: the *mlp* HSS. However, after two cell cycles (reaching N=8 cells), both differentiated cell types re-emerged (Fig. 4D, left). The other subpopulation of N=6 cells with low u-expression initially transiently revisited the *mlp* state (attracted and then repelled by

Development (2021) 148, dev197608. doi:10.1242/dev.197608

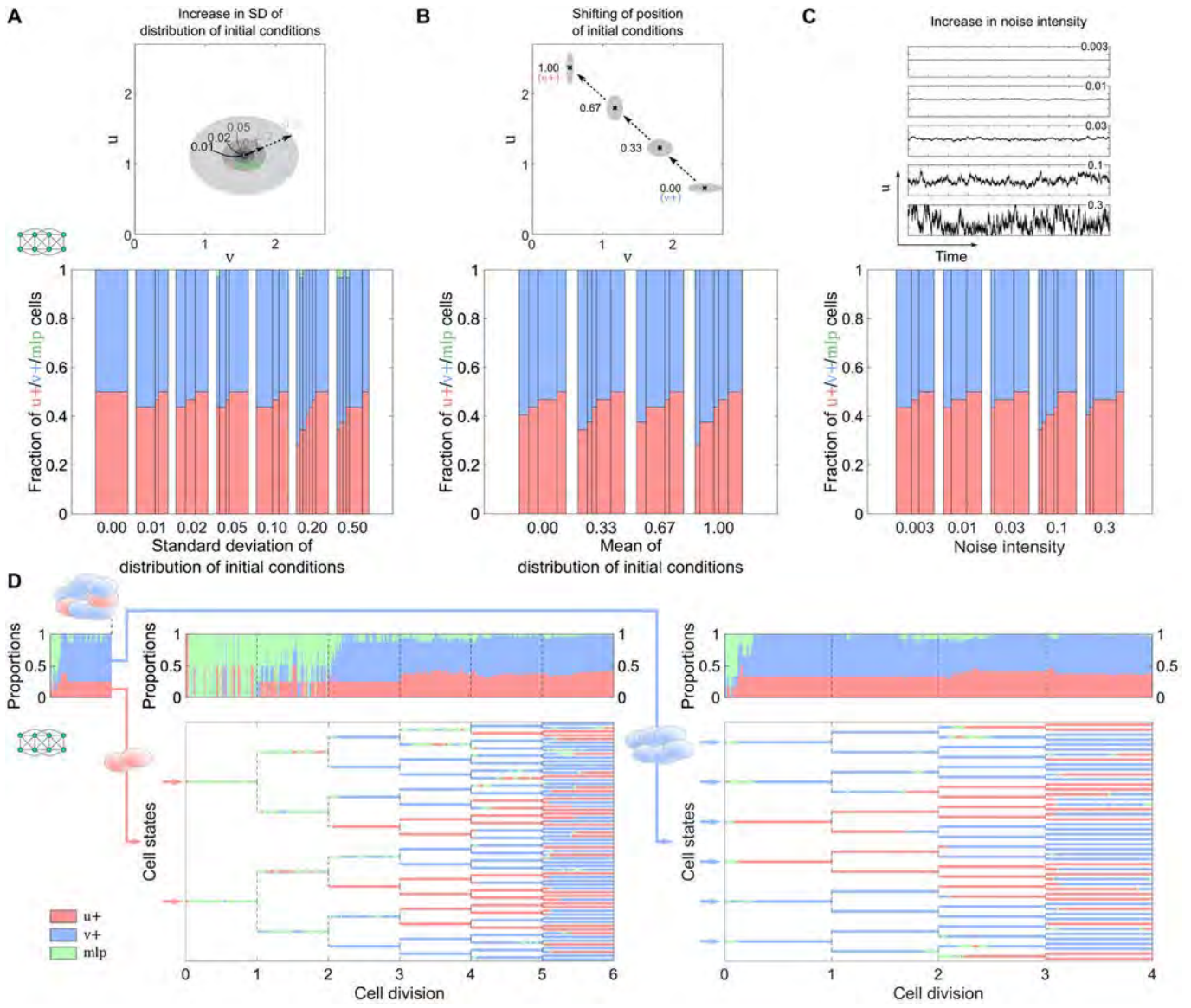


Fig. 4. Robustness in cell-type proportioning and plasticity in its re-establishment upon perturbations. (A-C) Cell-type proportions upon increase in the s.d. around the *mlp* initial conditions (as a fraction from the mean) (A); a shift in the mean of the distribution (μ_{ics}) value from high *v*-expression (μ_{v+}) to high *u*-expression cell state (μ_{u+}), with $\mu_{ics}=k\mu_{u+}+(1-k)\mu_{v+}$ for $k \in \left\{0, \frac{1}{3}, \frac{2}{3}, 1\right\}$ and $\sigma_{ics}=0.1\mu_{ics}$ (B); and an increase in the noise intensity (Materials and Methods) (C). Top: respective perturbations; 1-sigma ellipses of the distributions are depicted. Bottom: proportions from ten independent numerical realisations per condition are shown, estimated for N=32 short-range coupled cells on a 4×8 grid and organisation before the PB, α_u=2.3. (D) Lineage trees generated from homogeneous sub-populations of differentiated cells, separated at the fourth step of the lineage tree in Fig. 1D (N=8 cells). Left: lineage tree seeded from N=2 cells that previously had adopted a high *u*-expression state (*u*+). Right: lineage tree from the N=6 cells (2×3 grid) that before separation adopted the high *v*-expression state (*v*+). Upper panels show the respective cell-type proportions. Parameters are the same as in Fig. 1D.

the unstable saddle-type HSS) before both cell types stably re-emerged and the population settled in an IHSS during the first cell cycle (Fig. 4D, right).

The difference in timing between the two cases again points to a cell-number dependence in the triggering of the symmetry breaking (Fig. 2B). The cell-type ratios for both subpopulations were stabilised to a steady value similar to that of the full system before separation, and differed among the two subpopulations by ∼6%. This autonomous scaling and regenerating capability of the coupled system is a direct consequence of the properties of the IHSS solution: dynamically, it is not possible to stably populate the $u+$ without populating the $v+$ cell type. Thus, even when cells are separated such that only the cells of one type remain, the cell division and the cell-cell communication through which IHSS is established in the first place, will enable the system to recover both cell types with reliable ratios.

To confirm that both cell types and their proportions are indeed generated in a communication-dependent manner, we next investigated how inhibiting the cell-cell communication would affect the proportioning. We considered two different simulation scenarios, by implementing an increasing: (1) inhibition of α_s strength to mimic decreased production of the signalling molecules, and (2) inhibition of $\alpha_{u,s}$ strength, which effectively uncouples the dynamics of the intracellular circuit from the extracellular signalling. For this, α_s or $\alpha_{u,s}$ were effectively decreased by given multiplicative factors $(1-s_{inh,out})$ and $(1-s_{inh,in})$, respectively.

Development (2021) 148, dev197608. doi:10.1242/dev.197608

The simulations showed that in the first scenario, decreasing the production of the signalling molecule results in an increase of the *u*+ cells proportions, abruptly transiting to a homogeneous population of high *u*-expressing cells at 25%–30% of $s_{inh,out}$ (Fig. S6A). Decreasing $\alpha_{u,s}$ on the other hand, which decreases the overall transcription rate of *u*, reduces *u*+ cells proportions to abruptly transit to a homogeneous population of low *u*-expressing cells at $\sim 9\%$ of $s_{inh,in}$ (Fig. S6B). The single cells are now weakly coupled and, for the given parameters, the system is organised in the monostable low *u*-expressing state. The lack of robustness in cell-type proportioning observed under these conditions thus underlines the importance of population-level coordination in the system.

In summary, the presented results demonstrate that a subcritical organisation in conjunction with cell division within a communicating population ensures not only the robustness of the proportions of differentiated cell types with respect to initial conditions and noise, but can also enable re-establishment of this distribution upon perturbation.

IHSS leads to reproducible spatial patterns in growing cell populations

As the differentiation mechanism described here relies on cell-to-cell communication, we investigated whether the differentiated cells were randomly distributed or organised into spatial patterns. We performed stochastic lineage-tree simulations, with 13 cell cycles as in Fig. 3C, reaching a grid size of 64×64 cells. The *u*+/*v*+/*mlp* proportions were estimated from the stable steady states at the end of each cell cycle. To characterise the spatial distribution of cells, we quantified the extent of spatial clustering as the cluster radius of the *u*+ cells. Thus, for all *u*+ cells, the fraction of *u*+ cells in their surroundings with increasing distance was estimated (Fig. S7). The cluster radius was thereby the distance at which the fraction of *u*+ cells dropped to halfway between the fraction at zero distance (equal to 1), and the global fraction of *u*+ cells (dashed line in Fig. S7).

To systematically assess how the coupling range, i.e. the extent of signalling communication, contributes to the formation of patterns with distinct spatial frequencies, we used deterministic and probabilistic coupling with different ranges (1-, 2-, 3-, 5-, and 10-range, Materials and Methods). For the deterministic coupling schemes, distinct regular spatial patterns were formed, the features of which were strongly linked with the respective coupling range: local coupling generated checkerboard-like patterns, short-range coupling generated regular patterns of small *u*+ cell clusters (see also Fig. S8A, Movie 1), whereas stripe-like patterns were observed for increasing coupling (10-) range (see also Fig. S8B, Movie 2). Example patterns from the last stage of the lineage tree are shown in Fig. 5A. The clustering radius in these cases increased with the increase of the coupling range (Fig. 5B, left). In all the deterministic coupling scenarios, the patterns and their proportions were reliably reproduced for independent realisations of the lineage-tree simulations (Fig. 5C, left).

Interestingly, the spatial frequency of the formed stable patterns was preserved as the size of the population grew. This can be particularly exemplified for the case with 10-range coupling, in which doubling in the number of stripes followed the horizontal cell division events (Movie 2). This effectively rendered the width of the stripes, i.e. the *u*+ cluster radius and the distance between them, stable across the cycles (Fig. S8B). However, initially, at around the third cycle, a lower stable fraction of *u*+ cells was established and maintained by simple state propagation through cell division, which resulted in growth of the *u*+ cluster radius, until the critical system size was reached (eighth cycle) and a stable stripe-like pattern was formed (Fig. S8B). The pattern-directed *u*+ cluster size and cell-type proportions were further stably maintained. The simulations also showed that the cell population size at which the transition to the differentiated state was triggered was also preserved (Fig. S8A-D, lower left). Even more, the formed patterns and their characteristics were also preserved for a fixed population size as in the final grid, in which cells were randomly initialised (Fig. S9A-C).

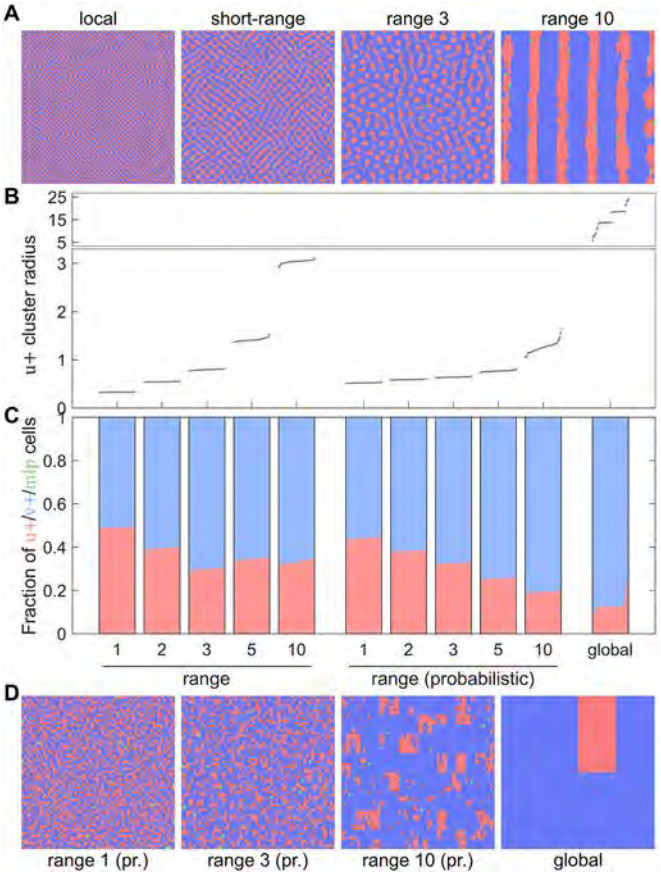


Fig. 5. Robust cell-type proportions are correlated with reproducible spatial patterns. (A) Exemplary spatial patterns for different coupling ranges: local, short-range, 3-range and 10-range coupling. (B) *u*+ cluster radius estimated at the final stable state of the lineage tree (64×64 size) for deterministic (left), probabilistic (middle) and global all-to-all coupling (right). Results from 100 independent stochastic realisations per coupling type are plotted in ascending order. (C) Corresponding *u*+/*v*+/*mlp* cell-type proportions. Bars are analogous to those in Fig. 3B. (D) Exemplary spatial patterns for range 1, 3 and 10 probabilistic coupling and global all-to-all coupling. Parameters are the same as in Fig. 1D.

On the other hand, when probabilistic distance-based coupling of different ranges was used, the lineage-tree simulations showed arrangement formation with less regular patterns (Movie 3 for 10-range coupling). Nevertheless, larger cluster size was again characteristic for larger communication ranges (Fig. 5B, middle; Fig. 5D). The proportioning was also reliably achieved over the different realisations (Fig. 5C, middle), in which the *u*+ fraction was maintained mainly by propagating the cell type from mother to daughter cells during division. Therefore, the *u*+ cluster radius is directed by the population growth, but is ultimately constrained by the communication range (Fig. S8C). That population growth highly regulated the formed patterns is also reflected through the formation of random spatial arrangements when populations of fixed size were randomly initialised (Fig. S9B-D).

Development (2021) 148, dev197608. doi:10.1242/dev.197608

The limiting case of these observations is the global all-to-all coupling scenario, which in most of the cases resulted in the formation of a single *u*+ cell-type cluster when population growth was considered (Fig. 5D, right; Fig. S8D, Movie 4). Such spatial arrangement is a direct consequence of the daughter cells inheriting their cell type from the mother cell along the lineage tree. In this case, the cell-type proportions were maintained globally, and the states of adjacent cells did not need to readjust to maintain the respective proportions locally. The relative size of the *u*+ cluster with respect to the population size was thereby only constrained by the cell type proportions. In contrast, for a fixed population size and starting from random initial conditions, a random arrangement of cell types across the grid was observed (Fig. S9D). Thus, in systems with global all-to-all or probabilistic coupling, the growing of the population is crucial for the observed spatial organisation.

Therefore, these results suggest that the IHSS not only represents a dynamical mechanism for generating stable proportions of differentiated cell types, but also enables their reproducible arrangement in regular patterns, the frequency of which is in turn constrained by the communication range.

DISCUSSION

We have argued here that intercellular communication, an integral process in developing mammalian embryos (Saiz et al., 2020; Lorthongpanich et al., 2012), gives rise to mutually exclusive differentiated cell types (Koseska and Bastiaens, 2017) as the population grows in size. We demonstrated that both the homogeneous undifferentiated and heterogeneous differentiated cell types, as well as the transition between them, can be described without changing the model parameters. This is valid when isolated cells are characterised with monostable steady state or oscillatory gene expression dynamics. The proposed simple but scalable mechanism enables robust cell-type proportions to emerge autonomously, but also describes their reliable recovery upon perturbation. This population-based heterogeneous solution is thereby distinct from the concept that single-cell multistability together with cell-to-cell variability are necessary to describe how differentiated cell types emerge. Such a single-cell scenario does not provide a mechanism for robust proportioning between differentiated cell types, but rather relies on tight regulation of initial states and signalling.

These two conceptual views have been recently subjected to an experimental test (Raina et al., 2020 preprint) using an *in vitro* model for robust proportioning of epiblast and primitive endoderm-like cell types in mouse embryonic stem cells. Starting from a broad range of initial gene expression profiles, the wild-type populations achieved robust epiblast and primitive endoderm cell-type proportions, in contrast to a communication-deficient mutant. The experiments also showed that the proportions could be reliably re-established in the wild type upon disproportionate removal of one cell type, in line with our predictions. Therefore, these observations corroborate the proposed hypothesis that balancing between robustness of differentiated cell-type proportions, while maintaining the plasticity of the system, such that it can recover to the exact proportioning upon perturbation, requires a population-based symmetry-breaking mechanism as realised by the IHSS.

Important insights regarding symmetry-breaking mechanisms unquestionably came from Turing's seminal work (Turing, 1952), and have been widely explored to describe the emergence of spatial organisation during development (Raspopovic et al., 2014; Economou et al., 2012). The IHSS similarly provides a dynamical transition from homogeneous *mlp* to a heterogeneous state of differentiated cell types within a population. Even more, the IHSS enables a reproducible spatial arrangement of the two differentiated cell types for a broad range of coupling scenarios, irrespective of the initial conditions. We have demonstrated that the formed pattern type is tightly related to the coupling range and the specified stable IHSS distribution. The spatial patterns identified in our simulations are broadly consistent with experimentally observed patterns in cell differentiation paradigms that could potentially be governed by an IHSS. Cell differentiation in the inner cell mass (ICM), for example, is triggered by a short-range communication signal (Raina et al., 2020 preprint). In mouse embryos and ICM organoids, the two cell types differentiating from the ICM form small clusters (Mathew et al., 2019; Fischer et al., 2020), similar to what we obtain in simulations with short-range coupling.

What is unique about the bifurcation transition presented here is not only the reliability of the differentiated cell type proportions and spatial organisation, but also the timing of the differentiation event that emerges as a result. The mechanism of differentiation timing is a unique property for organisation of the parameters of the system before the subcritical PB. A similar mechanism of population-based symmetry breaking, but via supercritical PB, has been previously suggested for the Delta-Notch lateral inhibition model, when the strength of the local interaction between the two cells is varied (Ferrell, 2012). However, in this case, the differentiation timing cannot be realised and the manifestation of only a specific cell-type proportioning with likely a spatial checkerboard-like pattern can be explained. A conceptually equivalent model to Delta-Notch has been recently described for cell differentiation in the mouse blastocyst, by relying on non-autonomous switching of gene expression circuits at a specific size of the population to recapitulate the emergence of differentiated cell types (Saiz et al., 2020). Thereby, the issue of self-organised differentiation timing characteristic for early embryo development has not been addressed.

The IHSS mechanism proposed here can be also taken one step further. An extension of the proposed mechanism can be envisioned that describes pluripotency and multipotency of stem cells. Conceptually, this would correspond to a finite cascade of subsequent PBs occurring simultaneously on both branches of the existing IHSS solutions (Zakharova et al., 2013; van Kekem and Sterk, 2019). IHSS therefore represents a cooperative dynamical mechanism through which a growing homogeneous cell population can break the symmetry, a prerequisite for novel information regarding different cellular types to emerge. Organisation in the vicinity of this dynamical transition allows the comprehensive capture of not only the differentiation timing, but also how robustness and accuracy during development are generated.

MATERIALS AND METHODS

Modelling a generic cell-cell communication system

To model the coupled system (Eqn 1), single cells were distributed spatially on a regular two-dimensional lattice, the dimensions of which are specified throughout the figures. The cells were coupled through paracrine signalling communication, by secreting and sensing the signalling molecule *s* within a certain distance. Four different communication ranges, *R*, were mainly considered, forming distinct network couplings (Fig. 1C): globally connected network (all-to-all communication, $R=\infty$); locally connected network (cells communicate only with adjacent cells on the lattice, $R=1$, i.e. within a unit lattice distance); short-range connected network (cells communicate with cells within distance $R=2$, i.e. with adjacent cells and second-adjacent cells on the lattice); and probabilistic distance-based coupling [cells establish communication with other cells with probability $e^{-(d^2/2R^2)}$, where *d* is the Euclidean cell-cell distance and $R=1$ in the default case]. No-flux or insulation boundary conditions on the lattices were used,

Development (2021) 148, dev197608. doi:10.1242/dev.197608

hence cells near the boundaries had fewer communicating neighbours, as exemplified by the schematic for short-range coupling on a 2×4 grid in Fig. 1C. In all cases, $s_{i,ext} = (1/(|N(i)|+1)) \sum_{j\in(N(i)\cup i)} s_j$ is the external amount of signal perceived by cell i, depicted as the averaged secreted signal from its communicating neighbourhood $N(i)$ (including itself) at every time instance, thus assuming an immediate mixing of s.

Each gene regulation term takes the renormalised Hill-type function form, $x^n/(1+x^n)$ for activation or $1/(1+x^n)$ for repression, and is therefore sensitive to values of the order of 1 of the transcription factor input x. The corresponding maximal transcription rates are α_u, α_v, α_s and $\alpha_{u,s}$, whereas β, γ, δ and η are the Hill coefficients. The reaction rates are globally scaled by λ. Values of α_u=2.3, α_v=3.5, α_s=2, $\alpha_{u,s}$=1, $\beta=\gamma=\delta=\eta$=2 and λ=50 were used throughout the study, unless noted otherwise. For the case of non-identical cells (Fig. S4A), the α_v parameter was uniformly varied between the cells in the range from −1% to 1% of its default value.

For the stochastic simulations (Fig. 1D, Fig. 2B, Fig. 3C, Fig. 4B-D, Fig. 5; Fig. S5B,C,E,F,H,I, Figs S8, S9, Movies 1-4), a stochastic differential equation model using Eqn 1 was constructed by adding a multiplicative noise $\sigma X \Delta W_t$, where ΔW_t is the Wiener process term, i.e. a normally distributed random variable with zero mean and variance Δt, whereas X is a state variable ($X\in\{u_1, v_1, s_1, \ldots, u_N, v_N, s_N\}$). The model was solved with Δt=0.01 using the Milstein method (Mil'shtein, 1974), by adding a second-order approximation term $0.5\sigma^2 X(\Delta W_t^2 - \Delta t)$. In the cases when the noise intensity σ was not varied, it was set to 0.1.

To discriminate between the multilineage-primed- (*mlp*), *u*-positive (*u*+) or *v*-positive (*v*+) cell types for a given realisation, each marginal cell state vector (u_i, v_i, s_i) within the converged state of the system (IHSS or HSS) was individually categorised as one of three cell types and the three-term *u*+/*v*+/*mlp* ratio (proportions) in the population was subsequently calculated. The reference *mlp* state vector (u_{mlp}, v_{mlp}, s_{mlp}) was predetermined for a given parameter set, i.e. for a specific value of α_u, from the respective steady-state value of the 1-cell monostable system realisation (as in Fig. S2A), as the bifurcation analysis demonstrated that the *mlp* HSS for single-cell and cell-to-cell coupled systems is equivalent. Every marginal cell state (u_i, v_i, s_i) for $i = \overline{1,N}$ of a deterministic realisation was categorised as *mlp* cell type when its value fell within 1% around the predetermined *mlp* state vector, whereas cell states whose u_i/v_i ratio was larger than the u_{mlp}/v_{mlp} ratio of the *mlp* state were assigned *u*+, and otherwise *v*+ cell types. Transient states during the stochastic realisations in Fig. 1D, Fig. 4D and Movies 1-4 were categorised as *mlp* type if they fell within 5% around the deterministic *mlp* state. End states of all stochastic realisations were allowed to converge to their deterministic attractor state in noise-free fashion before categorising (with 1% s.d. around the *mlp* state), as in Fig. 3C, Fig. 4B,C, Fig. 5; Figs S5B,C,E, F,H,I and Figs S8, S9.

Initial conditions for all cells and variables were independently randomly sampled from a normal distribution $\mathcal{N}(\mu_{ics}, \sigma^2_{ics})$, typically around the corresponding α_u-specific *mlp* state vector (u_{mlp}, v_{mlp}, s_{mlp}) as the mean (μ_{ics}), and with $\sigma_{ics}=0.1\mu_{ics}$ or $\sigma_{ics}=0.2\mu_{ics}$, as denoted in the respective figures. α_u-specific *mlp* states correspond to the steady-state values in the single-cell system (Fig. S2A). For Fig. 4A and Fig. S5A,D,G, σ_{ics} was increasing from 0 to 50% of the respective *mlp* value (μ_{mlp}). For Fig. 4B and Fig. S5B,E,H, μ_{ics} values were shifted stepwise along the line segment from a typical *v*+ cell state (μ_{v+}) towards a typical *u*+ cell state (μ_{u+}), using $\mu_{ics}=k\mu_{u+}+(1-k)\mu_{v+}$ for $k \in \left\{0, \frac{1}{3}, \frac{2}{3}, 1\right\}$, and with $\sigma_{ics}=0.1\mu_{ics}$.

Finally, for the results from ten repetitions, in which the identified IHSSs with equivalent proportions were grouped, each proportion was plotted as a stacked sub-bar within a bar plot, the width of which corresponded to the fraction of occurrence of that proportion in the simulations. Proportions were plotted in ascending order of their *u*+ cell-type fractions. The results from 100 repetitions (Fig. 5; Fig. S9) were plotted equivalently, without vertical lines separating the sub-bars of the proportions.

Estimating IHSS distributions as a function of the number of cells

In Fig. 2B, the different branches of the IHSS (and the proportions of cells in them) were estimated by analogy to Fig. 2A, but here using the number of cells as a bifurcation parameter. For this, exhaustive scanning was performed to locate the different fixed point attractors in the state space for each N. The scanning process involved 20 repetitive executions with different noise intensities (varying from 0 to 0.3). Each repetition consisted of 30 alternating cycles of stochastic execution (for exploration), followed by deterministic execution (for convergence to an attractor), when the reached state was tested for stability and subsequently recorded. For every distinctly detected attractor, the *u*+/*v*+/*mlp* proportion of cells was estimated, and attractors with equivalent proportions were grouped. The average *u* values were calculated separately for both *u*+ and *v*+ cell types for each proportion, and plotted as branches (*u*+ and *v*+ cells for IHSS, or *mlp* cells for HSS; colour-coding as in Fig. 2A), in analogy with the bifurcation diagrams in Fig. 2A and Fig. 3A. Chimera-like states were omitted from the diagram for simplicity.

Lineage tree generation

The generation of lineage trees was performed using stochastic simulations in which the system doubles in size at regular time intervals, starting from a single-cell system (Fig. 1D), unless otherwise specified (Fig. 4D). At every cell division, the final state of the mother cell is passed on to the initial conditions of the daughter cells. Cell divisions occur along the horizontal and vertical axes on the grid alternately, sequentially yielding lattices of 1×1, 1×2, 2×2, 2×4, 4×4, 4×8, 8×8, etc., as demonstrated in Fig. S1. Cellular states were categorised in every time instance to plot the single-cell temporal evolutions in the lineage trees (Fig. 1D, Fig. 4D). Furthermore, cellular proportions in the system were estimated from those values, and their temporal evolution was shown in the panels above the lineage trees. The fraction of *u*+ cells was plotted to yield Fig. 3C, Fig. 5C, Fig. S8 (lower left plots) and Fig. S9C.

In the cell-type separation case (Fig. 4D), the states of the cells at the end of the fourth cycle (from Fig. 1D) were categorised and the differentiated cells were then separated: *u*+ cells were given as seeds to a new lineage tree (1×2 grid), whereas *v*+ cells were seeds for a separate one (2×3 grid). Following this, multiple cell divisions were again performed and the cell proportions were estimated.

Analysis of spatial patterns

To characterise the spatial organisation of the system, lineage tree simulations with an extended duration of 13 cell cycles were performed, reaching a final size of a 64×64 grid. Both for deterministic and probabilistic distance-based coupling, communication with ranges of 1, 2, 3, 5 and 10 were considered. Additionally, global coupling was also used.

Spatial organisations were analysed at the end of each cell cycle, after the collective state reaches a steady state in a noise-free fashion. Fractions of *u*+/*v*+/*mlp* cells were quantified for the final spatial configuration (64×64 grid), and the results from 100 different repetitions were grouped as stacked bars. Moreover, the fractions of *u*+ cells were quantified for the end state of each cell cycle to track the development of the lineage tree (Fig. 3C; Fig. S8). The percentage of the numerical simulation realisations that showed symmetry breaking following a certain cell division cycle are also denoted on the plots (Fig. S8, lower left plots). Histograms depicting the proportion of numerical simulation realisations upon which a specific *u*+ cell-type fraction was reached in the final state of the lineage tree are also shown (Fig. S8, lower left plots).

To characterise the spatial clustering, the average *u*+ cluster radius was estimated from the spatial configurations by calculating for all *u*+ cells the fraction of *u*+ cells in the surround within increasing distance (Fig. S7). The cluster radius is therefore the distance by which the fraction of *u*+ cells drops to halfway between the fraction at zero distance (equal to 1), and the global fraction of *u*+ cells (dashed lines in Fig. S7), which is analogous to calculating half-life from exponential decay. The results from the realisations were plotted as single points in ascending order for each coupling range type in Fig. 5B and Fig. S9B. As with the fraction of *u*+ cells, the *u*+ cluster radius was also tracked over the development of the lineage tree, with a histogram being plotted for the final state as well. Both measures were also estimated when considering a population with a fixed size (64×64 grid), and starting from random initial conditions ($\mu_{ics}=\mu_{mlp}$, $\sigma_{ics}=0.1\mu_{ics}$) (Fig. S9).

The lineage tree simulations shown in Movies 1-4 were performed with a reduced number of time points (2000) per cell cycle to yield shorter movies. Spatial configurations with stochastic categorisation of the cell types were

Development (2021) 148, dev197608. doi:10.1242/dev.197608

saved for every tenth time point, and the saved frames were combined in a movie of 60 frames per second.

Decreasing cell-cell communication strength

For Fig. S6A, deterministic simulations were performed for a fixed population size on a 4×8 grid, short-range coupling (other parameters as in Fig. 1D) and starting from random initial conditions ($\mu_{ics}=\mu_{mlp}$, $\sigma_{ics}=0.1\mu_{ics}$). To model the decreased cell-cell communication strength, an inhibitory multiplicative term ($1-s_{inh,out}$), where $s_{inh,out}\in\{0, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3\}$, was introduced to the first term in the equation for s in Eqn 1. This enabled the effective reduction of the production strength of s by a given percentage. The corresponding results from ten independent realisations for each condition are presented in Fig. S6A.

Similarly, for Fig. S6B, an inhibitory multiplicative term ($1-s_{inh,in}$), where $s_{inh,in}\in\{0, 0.0153, 0.0307, 0.046, 0.0613, 0.0767, 0.092\}$, was introduced to the second term in the equation for u in Eqn 1. Other settings were the same as for Fig. S6A.

Additional cell-cell communication systems exhibiting population-based symmetry breaking

To demonstrate the generality of the population-based symmetry-breaking mechanism via a PB, we also tested several minimal network topologies with effective negative feedback coupling on u via s (Fig. S3A-C). Their dynamics were described using the following system of equations:

$$\begin{aligned}\frac{1}{\lambda}\frac{\mathrm{d}u_i}{\mathrm{d}t}&=\alpha_u\frac{1}{1+v_i^{\beta}}+\alpha_{u,s}\frac{1}{1+s_{i,ext}^{\eta}}-u_i\\ \frac{1}{\lambda}\frac{\mathrm{d}v_i}{\mathrm{d}t}&=\alpha_v\frac{1}{1+u_i^{\gamma}}+\alpha_{v,s}\frac{s_{i,ext}^{\eta}}{1+s_{i,ext}^{\eta}}-v_i\\ \frac{1}{\lambda}\frac{\mathrm{d}s_i}{\mathrm{d}t}&=\alpha_s\frac{u_i^{\delta}}{1+u_i^{\delta}}+\alpha_{s,v}\frac{1}{1+v_i^{\delta}}-s_i.\end{aligned}\tag{2}$$

For Fig. S3A, $\alpha_{u,s}=0$, $\alpha_s=1$, $\alpha_{v,s}=1$, $\alpha_{s,v}=0$ and $\alpha_v=2.75$. For Fig. S3B, $\alpha_{u,s}=0.5$, $\alpha_s=0$, $\alpha_{v,s}=0$, $\alpha_{s,v}=3$ and $\alpha_v=3$. For Fig. S3C, $\alpha_{u,s}=0$, $\alpha_s=0$, $\alpha_{v,s}=1$, $\alpha_{s,v}=2$ and $\alpha_v=3$.

Paradigmatic model mimicking the vertebrate neurogenesis process

It has been previously demonstrated that the presence of time delays in models of lateral inhibition can result in significant oscillatory transients before patterned steady states are reached. The impact of local feedback loops in a model of lateral inhibition based on the Notch signalling pathway, elucidating the roles of intracellular and intercellular delays in controlling the overall system behaviour, has also been proposed (Momiji and Monk, 2009). Here, we aimed to understand whether population-based PB can provide the dynamical background behind the observed symmetry-breaking phenomenon. As our aim was to demonstrate the validity of this concept, we omitted the molecular details of the Notch pathway example and modelled a generic case in which the gene expression dynamics in each cell were characterised by oscillatory behaviour, whereas intercellular communication between the N cells was realised in a global manner, for simplicity (N=2, Fig. S3D). The dynamics of the system were therefore described as:

$$\begin{aligned}\frac{\mathrm{d}p_i}{\mathrm{d}t}&=\alpha_p\frac{1}{1+q_i^{\beta}}+\alpha_{p,r}\frac{r_i^{\eta}}{1+r_i^{\eta}}-p_i\\ \frac{\mathrm{d}q_i}{\mathrm{d}t}&=\alpha_q\frac{1}{1+p_i^{\gamma}}-q_i\\ \frac{\mathrm{d}r_i}{\mathrm{d}t}&=\epsilon\left(\alpha_r\frac{1}{1+p_i^{\gamma}}-r_i\right)+2d(r_{i,ext}-r_i)\\ \frac{\mathrm{d}r_{i,ext}}{\mathrm{d}t}&=\frac{de}{N}\Sigma_j(r_j-r_{i,ext}).\end{aligned}\tag{3}$$

Here, p and q are two genes that mutually inhibit the expression of each other, r controls the production of the signalling molecule, the extracellular concentration of which is denoted as r_{ext}. This system has been demonstrated to exhibit synchronised oscillations in a population of communicating cells (Kuznetsov et al., 2004; Koseska et al., 2007). The parameters are as follows: α_q=5, $\alpha_{p,r}$=1, α_r=4, β=2, γ=2, η=2, ϵ=0.01, d=0.008 and de=1.

The numerical bifurcation analysis was performed using the XPP/AUTO software (http://www.math.pitt.edu/~bard/xpp/xpp.html). All simulations were performed using custom-made code in MATLAB (MATLAB and Statistics Toolbox Release R2020b, MathWorks).

Acknowledgements

We thank Philippe Bastiaens for numerous discussions, together with Peter Bieling and Dhruv Raina for critically reading the manuscript.

Competing interests

The authors declare no competing or financial interests.

Author contributions

Conceptualization: A.K.; Methodology: A.S., A.K.; Software: A.S.; Validation: A.S., C.S., A.K.; Formal analysis: A.S.; Investigation: A.S., A.K.; Writing - original draft: A.K.; Writing - review & editing: A.S., C.S., A.K.; Visualization: A.S.; Supervision: A.K.

Funding

The authors acknowledge funding from the Max-Planck-Gesellschaft.

Data availability

All data and code used in this study are available at www.github.com/astanoev/SymmetryBreaking.

Supplementary information

Supplementary information available online at https://dev.biologists.org/lookup/doi/10.1242/dev.197608.supplemental

Peer review history

The peer review history is available online at https://dev.biologists.org/lookup/doi/10.1242/dev.197608.reviewer-comments.pdf

References

Abrams, D. M. and Strogatz, S. H. (2004). Chimera states for coupled oscillators. *Phys. Rev. Lett.* **93**, 174102. doi:10.1103/PhysRevLett.93.174102

Andrecut, M., Halley, J. D., Winkler, D. A. and Huang, S. (2011). A general model for binary cell fate decision gene circuits with degeneracy: indeterminacy and switch behavior in the absence of cooperativity. *PLoS ONE* **6**, e19358. doi:10.1371/journal.pone.0019358

Bessonnard, S., De Mot, L., Gonze, D., Barriol, M., Dennis, C., Goldbeter, A., Dupont, G. and Chazaud, C. (2014). Gata6, Nanog and Erk signaling control cell fate in the inner cell mass through a tristable regulatory network. *Development* **141**, 3637-3648. doi:10.1242/dev.109678

Cherry, J. L. and Adler, F. R. (2000). How to make a biological switch. *J. Theor. Biol.* **203**, 117-133. doi:10.1006/jtbi.2000.1068

Chickarmane, V. and Peterson, C. (2008). A computational model for understanding stem cell, trophectoderm and endoderm lineage determination. *PLoS ONE* **3**, e3478. doi:10.1371/journal.pone.0003478

Collier, J. R., Monk, N. A. M., Maini, P. K. and Lewis, J. H. (1996). Pattern formation by lateral inhibition with feedback: a mathematical model of delta-notch intercellular signalling. *J. Theor. Biol.* **183**, 429-446. doi:10.1006/jtbi.1996.0233

Danino, T., Mondragón-Palomino, O., Tsmiring, L. and Hasty, J. (2010). A synchronized quorum of genetic clocks. *Nature* **463**, 326-330. doi:10.1038/nature08753

De Caluwé, J., Tosenberger, A., Gonze, D. and Dupont, G. (2019). Signalling-modulated gene regulatory networks in early mammalian development. *J. Theor. Biol.* **463**, 56-66. doi:10.1016/j.jtbi.2018.12.008

De Mot, L., Gonze, D., Bessonnard, S., Chazaud, C., Goldbeter, A. and Dupont, G. (2016). Cell fate specification based on tristability in the inner cell mass of mouse blastocysts. *Biophys. J.* **110**, 710-722. doi:10.1016/j.bpj.2015.12.020

Economou, A. D., Ohazama, A., Porntaveetus, T., Sharpe, P. T., Kondo, S., Basson, M. A., Gritli-Linde, A., Cobourne, M. T. and Green, J. B. A. (2012). Periodic stripe formation by a Turing mechanism operating at growth zones in the mammalian palate. *Nat. Genet.* **44**, 348. doi:10.1038/ng.1090

Enver, T., Pera, M., Peterson, C. and Andrews, P. W. (2009). Stem cell states, fates, and the rules of attraction. *Cell Stem Cell* **4**, 387-397. doi:10.1016/j.stem.2009.04.011

Ferrell, J. E. (2012). Bistability, bifurcations, and Waddington's epigenetic landscape. *Curr. Biol.* **22**, R458-R466. doi:10.1016/j.cub.2012.03.045

Fischer, S. C., Corujo-Simon, E., Lilao-Garzon, J., Stelzer, E. H. and Muñoz-Descalzo, S. (2020). The transition from local to global patterns governs the

differentiation of mouse blastocysts. *PLoS ONE* **15**, e0233030. doi:10.1371/journal.pone.0233030
Formosa-Jordan, P. and Ibañes, M. (2014). Competition in notch signaling with cis enriches cell fate decisions. *PLoS ONE* **9**, e95744. doi:10.1371/journal.pone.0095744
Garcia-Ojalvo, J., Elowitz, M. B. and Strogatz, S. H. (2004). Modeling a synthetic multicellular clock: repressilators coupled by quorum sensing. *Proc. Natl Acad. Sci. USA* **101**, 10955-10960. doi:10.1073/pnas.0307095101
Gupta, P. B., Fillmore, C. M., Jiang, G., Shapira, S. D., Tao, K., Kuperwasser, C. and Lander, E. S. (2011). Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. *Cell* **146**, 633-644. doi:10.1016/j.cell.2011.07.026
Hart, Y., Reich-Zeliger, S., Antebi, Y. E., Zaretsky, I., Mayo, A. E., Alon, U. and Friedman, N. (2014). Paradoxical signaling by a secreted molecule leads to homeostasis of cell levels. *Cell* **158**, 1022-1032. doi:10.1016/j.cell.2014.07.033
Hatakeyama, J., Bessho, Y., Katoh, K., Ookawara, S., Fujioka, M., Guillemot, F. and Kageyama, R. (2004). Hes genes regulate size, shape and histogenesis of the nervous system by control of the timing of neural stem cell differentiation. *Development* **131**, 5539-5550. doi:10.1242/dev.01436
Huang, S., Guo, Y.-P., May, G. and Enver, T. (2007). Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. *Dev. Biol.* **305**, 695-713. doi:10.1016/j.ydbio.2007.02.036
Jia, D., Jolly, M. K., Harrison, W., Boareto, M., Ben-Jacob, E. and Levine, H. (2017). Operating principles of tristable circuits regulating cellular differentiation. *Phys. Biol.* **14**, 035007. doi:10.1088/1478-3975/aa6f90
Kauffman, S. A. (1969). Metabolic stability and epigenesis in randomly constructed genetic nets. *J. Theor. Biol.* **22**, 437-467. doi:10.1016/0022-5193(69)90015-0
Koseska, A. and Bastiaens, P. I. H. (2017). Cell signaling as a cognitive process. *EMBO J.* **36**, 568-582. doi:10.15252/embj.201695383
Koseska, A., Volkov, E., Zaikin, A. and Kurths, J. (2007). Inherent multistability in arrays of autoinducer coupled genetic oscillators. *Phys. Rev. E* **75**, 031916. doi:10.1103/PhysRevE.75.031916
Koseska, A., Volkov, E. and Kurths, J. (2009). Detuning-dependent dominance of oscillation death in globally coupled synthetic genetic oscillators. *EPL* **85**, 28002. doi:10.1209/0295-5075/85/28002
Koseska, A., Ullner, E., Volkov, E., Kurths, J. and García-Ojalvo, J. (2010). Cooperative differentiation through clustering in multicellular populations. *J. Theor. Biol.* **263**, 189-202. doi:10.1016/j.jtbi.2009.11.007
Koseska, A., Volkov, E. and Kurths, J. (2013). Transition from amplitude to oscillation death via Turing bifurcation. *Phys. Rev. Lett.* **111**, 024103. doi:10.1103/PhysRevLett.111.024103
Kuramoto, Y. and Battogtokh, D. (2002). Coexistence of coherence and incoherence in nonlocally coupled phase oscillators. *Nonlinear Phenom. Complex Syst.* **5**, 380-385.
Kuznetsov, A., Kærn, M. and Kopell, N. (2004). Synchrony in a population of hysteresis-based genetic oscillators. *SIAM J. Appl. Math.* **65**, 392-425. doi:10.1137/S0036139903436029
Lorthongpanich, C., Doris, T. P. Y., Limviphuvadh, V., Knowles, B. B. and Solter, D. (2012). Developmental fate and lineage commitment of singled mouse blastomeres. *Development* **139**, 3722-3731. doi:10.1242/dev.086454
Martinez Arias, A., Nichols, J. and Schroeter, C. (2013). A molecular basis for developmental plasticity in early mammalian embryos. *Development* **140**, 3499. doi:10.1242/dev.091959
Mathew, B., Muñoz-Descalzo, S., Corujo-Simon, E., Schröter, C., Stelzer, E. and Fischer, S. C. (2019). Mouse icm organoids reveal three-dimensional cell fate clustering. *Biophys. J.* **116**, 127-141. doi:10.1016/j.bpj.2018.11.011
McMillen, D., Kopell, N., Hasty, J. and Collins, J. J. (2002). Synchronizing genetic relaxation oscillators by intercell signaling. *Proc. Natl Acad. Sci. USA* **99**, 679-684. doi:10.1073/pnas.022642299
Mil'shtein, G. N. (1974). Approximate integration of stochastic differential equations. *Teor. Veroyatn. Primen.* **19**, 583-588.
Momiji, H. and Monk, N. A. M. (2009). Oscillatory Notch-pathway activity in a delay model of neuronal differentiation. *Phys. Rev. E* **80**, 021930. doi:10.1103/PhysRevE.80.021930
Monk, N. A. M. (1997). Cell communities and robustness in development. *Bull. Mat. Biol.* **59**, 1183-1189. doi:10.1007/BF02460107
Nichols, J., Silva, J., Roode, M. and Smith, A. (2009). Suppression of Erk signalling promotes ground state pluripotency in the mouse embryo. *Development* **136**, 3215-3222. doi:10.1242/dev.038893
Niwa, H., Toyooka, Y., Shimosato, D., Strumpf, D., Takahashi, K., Yagi, R. and Rossant, J. (2005). Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. *Cell* **123**, 917-929. doi:10.1016/j.cell.2005.08.040
O'Dea, R. D. and King, J. R. (2012). Continuum limits of pattern formation in hexagonal-cell monolayers. *J. Math. Biol.* **64**, 579-610. doi:10.1007/s00285-011-0427-3
Raina, D., Stanoev, A., Bahadori, A., Protzek, M., Koseska, A. and Schröter, C. (2020). Cell-cell communication through FGF4 generates and maintains robust proportions of differentiated cell fates in embryonic stem cells. *bioRxiv*, p. 2020.02.14.949701.
Raspopovic, J., Marcon, L., Russo, L. and Sharpe, J. (2014). Digit patterning is controlled by a Bmp-Sox9-Wnt Turing network modulated by morphogen gradients. *Science* **345**, 566-570. doi:10.1126/science.1252960
Saiz, N., Williams, K. M., Seshan, V. E. and Hadjantonakis, A.-K. (2016). Asynchronous fate decisions by single cells collectively ensure consistent lineage composition in the mouse blastocyst. *Nat. Commun.* **7**, 13463. doi:10.1038/ncomms13463
Saiz, N., Mora-Bitria, L., Rahman, S., George, H., Herder, J., Garcia-Ojalvo, J. and Hadjantonakis, A.-K. (2020). Growth-factor-mediated coupling between lineage size and cell fate choice underlies robustness of mammalian development. *eLife* **9**, e56079. doi:10.7554/eLife.56079
Schröter, C., Rué, P., Mackenzie, J. P. and Martinez Arias, A. (2015). FGF/MAPK signaling sets the switching threshold of a bistable circuit controlling cell fate decisions in embryonic stem cells. *Development* **142**, 4205-4216. doi:10.1242/dev.127530
Simon, C. S., Hadjantonakis, A.-K. and Schröter, C. (2018). Making lineage decisions with biological noise: Lessons from the early mouse embryo. *Wiley Interdiscipl. Rev. Dev. Biol.* **7**, e319. doi:10.1002/wdev.319
Snoussi, E. H. (1998). Necessary conditions for multistationarity and stable periodicity. *J. Biol. Syst.* **6**, 3-9. doi:10.1142/S0218339098000042
Soldatov, R., Kaucka, M., Kastriti, M. E., Petersen, J., Chontorotzea, T., Englmaier, L., Akkuratova, N., Yang, Y., Häring, M., Dyachuk, V. et al. (2019). Spatiotemporal structure of cell fate decisions in murine neural crest. *Science* **364**, eaas9536. doi:10.1126/science.aas9536
Sprinzak, D., Lakhanpal, A., LeBon, L., Santat, L. A., Fontes, M. E., Anderson, G. A., Garcia-Ojalvo, J. and Elowitz, M. B. (2010). Cis-interactions between notch and delta generate mutually exclusive signalling states. *Nature* **465**, 86-90. doi:10.1038/nature08959
Suzuki, N., Furusawa, C. and Kaneko, K. (2011). Oscillatory protein expression dynamics endows stem cells with robust differentiation potential. *PLoS ONE* **6**, e27232. doi:10.1371/journal.pone.0027232
Taga, M. E. and Bassler, B. L. (2003). Chemical communication among bacteria. *Proc. Natl Acad. Sci. USA* **100**, 14549-14554. doi:10.1073/pnas.1934514100
Teomy, E., Kessler, D. A. and Levine, H. (2019). Ordered hexagonal patterns via notch-delta signaling. *arXiv preprint arXiv*, 1902.04917. doi:10.1101/550657
Thomas, R. (1981). On the relation between the logical structure of systems and their ability to generate multiple steady states or sustained oscillations. In *Numerical Methods in The Study Of Critical Phenomena*, Vol. 9 (ed. J. Della Dora, J. Demongeot and B. Lacolle), pp. 180-193. Berlin: Springer.
Turing, A. M. (1952). The chemical basis of morphogenesis. *Philos. Trans. R. Soc. Lond. Ser. B* **237**, 37-72. doi:10.1098/rstb.1952.0012
Ullner, E., Zaikin, A., Volkov, E. I. and García-Ojalvo, J. (2007). Multistability and clustering in a population of synthetic genetic oscillators via phase-repulsive cell-to-cell communication. *Phys. Rev. Lett.* **99**, 148103. doi:10.1103/PhysRevLett.99.148103
Ullner, E., Koseska, A., Kurths, J., Volkov, E., Kantz, H. and García-Ojalvo, J. (2008). Multistability of synthetic genetic networks with repressive cell-to-cell communication. *Phys. Rev. E* **78**, 031904. doi:10.1103/PhysRevE.78.031904
van Kekem, D. L. and Sterk, A. E. (2019). Symmetries in the Lorenz-96 model. *Int. J. Bifurcation Chaos* **29**, 1950008. doi:10.1142/S0218127419500081
Wang, J., Zhang, K., Xu, L. and Wang, E. (2011). Quantifying the Waddington landscape and biological paths for development and differentiation. *Proc. Natl Acad. Sci. USA* **108**, 8257-8262. doi:10.1073/pnas.1017017108
Yamanaka, Y., Lanner, F. and Rossant, J. (2010). FGF signal-dependent segregation of primitive endoderm and epiblast in the mouse blastocyst. *Development* **137**, 715-724. doi:10.1242/dev.043471
You, L., Cox, R. S., III, Weiss, R. and Arnold, F. H. (2004). Programmed population control by cell-cell communication and regulated killing. *Nature* **428**, 868-871. doi:10.1038/nature02491
Youk, H. and Lim, W. A. (2014). Secreting and sensing the same molecule allows cells to achieve versatile social behaviors. *Science* **343**, 6171. doi:10.1126/science.1242782
Zakharova, A., Schneider, I., Kyrychko, Y. N., Blyuss, K. B., Koseska, A., Fiedler, B. and Schöll, E. (2013). Time delay control of symmetry-breaking primary and secondary oscillation death. *EPL (Europhysics Letters)* **104**, 50004. doi:10.1209/0295-5075/104/50004
Zhang, H. T. and Hiiragi, T. (2018). Symmetry breaking in the mammalian embryo. *Annu. Rev. Cell Dev. Biol.* **34**, 405-426. doi:10.1146/annurev-cellbio-100617-062616

OXFORD

human reproduction update

Human Reproduction Update, 2023, **00(0)**, 1–17

https://doi.org/10.1093/humupd/dmad017

Dissecting mammalian reproduction with spatial transcriptomics

Xin Zhang[1,2,†], Qiqi Cao[1,2,†], Shreya Rajachandran[1,2], Edward J. Grow[1,2], Melanie Evans[2], and Haiqi Chen [1,2,*]

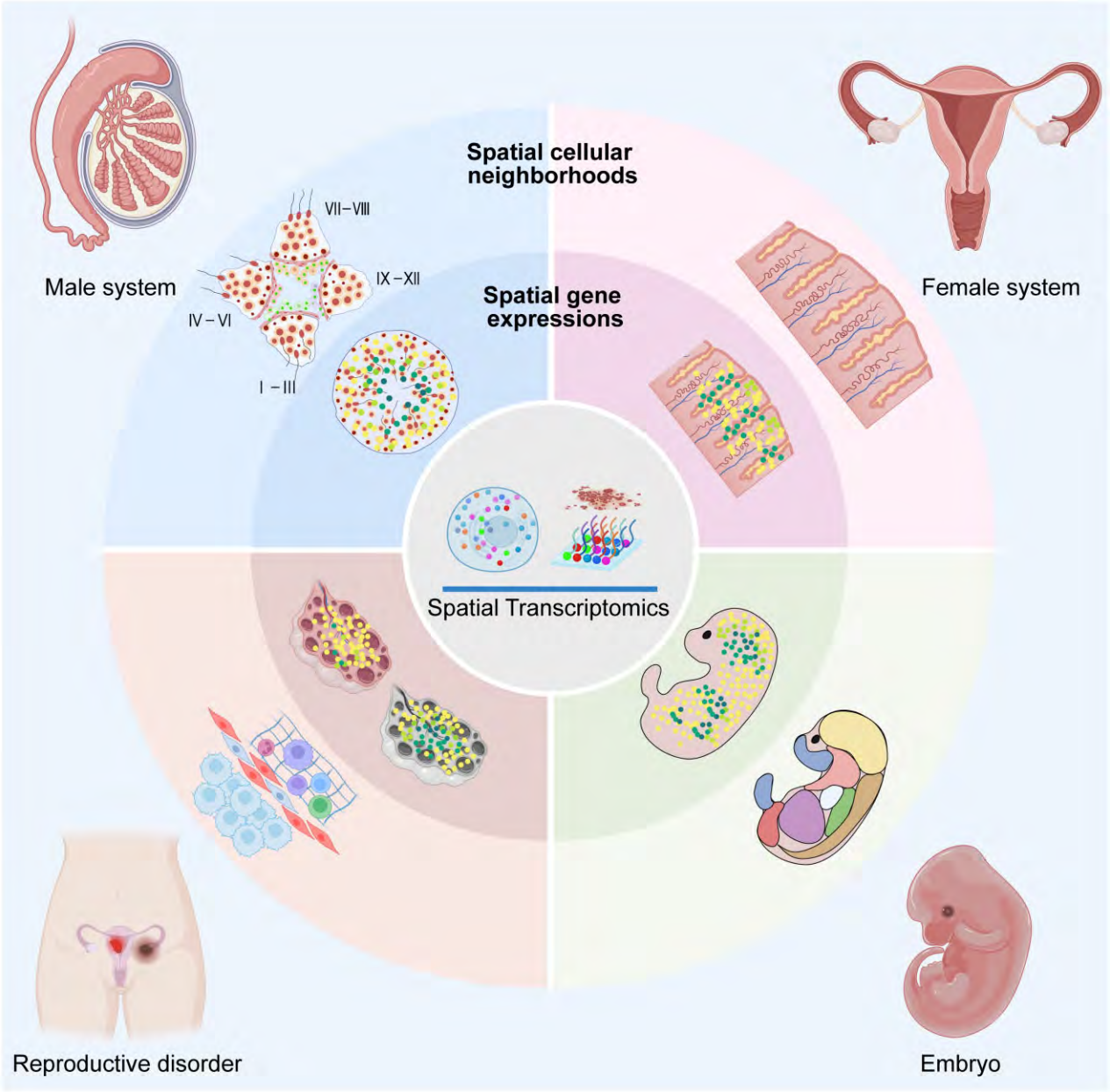
[1]Cecil H. and Ida Green Center for Reproductive Biology Sciences, University of Texas Southwestern Medical Center, Dallas, TX, USA
[2]Department of Obstetrics and Gynecology, University of Texas Southwestern Medical Center, Dallas, TX, USA

*Correspondence address. Cecil H. and Ida Green Center for Reproductive Biology Sciences, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX 75390, USA. E-mail: haiqi.chen@utsouthwestern.edu https://orcid.org/0000-0001-7179-3229
†These authors contributed equally to this work.

TABLE OF CONTENTS

Received: March 19, 2023. **Revised:** May 15, 2023. **Editorial decision:** May 28, 2023.

GRAPHICAL ABSTRACT



Applications of spatial transcriptomics technologies in mammalian reproductive systems in the context of gametogenesis, embryogenesis, and reproductive pathologies.

ABSTRACT

BACKGROUND: Mammalian reproduction requires the fusion of two specialized cells: an oocyte and a sperm. In addition to producing gametes, the reproductive system also provides the environment for the appropriate development of the embryo. Deciphering the reproductive system requires understanding the functions of each cell type and cell–cell interactions. Recent single-cell omics technologies have provided insights into the gene regulatory network in discrete cellular populations of both the male and female reproductive systems. However, these approaches cannot examine how the cellular states of the gametes or embryos are regulated through their interactions with neighboring somatic cells in the native tissue environment owing to tissue disassociations. Emerging spatial omics technologies address this challenge by preserving the spatial context of the cells to be profiled. These technologies hold the potential to revolutionize our understanding of mammalian reproduction.

OBJECTIVE AND RATIONALE: We aim to review the state-of-the-art spatial transcriptomics (ST) technologies with a focus on highlighting the novel biological insights that they have helped to reveal about the mammalian reproductive systems in the context of gametogenesis, embryogenesis, and reproductive pathologies. We also aim to discuss the current challenges of applying ST technologies in reproductive research and provide a sneak peek at what the field of spatial omics can offer for the reproduction community in the years to come.

SEARCH METHODS: The PubMed database was used in the search for peer-reviewed research articles and reviews using combinations of the following terms: 'spatial omics', 'fertility', 'reproduction', 'gametogenesis', 'embryogenesis', 'reproductive cancer', 'spatial

transcriptomics', 'spermatogenesis', 'ovary', 'uterus', 'cervix', 'testis', and other keywords related to the subject area. All relevant publications until April 2023 were critically evaluated and discussed.

OUTCOMES: First, an overview of the ST technologies that have been applied to studying the reproductive systems was provided. The basic design principles and the advantages and limitations of these technologies were discussed and tabulated to serve as a guide for researchers to choose the best-suited technologies for their own research. Second, novel biological insights into mammalian reproduction, especially human reproduction revealed by ST analyses, were comprehensively reviewed. Three major themes were discussed. The first theme focuses on genes with non-random spatial expression patterns with specialized functions in multiple reproductive systems; The second theme centers around functionally interacting cell types which are often found to be spatially clustered in the reproductive tissues; and the thrid theme discusses pathological states in reproductive systems which are often associated with unique cellular microenvironments. Finally, current experimental and computational challenges of applying ST technologies to studying mammalian reproduction were highlighted, and potential solutions to tackle these challenges were provided. Future directions in the development of spatial omics technologies and how they will benefit the field of human reproduction were discussed, including the capture of cellular and tissue dynamics, multi-modal molecular profiling, and spatial characterization of gene perturbations.

WIDER IMPLICATIONS: Like single-cell technologies, spatial omics technologies hold tremendous potential for providing significant and novel insights into mammalian reproduction. Our review summarizes these novel biological insights that ST technologies have provided while shedding light on what is yet to come. Our review provides reproductive biologists and clinicians with a much-needed update on the state of art of ST technologies. It may also facilitate the adoption of cutting-edge spatial technologies in both basic and clinical reproductive research.

Keywords: spatial transcriptomics / reproduction / gametogenesis / pregnancy / embryogenesis / cancer

Introduction

Reproduction ensures the transmission of genetic and epigenetic information to the next generation and the continuity of species. The maintenance of the reproductive systems, the generation of gametes, and embryonic development are some of the central focuses of reproductive biology. A deep understanding of mammalian reproduction could facilitate the diagnosis and treatment of infertility, cancer, and other reproductive pathologies, as well as the development of contraceptives.

Mammalian reproduction is regulated by numerous biological pathways and involves many cell types. For example, the development of gametes is a highly regulated process, which includes, but is not limited to, chromatin remodeling, epigenetic reprogramming, cell cycle regulation, meiosis, and cell migration (Marston and Amon, 2004; Richardson and Lehmann, 2010; Yosefzon *et al.*, 2017; Cabot and Cabot, 2018; Larose *et al.*, 2019; Fang *et al.*, 2022). It also requires cellular and molecular interactions between developing gametes and surrounding somatic cell types (Wassarman, 2002; Mruk and Cheng, 2004; Gershon *et al.*, 2008; Hofmann and McBeath, 2022). Deciphering such complexity requires technologies capable of characterizing molecular and cellular processes at scale. While recent single-cell technologies offer a high throughput solution (Vitak *et al.*, 2017; Wang *et al.*, 2018b, 2019, 2021; Green *et al.*, 2018; Guo *et al.*, 2018; Argelaguet *et al.*, 2019; Ferrero *et al.*, 2019; McGinnis *et al.*, 2019; Zhao *et al.*, 2020; Cheung *et al.*, 2021; Li *et al.*, 2018, 2021; Mittnenzweig *et al.*, 2021; Yan *et al.*, 2021; Zhang *et al.*, 2021; Garcia-Alonso *et al.*, 2022), they require tissue dissociation, which results in the loss of spatial context and significant cellular information such as cell–cell and cell–extracellular matrix interactions.

Spatial transcriptomics (ST) technologies have emerged as tools that can not only provide the information on the abundance of mRNA molecules in the cells but also capture their spatial locations within the tissue (Marx, 2021; Rao *et al.*, 2021; Tian *et al.*, 2023). These technologies range from laser capture microdissection (LCM), *in situ* hybridization (ISH) and *in situ* sequencing (ISS) to solid phase-based capturing technologies (Marx, 2021; Rao *et al.*, 2021; Tian *et al.*, 2023). Together, they play a crucial role in exploring the spatial distribution of RNA, the spatial location of cell populations, and cell–cell interactions. In this review, we introduce major ST technologies that have been applied to mammalian reproductive systems, discuss in detail the biological insights that have been revealed by studies using ST, and offer an outlook for the future of ST technologies and how they can further benefit the field of reproductive biology in the near future.

Overview of ST technologies

ST technologies can be primarily categorized into two classes based on their design principles. The first class relies on the imaging of pre-determined mRNA targets. These targeted approaches include ISH-based methods and ISS-based methods. The second class of ST technologies includes unbiased approaches that build on spatial isolation/capture of RNA molecules followed by next-generation sequencing (NGS). For readers who are interested in the current landscape of ST technologies, we have compiled a list of representative ST technologies in Table 1. For readers who are interested in learning about computational approaches to analyze ST data, we recommend these excellent reviews (Dries *et al.*, 2021; Longo *et al.*, 2021; Zeng *et al.*, 2022). In the next two sections, we focus on ST technologies that have been applied to the reproductive systems.

Targeted approaches

The ISH method visualizes the target molecules in tissue sections by using imaging probes to sequentially hybridize the targets. Representative techniques include single-molecule fluorescence *in situ* hybridization (smFISH) (Femino *et al.*, 1998), sequential single-molecule FISH (seqFISH) (Lubeck *et al.*, 2014), and multiplexed error-robust fluorescence *in situ* hybridization (MERFISH) (Chen *et al.*, 2015) (Fig. 1A).

In smFISH, multiple fluorescent probes target specifically complementary sites of an individual mRNA to generate high-intensity signals for visualization. This methodology accurately quantifies and visualizes the expression of RNA molecules within the cells. However, owing to a limited number of available fluorescent channels, multiple RNA targets cannot be measured simultaneously. To overcome this limitation, a technology that combines the smFISH technique with combinatorial labeling was developed. This technology, termed seqFISH (Lubeck *et al.*, 2014), decodes mRNAs by sequential rounds of hybridization, imaging, and probe stripping. Specifically, during each round of hybridization, each transcript labeled by FISH probes with a single type of

Table 1. A list of representative spatial transcriptomics technologies.

TARGETED APPROACHES

***In situ* hybridization (ISH)**

Year	Methods	Authors	Features	Number of targets	Spatial resolution	Limitations	Estimated cost*
1998	smFISH (Femino *et al.*, 1998)	Andrea M. Femino	Multiple oligonucleotide probes hybridize with the same transcript.	4–5 targets	Subcellular resolution	The spectral overlap limits the simultaneous detection of multiple transcripts.	~$120/sample; Microscope needed.
2014	seqFISH (Lubeck *et al.*, 2014)	Eric Lubeck *et al.*	The mRNA molecules are barcoded by sequential rounds of hybridization, imaging, and probe stripping.	Dozens of targets within individual cell	Subcellular resolution	Complex experimental setup: data analysis is challenging.	Custom microfluidics and flow cell: ~$5000; consumable: ~$500/sample; microscope needed.
2015	MERFISH (Chen *et al.*, 2015)	Kok Hao Chen *et al.*	Using the binary words of modified Hamming code to encode the RNA molecules.	~100–1000 genes within individual cell	Subcellular resolution	Costly instrument	Commercial Vizgen MERSCOPE Instrument: ~$300 000; consumable: ~$600/sample.
2019	MERFISH (Xia *et al.*, 2019)	Chenglong Xia *et al.*	Enhanced throughput of the original MERFISH.	~10 000 genes within individual cell	Subcellular resolution	Complex experimental setup; data analysis is challenging.	Custom microfluidics and flow cell: ~$5000; consumable: ~$500/sample; microscope needed.
2019	seqFISH+ (Eng *et al.*, 2019)	Chee-Huat Linus Eng *et al.*	Using 60 'pseudocolor' channels to dilute mRNA molecules.	~10 000 within individual cell	Sub-diffraction limit resolution	Complex experimental setup; data analysis is challenging.	Custom microfluidics and flow cell: ~$5000; consumable: ~$600/sample. microscope needed.
2019	GeoMix (Nanostring, 2019)	Nanostring	Based on probes linked to indexing oligo barcodes via a photocleavable linker.	A few hundred genes	Single-cell resolution	Costly instrument	Commercial GeoMix instrument: ~$290 000; consumable: ~$500/sample. Additional NGS required.

***In situ* sequencing (ISS)**

Year	Methods	Authors	Features	Number of targets	Spatial resolution	Limitations	
2013	ISS (Ke *et al.*, 2013)	Rongqin Ke *et al.*	Based on padlock probes, RCA, and sequencing-by-ligation chemistry.	256 targets within a single cell	Subcellular resolution	Laborious and low efficiency	Commercial microfluidics and flow cell: ~$5000; consumable: ~$500/sample. Microscope needed.
2014	FISSEQ (Lee *et al.*, 2014)	Je Hyuk Lee *et al.*	RNA is reverse transcribed with tagged random hexamers; RCA; SOLiD sequencing	Entire transcriptome	Subcellular resolution	Low sequencing depth; time-consuming protocol; SOLiD sequencing reagents have been discontinued	Commercial microfluidics and flow cell: ~$5000; consumable: ~$500/sample. Microscope needed.
2018	BaristaSeq (Chen *et al.*, 2018)	Xiaoyin Chen *et al.*	Based on padlock probes and RCA; using *in situ* barcode sequencing compatible with Illumina sequencing chemistry	Entire transcriptome	Subcellular resolution	Low sequencing depth	Custom microfluidics and flow cell: ~$5000; consumable: ~$600/sample. Microscope needed.
2018	STARmap (Wang *et al.*, 2018a)	Xiao Wang *et al.*	Based on SNAIL probes, RCA, and fluorescent *in situ* sequencing	Detecting ~1000 transcripts in a cell	Subcellular resolution	Complex experimental setup; data analysis is challenging.	Custom microfluidics and flow cell: ~$5000; consumable: ~$600/sample. Microscope needed.
2021	ExSeq (Alon *et al.*, 2021)	Shahar Alon *et al.*	Based on expansion microscopy	3039 genes	Nanoscale resolution	Complex experimental setup; data analysis is challenging.	Custom microfluidics and flow cell: ~$5000; consumable: ~$600/sample. Microscope needed.

(continued)

Table 1. (continued)

In situ sequencing (ISS)							
Year	**Methods**	**Authors**	**Features**	**Number of targets**	**Spatial resolution**	**Limitations**	
2023	Xenium (10× Genomics, 2023)	10× Genomics	Padlock probe-based RCA and fluorescent *in situ* sequencing.	300 genes	Subcellular resolution	Costly instrument	Commercial Xenium instrument: ~$350 000; consumable: ~$1000/sample.
UNBIASED APPROACHES							
Year	**Methods**	**First author**	**Features**	**Number of targets**	**Spatial resolution**	**Limitations**	**Estimated cost**
1996	LCM (Emmert-Buck *et al.*, 1996)	Michael R. Emmert-Buck *et al.*	Using laser to accurately obtain target cell subgroups or single cells	Entire transcriptome	Single-cell resolution	Low throughput; time-consuming protocol	Commercial LCM instrument: ~$150 000; consumable: ~$100/sample; additional NGS required.
2016	Spatial Transcriptomics (Ståhl *et al.*, 2016)	Patrik L. Ståhl *et al.*	Capturing RNA using spatially indexed poly(dT) oligo arrays.	Entire transcriptome	100 μm	Low sensitivity for low abundant transcripts; low spatial resolution	~$650/sample; additional NGS required.
2017	GEO-seq (Chen *et al.*, 2017)	Jun Chen *et al.*	Combining LCM with single-cell RNA-seq	Entire transcriptome	Single-cell resolution	Low throughput; time-consuming protocol	Commercial LCM instrument: ~$150 000; consumable: ~$1000/sample; additional NGS required.
2019	Visium (10× Genomics, 2019)	10× Genomics	Based on the spatial transcriptomics technology	Entire transcriptome	~55 μm	Low spatial resolution	~$3000/sample; additional NGS required.
2019	Slide-seq (Rodriques *et al.*, 2019)	Samuel G. Rodriques *et al.*	Capturing mRNA using spatially indexed 10-μm-diameter beads	Entire transcriptome	10 μm	Low sensitivity for low abundant transcripts	~$1000/sample; additional NGS required.
2020	DBiT-seq (Liu *et al.*, 2020)	Yang Liu *et al.*	Using microfluidic chip and DNA barcodes to spatially index molecules in the tissue	Entire transcriptome	10 μm	Requiring a custom microfluidic device	Custom microfluidics: ~$7000; consumable: ~$1400/sample; additional NGS required.
2021	Slide-seqV2 (Stickels *et al.*, 2021)	Robert R. Stickels *et al.*	Enhanced capture efficiency of the original Slide-seq.	Entire transcriptome	10 μm	Cannot be applied to FFPE tissue blocks	Commercial Curio Seeker product line: ~$1300/sample; additional NGS required.
2021	Seq-Scope (Cho *et al.*, 2021)	Chun-Seok Cho *et al.*	Repurposing of an illumina sequencing flow cell.	Entire transcriptome	~0.5–0.8 μm	Exposing the cluster surface of the flow cell is challenging.	~$1500/sample; additional NGS required.
2021	Stereo-seq (Chen *et al.*, 2022)	Ao Chen *et al.*	Using DNA nanoball-patterned chips to capture RNA molecules	Entire transcriptome	~220 nm	Low sensitivity for low abundant transcripts.	Commercial BGI Platform: ~$1500/sample; additional NGS required.
2021	Sci-Space (Srivatsan *et al.*, 2021)	Sanjay R. Srivatsan *et al.*	Using unmodified DNA oligos to label nuclei	Entire transcriptome	Single-cell resolution	Cannot capture cytoplasmic RNA; dropout during nucleus collection.	~$5000/sample; additional NGS required.
2022	Pixel-seq (Fu *et al.*, 2022)	Xiaona-n Fu *et al.*	Polony gel stamping enables scalable replication of DNA cluster arrays	Entire transcriptome	~1 μm	Require specialized expertise in gel fabrication and transfer.	~$105/sample; additional NGS required.

* Please note that this is only a rough estimate. Many of the technologies have not yet been commercialized and, therefore, the accurate information about their costs is not fully available. Furthermore, the actual cost of an experiment can vary significantly depending on experimental designs (e.g. tissue type and size, number of cells in a tissue slice, number of genes profiled, and pre-designed versus custom gene panels). Additional costs of obtaining a microscope for imaging-based technologies and next-generation sequencing for array-based technologies are not included in our calculation.

BaristaSeq: barcode *in situ* targeted sequencing; DBiT-seq: deterministic barcoding in tissue for spatial omics sequencing; FFPE: formalin fixed paraffin embedded; FISSEQ: fluorescent *in situ* sequencing; GEO-seq: geographical position sequencing; ISH: *in situ* hybridization; ISS: *in situ* sequencing; LCM: laser capture microdissection; MERFISH: multiplexed error-robust fluorescence *in situ* hybridization; NGS: next-generation sequencing; RCA: rolling circle amplification; smFISH: single-molecule fluorescence *in situ* hybridization; STARmap: spatially resolved transcript amplicon readout mapping; Stereo-seq: spatial enhanced resolution omics-sequencing.

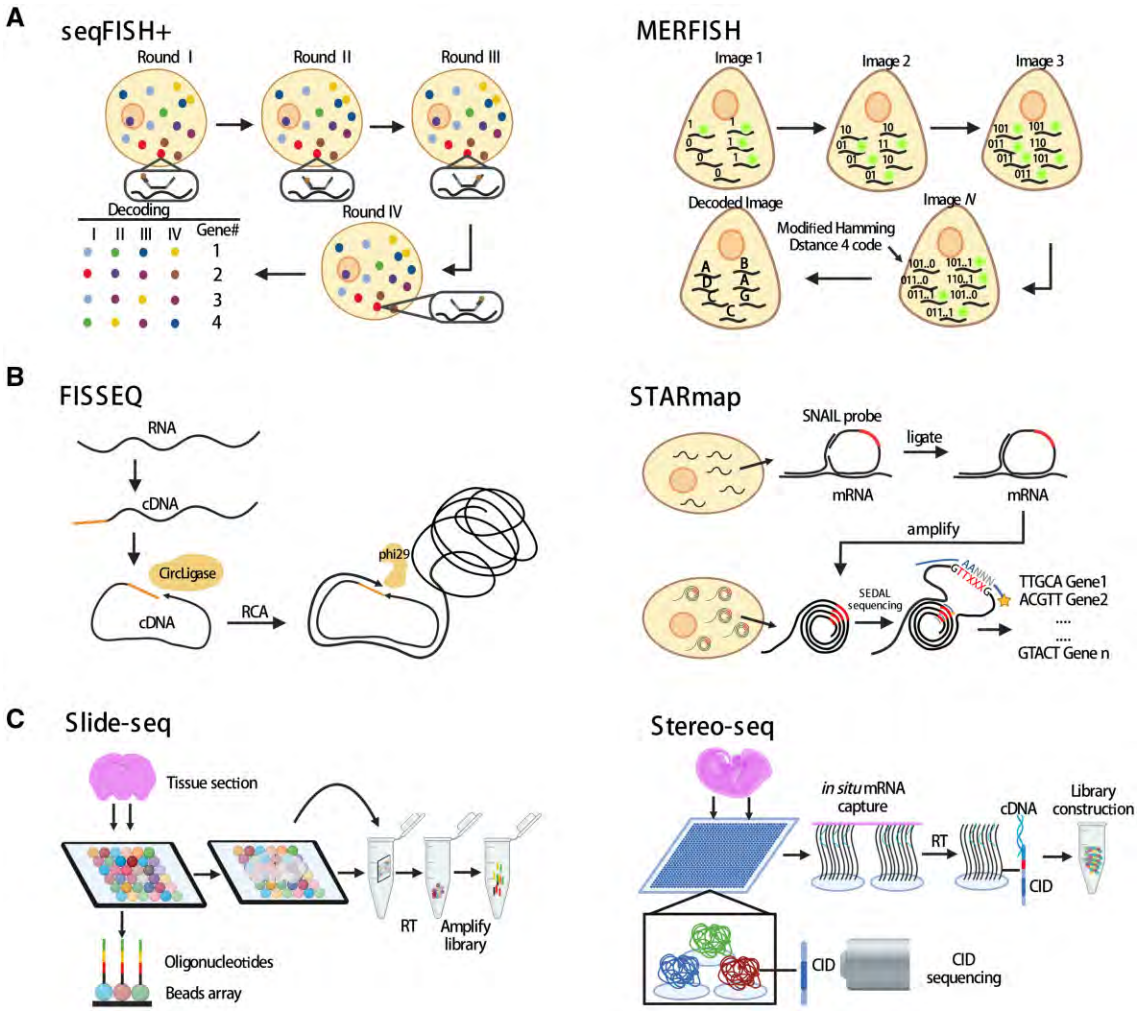


Figure 1. Representative spatial transcriptomics technologies. (**A**) Schematics of ISH-based ST technologies seqFISH+ and MERFISH. (**B**) Schematics of ISS-based ST technologies FISSEQ and STARmap. (**C**) Schematics of solid phase capture-based ST technologies Slide-seq and Stereo-seq. CID: coordinate identity; FISSEQ: fluorescent *in situ* sequencing; ISH: *in situ* hybridization; ISS: *in situ* sequencing; MERFISH: multiplexed error-robust fluorescence *in situ* hybridization; RCA: rolling circle amplification; RT: reverse transcription; seqFISH: sequential single-molecule fluorescence *in situ* hybridization; ST: spatial transcriptomics; STARmap: spatially resolved transcript amplicon readout mapping.

fluorophore is visualized, and then, the FISH probes are removed by treatment with DNase. In a subsequent round, the same transcript is hybridized with the same FISH probes but now labeled with a different dye. Thus, four dyes and eight rounds of hybridization can theoretically cover the entire transcriptome of the mouse or the human (4^8 = 65 536). However, global profiling of hundreds or thousands of mRNA is hindered by optical crowding. To overcome this challenge, seqFISH+ was developed (Eng *et al.*, 2019). seqFISH+ expands the barcode base palette from 4 to 5 colors to 60 pseudocolors per image cycle, resulting in the detection of ~10 000 genes per cell by repeating the cycle of pseudocolor imaging only four times (Fig. 1A, left).

Another ISH-based method is MERFISH (Chen *et al.*, 2015), which uses the modified Hamming code to encode the RNA molecules (Fig. 1A, right). Using a two-step labeling scheme, MERFISH dramatically decreases the probe hybridization time and reduces the error rate of barcode identification. Furthermore, the detection efficiency of RNA molecules by MERFISH can be increased through a combination with expansion microscopy, which effectively increases the distances between neighboring RNA molecules and helps substantially increase the RNA density measurable by MERFISH (Wang *et al.*, 2018c). Thus, the gene throughput of MERFISH has been increased from the original ~1000 transcripts to ~10 000 transcripts in individual cells (Xia *et al.*, 2019).

Besides ISH-based approaches, ISS-based methods are also frequently used to yield spatial transcriptome information (Fig. 1B). In 2013, an ISS technology that combines padlock probing, rolling circle amplification (RCA), and sequencing-by-ligation chemistry was used to sequence RNAs *in situ* for the first time (Ke *et al.*, 2013). In ISS, the padlock probes that carry transcript-specific barcodes hybridize to the RNA targets and are circularized via ligation of the 5′ and 3′ ends of the probes. Then, the circularized padlock probes are amplified by RCA and the probe barcodes are sequenced *in situ* using fluorescent oligos. Similarly, fluorescent *in situ* sequencing (FISSEQ) (Lee *et al.*, 2014), another ISS-based method, first generates cDNA from RNA using reverse transcription with tagged random hexamers. Then, the cDNA fragments are circularized by circligase and amplicons are formed after RCA. This procedure ensures that RNA molecules are profiled in a non-targeted manner. Spatially resolved transcript amplicon readout mapping (STARmap) is another technology based on ISS (Wang *et al.*, 2018a). It uses a pair of primer and padlock probes (called SNAIL probes) to ensure target-specific signal amplification. STARmap bypasses the step of reverse transcription to increase the efficiency of amplicon generation. In

addition, an improved ISS chemistry called SEDAL was devised specifically for STARmap. SEDAL eliminates error accumulation as sequencing proceeds and exhibits minimal background. With these improvements, STARmap reads >1000 genes per cell in a mouse brain.

Unbiased approaches

In the early days of ST, physical microdissection techniques were used to isolate molecules at specific spatial locations, such as those in LCM (Emmert-Buck *et al.*, 1996), Tomo-Seq (Junker *et al.*, 2014), and spatial transcriptomics by reoriented projections and sequencing (STRP-seq) (Schede *et al.*, 2021). LCM can efficiently and accurately obtain target cell subgroups or single cells within tissues (Emmert-Buck *et al.*, 1996), and is often used to analyze the transcriptome of tissue regions in combination with other sequencing methods. For example, geographical position sequencing (Geo-seq) captures cell heterogeneities and spatial variance simultaneously by combining LCM with single-cell RNA sequencing (scRNA-seq) technology (Chen *et al.*, 2017). Similarly, LCM-seq combines LCM with poly A-based Smart-seq2 RNA sequencing (Nichterwitz *et al.*, 2016, 2018).

Although LCM combined with scRNA-seq can provide ST information at cellular resolution, its low throughput makes it difficult to scale to large tissue areas. To overcome this limitation, a solid-phase capture technology named Spatial Transcriptomics was developed in 2016 (Ståhl *et al.*, 2016). Its innovation lies in the introduction of spatial barcodes before sequencing library preparation (Jemt *et al.*, 2016). Specifically, the mRNA molecules of tissue sections are captured with spatially barcoded oligo(dT) primers anchored on glass slides. The subsequent reverse transcription enables the resulting cDNAs to be coupled to the arrayed oligo(dT) primers on the glass slides. By using NGS, the mRNA identity and the coupled spatial barcode can be identified. Each gene can then be unbiasedly mapped to the tissue sections based on the unique spatial barcode. Thus, Spatial Transcriptomics quantifies the gene expression and visualizes the distribution of mRNAs within tissue sections. The spatial resolution of Spatail Transcriptomics is 100 μm with a center-to-center distance of 200 μm between two adjacent 'spots'. Building upon the ST technology, the commercially available 10× Genomics Visium technology (10× Genomics, 2019) increases the cellular resolution to 55 μm with a 100-μm center-to-center distance between spots and a sensitivity of >10 000 transcripts per spot.

Slide-seq is another ST technology that combines spatial barcoding with solid-phase RNA capture (Rodriques *et al.*, 2019) (Fig. 1C). The Slide-seq array is generated by packing DNA-barcoded beads onto a glass surface. The position of each bead is determined by ISS. Using the spatially indexed arrays, Slide-seq captures mRNA molecules from fresh frozen tissue sections and enables unbiased mapping of the mRNA molecules back to the original locations. Compared with Spatial Transcriptomics and 10× Genomics Visium, Slide-seq provides a higher spatial resolution (10-μm bead diameter) and lower experimental cost. Slide-seqV2 (Stickels *et al.*, 2021), an improved version of Slide-seq, has an RNA capture efficiency of ∼10-fold greater than the original Slide-seq, resulting from the improved workflow of library generation, bead synthesis, and array indexing.

Recently, spatial enhanced resolution omics-sequencing (Stereo-seq) has achieved nanoscale resolution (220-nm spot diameter with ∼500-nm center-to-center distance) by using spatially barcoded DNA nanoball (DNB) chips (Fig. 1C). The spatial location of each DNB can be read out by sequencing. The high sensitivity and resolution allow Stereo-seq to be used to visualize nuclear versus cytoplasmic transcripts. Other examples of ST technologies based on solid-phase capture and spatial barcoding include deterministic barcoding in tissue for spatial omics sequencing (DBiT-seq) (Liu *et al.*, 2020), sci-Space (Srivatsan *et al.*, 2021), and Pixel-seq (Fu *et al.*, 2022).

Biological insights of mammalian reproduction revealed by ST technologies

ST technologies have been widely applied for the visualization of molecular spatial structures within various tissues (Garcia-Alonso *et al.*, 2022; Guilliams *et al.*, 2022; Hwang *et al.*, 2022; Kuppe *et al.*, 2022; Ratz *et al.*, 2022). By capturing the spatial context of RNA molecules, ST complements scRNA-seq for biological discoveries. In the following sections, we systematically review the novel biological insights of mammalian reproduction revealed by ST technologies, including the identification of genes with non-random spatial expression patterns and specialized functions; the characterization of cellular neighborhoods under reproductive homeostasis; and the examination of tissue microenvironment under pathological conditions.

Spatially patterned gene expression and functions in reproductive systems

Genes with non-random spatial distributions within a tissue often play important roles in cellular functions. To this end, ST technologies offer a unique opportunity to identify these genes at scale.

For example, seminiferous tubules are the functional units of spermatogenesis in mammalian testes (Hess and Renato de Franca, 2008) (Fig. 2A). In a recent study, Slide-seqV2 was used to capture the spatial distribution of testicular genes in the mouse and human testis at a high throughput (Chen *et al.*, 2021b). Computational analysis of the Slide-seqV2 data systematically revealed genes with non-random spatial distribution in seminiferous tubules such as genes enriched at the periphery of a tubule versus genes enriched near the center of a tubule. The analysis also identified genes whose expression is restricted to a subset of seminiferous tubules. Among these genes, *Habp4* (hyaluronan binding protein 4) was discovered as a potential novel regulator of the chromatin remodeling process during male germ cell development. Furthermore, by comparing the gene expression profiles of Leydig cells (the testosterone-producing somatic cells in the interstitial space), the authors showed that Leydig cells that are spatially adjacent to a subset of seminiferous tubules express a high level of *1700017N19Rik*. These *1700017N19Rik*-expressing Leydig cells also express the stem Leydig cell marker *Nr2f2* (nuclear receptor subfamily 2 group F member 2), indicating that *1700017N19Rik* may be involved in the regulation of stem Leydig cell functions. A similar analysis was performed on testicular macrophages, which identified two spatially distinct macrophage subpopulations. One population localizes in the interstitial space and the other is enriched in the peritubular space. These two populations can be distinguished by the expression of *H2-Ab1* (histocompatibility 2, class II antigen A, beta 1) and *Il1b* genes that are exclusively expressed in peritubular macrophages. Besides Leydig cells and macrophages, it is likely that other testicular somatic cells, such as Sertoli cells and myoid cells, also exhibit spatially dependent gene expression patterns. However, the spindle shape of the myoid cells as well as the spatial proximity of Sertoli cell cytoplasm and that of germ cells make it challenging to capture the myoid cell-specific or Sertoli cell-specific

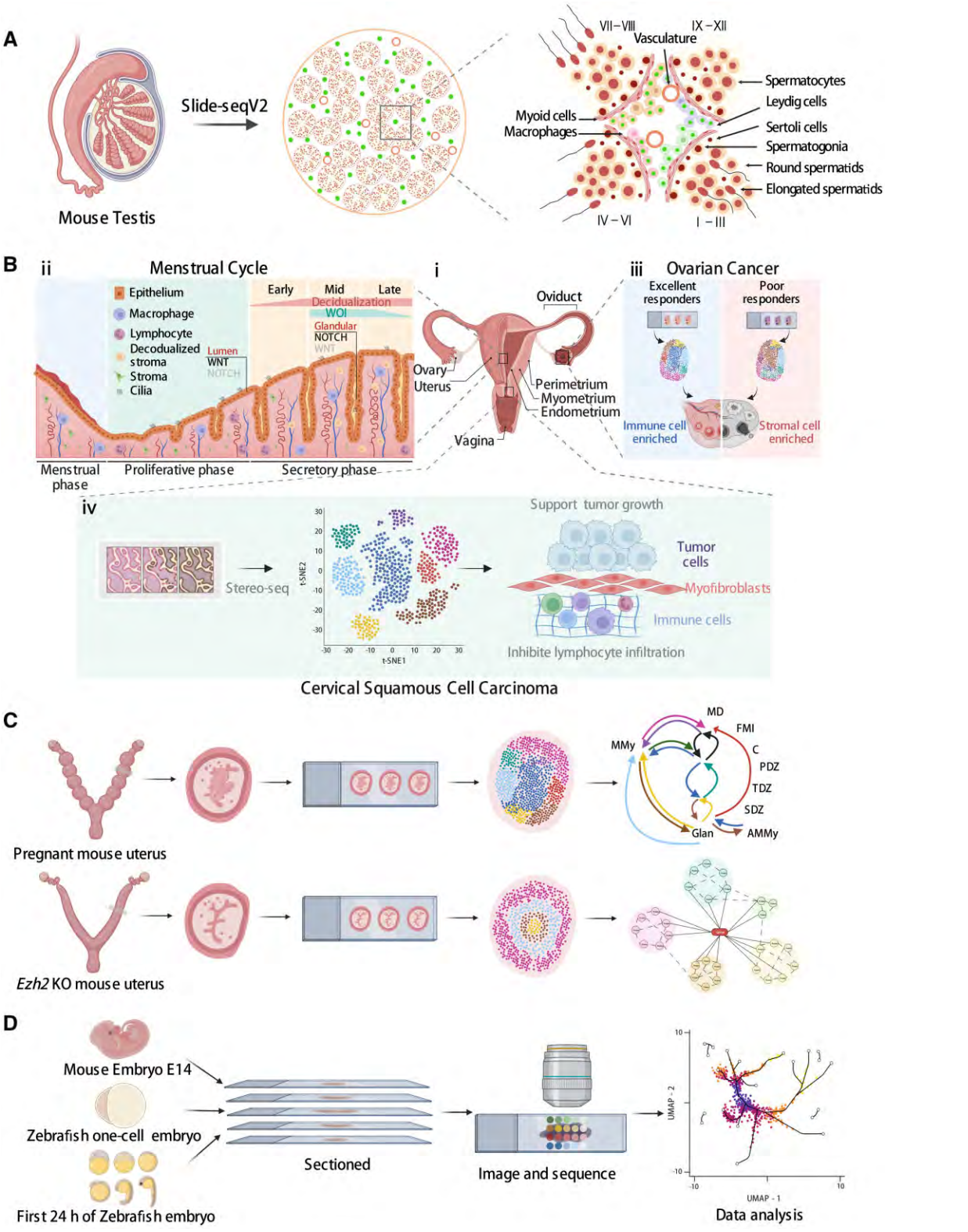


Figure 2. Applications of spatial transcriptomics technologies in reproductive systems. (**A**) The establishment of an unbiased spatial transcriptome atlas of mammalian spermatogenesis using Slide-seqV2. (**B**) (i) An overview of the human female reproductive system. (ii) The cellular structure and molecular signaling of the human endometrium throughout the menstrual cycle. (iii) Spatial characterization of high-grade serous ovarian carcinoma tumor tissue from poor and excellent responders to neoadjuvant chemotherapy. (iv) Stereo-seq identifies cancer-associated myofibroblasts, which may play a supporting role in tumor growth and metastasis by inhibiting lymphocyte infiltration and remodeling tumor extracellular matrix in cervical squamous cell carcinoma. (**C**) Upper panel: spatial cellular neighborhoods of the mouse uterus at the embryo implantation site. AMMy: anti-mesometrial myometrium; E: embryo; FMI: fetal–maternal interface; Glan: uterine glands; MD: mesometrial decidua; MMy: mesometrial myometrium; PDZ: primary decidual zone; SDZ: secondary decidual zone; TDZ: transition decidual zone. Lower panel: the spatial gene expression profile of the enhancer of zeste homolog 2 (*Ezh2*) knockout (KO) mouse uterus. (**D**) Dissecting the developmental processes of mouse and zebrafish embryos using spatial transcriptomics technologies.

transcriptome using Slide-seqV2. This is because one spot on the Slide-seqV2 array may capture mRNA transcripts from two to three adjacent cells. Thus, single-cell-level ST approaches may be better at resolving the spatial transcriptome in cells with irregular shapes or small sizes. For example, seqFISH, a single-cell ST approach, was employed to probe the spatial expression patterns of marker genes of human spermatogonial sub-states identified by scRNA-seq (Guo *et al.*, 2018). This analysis showed that *PIWIL4* (piwi-like RNA-mediated gene silencing 4) and *ETV5* (ETS variant transcription factor 5)/*L1TD1* (LINE1-type transposase domain containing 1) are enriched in spatially distinct spermatogonium (SPG) subpopulations.

The human endometrium is another example in which ST technologies have been employed to reveal spatial gene expression pattens (Fig. 2B, ii). Studying human endometrial homeostasis and pathology has been challenging owing to a lack of model systems (Maurya *et al.*, 2021). One study combined scRNA-seq with ISH to characterize the human endometrium across the menstrual cycle (Wang *et al.*, 2020). Besides the canonical cell types such as stromal fibroblast, endothelium, macrophage, and lymphocyte, scRNA-seq analysis also identified an epithelium-associated cell type the authors called 'ciliated epithelium'. Four genes were found to be highly discriminatory for these ciliated cells (*C11orf88*, *C20orf85*, *FAM183A* (family with sequence similarity 183 member A), and *CDHR3* (cadherin-related family member 3)). ISH targeting of these four genes revealed their consistent co-expression with FOXJ1 protein (forkhead box J1, a master regulator for motile cilia with epithelial lineage identity) in both glandular and luminal epithelia on Days 17 and 25 of the menstrual cycle. In another study, Garcia-Alonso *et al.* (2021) examined the human endometrial epithelium using 10× Genomics Visium. They spatially resolved five cell clusters corresponding to cells in the luminal, functional, and basal layers. Gene signatures of WNT and NOTCH signaling pathways were found to be present in distinct endometrial locations. For example, genes encoding the WNT pathway components, *FOXJ1* and *LGR5* (leucine-rich repeat-containing G protein-coupled receptor 5), are enriched at the luminal surface while *NOTCH2* (notch receptor 2) is mainly expressed in glands in the functional layer. Furthermore, the authors found that *NOTCH2* expression increases in glands moving away from the lumen while *WNT7A* (wnt family member 7A) expression is higher in the luminal epithelium compared with glands. By contrast, the noncanonical WNT gene *WNT5A* is enriched in stromal cells surrounding the glands. These findings suggest an almost mutually exclusive spatial expression pattern between the canonical and noncanonical WNT pathways in the glandular microenvironment.

Besides the human, ST technologies have also been applied to study the mouse uterus in a uterine *Ezh2* (enhancer of zeste homolog 2) knockout (KO) model (Mesa *et al.*, 2021) (Fig. 2C, lower panel). EZH2 is an epigenetic modifier that methylates histone lysine residue 27 (Trevino *et al.*, 2015). Conditional deletion of *Ezh2* in the uterus results in an increased proliferation of luminal and glandular epithelial cells and affects the estrogen signaling pathway (Fang *et al.*, 2019; Nanjappa *et al.*, 2019). Spatial analysis of the uterine *Ezh2* KO model versus the wild type (WT) using 10× Genomics Visium allowed a specific selection of epithelial cells for downstream analyses. Differential expression analysis identified up-regulated (*Asb4* (ankyrin repeat and SOCS box-containing 4), *Cxcl14* (chemokine (C-X-C motif) ligand 14), *Dio2* (deiodinase, iodothyronine, type II), and *Igfbp5* (insulin-like growth factor-binding protein 5)) and down-regulated (*Sult1d1* (sulfotransferase family 1D, member 1), *Mt3* (metallothionein 3), and *Lcn2* (lipocalin 2)) genes in *Ezh2* KO versus WT uterine epithelium.

Finally, multiple ST technologies have been used to spatially profile gene expression during embryogenesis (Fig. 2D). In one study, spatial analysis of mouse E14.0 embryos using sci-Space revealed spatially patterned, cell-type-specific gene expression across the embryo (Srivatsan *et al.*, 2021). Follow-up analyses distinguished genes whose spatial pattern of expression is contributed by multiple cell subtypes from genes whose spatial pattern of expression is contributed by the presence of a single spatially restricted, unannotated cell subtype. For example, the spatial expression pattern of *Hox* genes, a class of homeotic transcription factors that specify the body plan, could not be explained solely by spatial restriction of a single-cell subtype. The spatial expression of *Cyp26b1*, a gene encoding a member of the cytochrome P450 superfamily, is restricted to the brainstem with expression observed in multiple neuronal subclusters. In another study, LCM-based Tomo-seq was applied to zebrafish embryos (Holler *et al.*, 2021). The authors sorted genes based on their spatial expression patterns along the animal-to-vegetal axis of the embryo. As a result, three major groups of spatially patterned genes were identified. One group localizes to the animal side of the embryo, one group of genes is equally distributed across all sections, and a third group of genes is spatially confined to the most vegetal part of the yolk sac. By combining Tomo-seq data of *Xenopus laevis* and *Xenopus tropicalis* embryos with that of the zebrafish embryos, the authors identified nine genes, such as *dazl* (deleted in azoospermia like) and *camk2g1* (calcium/calmodulin-dependent protein kinase II gamma 1), that localize vegetally in all three species, suggesting their conserved function in germ cell development or dorsoventral axis development.

Cellular neighborhoods and their functional implications in reproductive physiology

Another key feature of ST is the ability to identify spatial clustering of interacting cell populations (i.e. cell neighborhoods) within the tissue context.

In the testis, developing gametes are regulated by the tissue microenvironment consisting of various somatic cell types (Sertoli cells, Leydig cells, myoid cells, endothelial cells, macrophages, etc.) (Phillips *et al.*, 2010; Wu *et al.*, 2020) (Fig. 2A). Therefore, understanding the interplay between the developing gametes and the somatic cells is essential to the understanding of spermatogenesis. Using the Slide-seqV2 data, one study calculated the cellular compositions of the tissue microenvironment surrounding SPG (the stem cell-containing developing gametes) by identifying their cellular neighborhoods (Chen *et al.*, 2021b). It was found that in both the mouse and human testis, undifferentiated and differentiating SPG self-aggregate while also spatially segregating from each other. Furthermore, no difference in the spatial compositions of the microenvironment surrounding the undifferentiated versus differentiating mouse SPG was found, which was further validated by an independent ISS experiment targeting 22 testicular marker genes. Of interest, in contrast to the mouse, significant differences in the spatial cellular compositions of the microenvironment surrounding the human undifferentiated versus differentiating SPG were identified. For example, a differential enrichment of endothelial cells in the microenvironment surrounding the human undifferentiated versus differentiating SPG was noted. Together, this study revealed differences in the spatial structure of the spermatogonial microenvironment between the mouse and the human, indicating

differential regulatory mechanisms governing the early stage of spermatogenesis between the two species.

During pregnancy, the relationship between the placenta and the decidua of the uterus is essential to nurture and protect the fetus. To this end, one study focused on the development of trophoblast cells at the human maternal–fetal interface by integrating single-nucleus RNA sequencing (snRNA-seq) with 10× Genomics Visium (Arutyunyan et al., 2023). Different trophoblast subsets were annotated and spatially mapped in the tissue. These trophoblast cells were further grouped into five predefined microenvironments in the tissue based on histological features, and distinct trophoblast subsets were found in different microenvironments. Furthermore, by ordering trophoblast states based on their proximity in both the gene expression and physical space, the most likely trajectory for the emergence and differentiation of invasive extravillous trophoblast cells (EVTs) was inferred (Arutyunyan et al., 2023). This analysis showed that a subset of EVTs (EVTs-2) can transit either into interstitial EVTs that invade through decidual stroma or into endovascular EVTs that move down inside the arteries. Thus, this study demonstrates how high-quality single-cell and spatial data can be integrated to identify the spatial organization of cell-type subsets and their developmental relationships.

Similarly, normal uterine functions are required during pregnancy and depend on crosstalk among multiple cell types in uterine microenvironments. By applying 10× Genomics Visium to the embryo implantation site of the mouse uterus on pregnancy Day 7.5, Li et al. (2022) identified 11 cellular neighborhoods, including a mesometrial myometrium (MMy), an anti-mesometrial myometrium, a mesometrial decidua (MD) enriched with natural killer (NK) cells, a vascular sinus zone (VSZ) for maternal vessel remodeling, a fetal–maternal interface (FMI), a primary decidual zone (PDZ), a transition decidual zone (TDZ), a secondary decidual zone (SDZ), undifferentiated stroma (udStr), uterine glands, and the embryo (Fig. 2C, upper panel). Consistent with previous histological studies, the authors found that the MMy is located next to the mesometrial uterine artery, and the embryo environment is surrounded by the decidual zone composed of polyploid decidual cells. Analysis of gene expression further showed that the decidual zone can be divided into the PDZ, the SDZ, and the TDZ. The PDZ consists of cells that express high levels of prolactin genes *Prl3c1*, *Prl8a2*, and troponin gene *Tnnc1*. In contrast, the SDZ expresses metallothionein genes *Mt3*, *Mt4*, and cochlin. Using the ST data, the authors identified three major communication regions among the uterine neighborhoods. The first one consists of the embryo and its adjacent FMI and PDZ microenvironments. The other two regions can be divided into the mesometrial pole and the anti-mesometrial pole. The mesometrial pole is made up of the MD, VSZ, and FMI, while the anti-mesometrial pole consists of PDZ, TDZ, SDZ, and udStr. These neighborhoods communicate not only within the regions but also between regions. Together, this study demonstrates the complex molecular and cellular interactions that occur during early pregnancy. Given the low spatial resolution of the 10× Genomics Visium technology, future studies using single-cell-level ST approaches may better resolve the spatial heterogeneity of cell-type distributions and communications in the uterine microenvironment.

Besides the testis and the uterus, the mouse placenta has also been studied using ST to reveal its cellular neighborhoods. By using STARmap to target 903 placental genes, He et al. (2021) discovered distinct spatial patterns of placental cell types. The authors found that a subset of maternal decidua cells (MD-1), a subset of trophoblast giant cells (TG-2), and maternal NK cells mainly self-aggregate, while a subset of glandular trophoblast cells (GT-2), TG-1, TG-3, endothelial, and stromal cells exhibit high spatial mixing with each other. Furthermore, to investigate if the neighbors of a cell influences the gene expression of the cell, the authors performed clustering of MD-1 cells based on their gene expression and cellular neighborhood compositions, respectively. Both clustering results identified the same two subtypes, suggesting that the spatial environment may shape the gene expression landscape of MD-1 cells. Future functional studies are needed to go beyond this correlation analysis to establish causality between cellular microenvironment and gene expression.

Finally, ST technologies have also been applied to embryos for tissue structure analysis (Fig. 2D). For example, Chen et al. (2022) applied Stereo-seq to mouse embryos. In the spinal cord region of an E13.5 embryo, the authors identified the *Hopx* (HOP homeobox)$^+$ ventricular zone, *Slc5a7* (solute carrier family 5, member 7)$^+$ marginal zone, *Vsnl1* (visinin-like 1)$^+$ basal plate, *Fut9* (fucosyltransferase 9)$^+$ ventral and *Hoxb8*$^+$ lateral parts of the spinal alar plate, and *Pdyn* (prodynorphin)$^+$ superficial stratum of spinal basal plate. Furthermore, spatial clustering of cell types in the embryonic brain recapitulated known anatomically defined brain regions including the ventricular and mantle zones of the pallium, subpallium, midbrain, hindbrain, diencephalon, cerebellum, hypothalamus, olfactory bulb, and choroid plexus. In another study, Srivatsan et al. (2021) used sci-Space data of mouse embryos to delineate the spatial gradients of cellular differentiation and neuronal migration. The authors found that in the pallium, immature neurons migrate and differentiate radially outward, leading to the inside-out development of the cortical layers. In the subpallium, cortical interneurons born in the ganglionic eminences migrate tangentially to populate the developing cortex and olfactory bulb. Moreover, midbrain neurons seem to migrate both radially, toward the pial surface, and tangentially, parallel to the pial surface, to populate this region. Together, these studies demonstrate the ability of ST technologies to resolve complex tissue structure, such as the embryo.

Reproductive pathology-associated spatial microenvironments

Pathological states in a tissue are often associated with altered cellular microenvironments. By identifying genes with spatially patterned expressions and spatially clustered cell populations, ST is well positioned to detect and characterize such alterations.

Diabetes mellitus has been known to impact male fertility through multiple mechanisms, such as disruption of spermatogenesis, testicular degeneration and apoptotic changes, and endocrine disorders (Bhat et al., 2006; Agbaje et al., 2007; Ricci et al., 2009; Schoeller et al., 2012; Jangir and Jain, 2014; Maresch et al., 2018). Applying Slide-seqV2 to testis samples from leptin-deficient diabetic mice (*ob/ob*) and WT mice identified genes with altered spatial expression patterns such as *Smcp* (sperm mitochondria-associated cysteine-rich protein) and *Malat1* (metastasis-associated lung adenocarcinoma transcript 1). Further analysis of Slide-seqV2 data showed a significant increase in the extent of spatial mixing between haploid spermatids and other testicular cell types in *ob/ob* seminiferous tubules (Chen et al., 2021b), suggesting that the disruption of the spatial structure of seminiferous tubules is a potential mechanism of diabetes-induced testicular injuries.

Besides the testis, other male reproductive organs, such as the prostate, have also been investigated using ST approaches. Hirz et al. (2023) applied Slide-seqV2 to study the prostate tumor microenvironment. First, compared to the spatial configuration of

the healthy prostate tissue in which well-organized prostate epithelial glands are surrounded by immune and non-immune stromal cells including fibroblasts, pericytes, and endothelial cells, the tissue architecture was notably disrupted in the cancerous prostate. The authors found that the spatial distributions of fibroblasts, endothelial cells, and pericytes became more dispersed compared to those in healthy tissues. Second, compared to the healthy prostate tissue in which an organized glandular epithelium contains a well-structured bilayer of basal and luminal cells, there was an expansion of the luminal epithelial population and loss of the well-organized glands in tumor-adjacent normal sample. Third, among the four epithelial subpopulations, the spatial organizations of the club and hillock cells were disrupted in the tumor and tumor-adjacent normal tissues. Finally, to infer cell–cell communications (CCC), the Slide-seqV2 data were used to construct a graph of physically adjacent cells, which permitted testing of whether a ligand–receptor (LR) score, defined as a product of the two corresponding expression levels, was significantly higher in physically adjacent cells than would be expected from a randomized spatial arrangement. This analysis revealed 405 statistically significant potential LR interactions. Focusing on tumor–stromal communication, the authors found that tumor cells expressing vascular endothelial growth factors (VEGFA and VEGFB) can stimulate a subpopulation of endothelial cells through VEGF receptors, FLT143 and beta-1 integrin. Potential interactions between tumor cells and fibroblasts (COL9A2-ITGA1) and tumor cells with a subpopulation of pericytes cells (COL12A1-ITGA1) were also identified. Together, this study demonstrates the power of ST in dissecting the prostate tumor microenvironment as well as tumor–stromal cell interactions.

In the human uterus, Garcia-Alonso *et al.*, identified four main groups of human endometrial epithelial cells based on their marker gene expression using scRNA-seq: a *SOX9* (SRY-box transcription factor 9)$^+$ population; *PIFO* (primary cilia formation)$^+$*TPPP3* (tubulin polymerization promoting protein family member 3)$^+$ ciliated cells; *LGR5*$^+$ lumenal cells; and *SCGB2A2* (secretoglobin family 2A member 2)$^+$ glandular cells. Further analysis of the *SOX9*$^+$ population revealed three cell clusters: *SOX9*$^+$*LGR5*$^+$ cells; *SOX9*$^+$*LGR5*$^-$ cells; and proliferative *SOX9*$^+$ cells. By integrating scRNA-seq and 10× Genomics Visium data, the authors showed that *SOX9*$^+$*LGR5*$^+$ cells are spatially enriched in the surface epithelium; *SOX9*$^+$*LGR5*$^-$ cells locate in the basal glands; and proliferative *SOX9*$^+$ cells are spatially mapped to glands in the regenerating superficial layer. ISH experiments further confirmed the spatial distribution of proliferative *SOX9*$^+$ cells by showing high expression of the proliferative marker *MKI67* in the superficial layer of the endometrium during the proliferative phase (Garcia-Alonso *et al.*, 2021). In the same study, the authors correlated the clinical stages of endometrial adenocarcinomas with the three clusters of the *SOX9*$^+$ population. The more advanced stages of endometrial adenocarcinomas (stages III and IV) were found to have a greater *SOX9*$^+$LGR5$^+$ signal. This *SOX9*$^+$*LGR5*$^+$ signal is also stronger in endometrial tumors characterized by high copy number alterations and is linked with a worse prognosis (Garcia-Alonso *et al.*, 2021). These data demonstrate the importance of pinpointing the molecular and spatial identity of cellular subtypes for disease diagnosis and treatment. In another study, Fonseca *et al.* (2023) applied 10× Genomics Visium to study endometriosis—a disease characterized by endometrial-like tissue growing outside of the uterine cavity. *FOXJ1*$^+$ ciliated cells and *LGR5*$^+$ and *SOX9*$^+$ cells were found to be surrounded by *KRT10* (keratin 10)$^+$ cells both within and outside of the endometriosis lesion proper. Furthermore, *ECM1* (extracellular matrix protein 1)$^+$ and *MMP11* (matrix metallopeptidase 11)$^+$ endometrial-type stroma cells were detected in the endometriosis lesions. *CFD* (complement factor D)$^+$ peritoneal fibroblasts were separated from the lesions by a region of *C7* (complement C7)$^+$ fibroblasts scattered with *FAP* (fibroblast activation protein alpha)$^+$ cells. This study demonstrates the cellular and spatial heterogeneity in the endometriosis lesion.

In the ovary, high-grade serous ovarian carcinoma (HGSC) is the most common type of ovarian cancer and is also highly chemosensitive. Platinum-based combination chemotherapy is an important treatment for this disease, but patient responses to the treatment vary significantly (Matulonis *et al.*, 2016). The mechanisms behind this diversity in response to treatment are unclear (Peres *et al.*, 2020). Stur *et al.* (2022) applied 10× Genomics Visium to investigate the reasons behind the different responses to neoadjuvant chemotherapy from patients with HGSC. The authors uncovered more stromal-dominated cell groups, largely formed by myofibroblasts rather than conventional cancer-associated fibroblasts in the tumor samples from the poor responder (PR) group. By contrast, tumors of excellent responders (ER) contain a high proportion of immune cells, including T cells, B cells, and NK cells. Unsupervised clustering of the ST data revealed nine cell clusters. Significant differences in the spatial distribution of these clusters were observed. For example, in the PR group, the clusters are physically larger and distributed throughout the whole tissue area, whereas in the ER group, clusters are smaller and more compact. Furthermore, some clusters are located close to each other in one group but significantly farther from each other in the other group, indicating differential cell-to-cell contacts in different groups (Fig. 2B, iii). Follow-up studies on how the cell-to-cell contacts differ in the two patient groups at the molecular level (e.g. LR interactions) would provide mechanistic insights into the differential responses to chemotherapy from patients with HGSC.

In addition to ovarian cancer, cervical cancer also threatens the reproductive health of women worldwide. Recently, Ou *et al.* (2022) used snRNA-seq and Stereo-seq to analyze the gene expression patterns and cellular interactions in cervical squamous cell carcinoma tumors (Fig. 2B, iv). The authors identified six tissue clusters based on gene expression patterns: tumor, stroma (without obvious inflammation), inflammation (stroma with diffuse inflammation or focal inflammation), gland, blood vessel, and necrosis. The tumor cluster was further divided into hypermetabolic tumor and hypometabolic tumor based on the expression level of genes associated with oxygen status and energy production pathways. Of interest, a unique spatial cluster largely composed of cancer-associated myofibroblasts (myCAFs) was found outside the hypermetabolic tumor regions. Differential gene expression analysis showed that the myCAF$^+$ tumors are more active in energy usage, metabolism, mitosis, and cell growth than myCAF$^-$ tumors, whereas signaling activities associated with cellular adhesion, apoptosis, and immune responses are down-regulated in myCAF$^+$ tumors. These observations indicate that the presence of myCAFs may play important roles in supporting cervical cancer progression.

Challenges and outlook of applying ST technologies to studying mammalian reproduction

A growing number of studies, as reviewed above, have demonstrated the utility of ST technologies in revealing the biological

regulation of reproductive physiology and pathology. However, challenges remain to apply ST technologies to study reproductive systems.

Experimental challenges

The number of unique molecules per cell captured by the current ST technologies is, in most cases, less than that captured by the state-of-art scRNA-seq technologies. This has hindered the spatial profiling of lowly expressed genes and rare cell types in reproductive systems. At the RNA level, the current ST methods mostly focus on the detection of mRNAs, whereas the spatial information of non-coding RNAs is rarely resolved, even though non-coding RNAs play important roles in reproductive systems (Bourc'his and Voinnet, 2010; Cabili *et al.*, 2011; Pauli *et al.*, 2011; McIver *et al.*, 2012; de Mateo and Sassone-Corsi, 2014). A combination of *in situ* RNA polyadenylation with existing ST technologies may solve the issue (McKellar *et al.*, 2022). Furthermore, many ST technologies are optimized for fresh frozen tissue specimens and have a high requirement for RNA integrity, preventing their applications in clinical research where clinical specimens are often formalin fixed and paraffin embedded and contain a large quantity of fragmented RNAs. Thus, for samples with low RNA quality, a targeted gene panel may be applied to enhance the capture efficiency of the ST technologies (Mirzazadeh *et al.*, 2023). Finally, many ST technologies rely on specialized equipment or custom-made arrays. Although several commercial ST solutions are available, the high costs of these solutions limit their accessibility to non-specialist laboratories (Table 1). Institutional or regional core facilities that provide these commercial solutions on a fee-for-service basis would help to democratize the use of ST technologies.

Computational challenges

Current ST approaches span a wide range of spatial resolution, from broad tissue regions to subcellular localization (Table 1). In reproductive systems, cell sizes vary significantly among different cell types and even within cells belonging to the same cell type but at different developmental stages. The ability to accurately perform cellular segmentation on the measured molecules is, therefore, crucial to many downstream applications, such as quantifying cell-type composition and tissue organization. For example, in ISH- and ISS-based ST approaches, individual transcripts need to be grouped into cells from microscopy images based on image masks generated by a segmentation algorithm. This algorithm often needs extensive customization and fine tuning for each tissue type. Thus, innovations in computer vision, such as the recent machine learning-based approaches (Berg *et al.*, 2019; Pachitariu and Stringer, 2022), will greatly accelerate the ability of ST tools to be applied to various reproductive organs.

Another key computational challenge is to analyze CCC specifically for ST data. While many analyses focus on the structural relationship of cells, such as calculating the frequencies or pairwise co-occurrence of cell types in different tissue regions, few tools are available to model CCC at the molecular level. Current methods to examine molecular CCC do so in a pairwise and local manner, focusing on information between cells or in the neighborhoods of individual cells (Cang and Nie, 2020; Dries *et al.*, 2021; Garcia-Alonso *et al.*, 2021; Shao *et al.*, 2022). As a result, the collective or global information in CCC, such as the competition between cells, and long-range cell–cell interactions, such as the endocrine and telecrine signaling (both are common regulatory mechanisms in the reproductive organs), are neglected. Incorporating prior knowledge of cell–cell competition and classification of LR interactions into short-range and long-range communications might be helpful to infer the comprehensive communication categories computationally.

The outlook

The rapid progress in the development of ST technologies will open new possibilities for the study of reproductive systems and beyond. We anticipate several exciting new directions the field is heading.

First, going beyond capturing a snapshot of molecular abundancy in a spatially resolved manner, ST technologies can be applied to measure cellular dynamics. For example, by combining ethynyl-2′-deoxyuridine labeling of the transcriptome with STARmap, temporally resolved *in situ* sequencing and mapping was recently developed to simultaneously profile the age and location of individual RNA molecules within intact cells and tissues (Ren *et al.*, 2023). Furthermore, novel temporal recording technologies have enabled the encoding of cellular lineages (Shipman *et al.*, 2016; Frieda *et al.*, 2017; Chen *et al.*, 2020; Choi *et al.*, 2022) and transcriptomic states (Chen *et al.*, 2021a; Rodriques *et al.*, 2021) in the form of DNA or RNA mutations. Combining these recording approaches with ST technologies may reveal cellular histories and dynamics during gamete development and embryonic development within the native tissue context.

Second, progress in single-cell technologies has already enabled multi-modal profiling of the transcriptome, the proteome, and the epigenome (Zhu *et al.*, 2020; Ogbeide *et al.*, 2022). Spatial multi-omics technologies may provide solutions for spatially resolved muti-modal profiling. Recent developments in DNA-tagged antibodies and application of LCM have enabled highly multiplexed protein or whole proteome readouts (Goltsev *et al.*, 2018; Merritt *et al.*, 2020; Mund *et al.*, 2022), respectively. The protein A-Tn5 transposase fusion has enabled highly multiplexed spatial readouts of the epigenome (Deng *et al.*, 2022a,b; Lu *et al.*, 2022). These approaches can be readily coupled with ST measurements. For instance, spatially resolved co-capture of the transcriptome and the epigenome in E13 mouse embryo has been proven feasible (Zhang *et al.*, 2023). Soon, whole proteome-targeting antibody/nanobody libraries may be developed for *in situ* measurements.

Finally, ST technologies may offer an opportunity to dissect gene functions at scale within the native tissue context. For biological processes like gametogenesis, thousands of genes are involved, which makes it difficult to pinpoint the functional contribution of each gene. Traditionally, the *in vivo* functions of a gene can be analyzed by generating KO mouse lines. However, this approach demands significant time and resources, making it challenging to scale. Emerging technologies, such as clustered regularly interspaced short palindromic repeats (CRISPR) screens coupled with scRNA-seq, can examine gene functions at scale (Dixit *et al.*, 2016; Datlinger *et al.*, 2017). While cell-intrinsic effects of a gene perturbation may be read out using scRNA-seq, the extracellular effects of a gene perturbation cannot be assessed owing to tissue disassociation. This excludes using CRISPR screens to identify genes controlling phenotypes that require spatial resolution to assess, such as genes encoding secreted factors. Therefore, future efforts to develop a CRISPR screen approach that retains the spatial context of a biological process will enable profiling of phenotypes that cannot be accessed in the absence of tissue context, such as cellular localization and cell–cell interactions.

Conclusion

Reproduction is essential for the continuation of our species, as it ensures that parental genetic and epigenetic information are passed on to the next generation. Besides producing gametes, the reproductive system also provides the environment for the appropriate development of the embryo. New genomic and computational tools offer unique opportunities to study the intricate spatiotemporal regulatory mechanisms that are required for mammalian reproduction.

Like scRNA-seq, ST technologies hold tremendous potential for clinical applications. First, the identification of signaling pathways regulating human germ cell differentiation and proliferation will enable the development of protocols for human *in vitro* spermatogenesis—a technology that would have tremendous impact on fertility preservation. Second, spatially altered genes, cell types, and spatial neighborhoods under pathological conditions identified by ST technologies may be considered as markers for infertility, cancer, or other reproductive disorders. Finally, CCC through LR interactions revealed by ST technologies may also serve as new therapeutic targets for either treating reproductive disorders or developing novel contraceptive approaches.

In summary, our review discusses the novel biological insights that have been revealed by studies using ST technologies, while also shedding light on what is yet to come. We hope that this review will provide reproductive biologists and clinicians with a much-needed update on the state of art of ST technologies. This review may also facilitate the adoption of cutting-edge spatial omics technologies in both basic and clinical reproductive research.

Data availability

No new data were generated or analyzed in support of this research.

Authors' roles

X.Z., Q.C., and H.C. wrote the article, with input from S.R., E.J.G., and M.E. All authors read and approved the final version. Figures were generated using BioRender.

Funding

H.C. acknowledges support from the National Institute of Health (R21HD110878) and Green Center for Reproductive Biology Endowment. M.E. is supported by the UT Southwestern Medical Center REI fellow training fund.

Conflict of interest

The authors declare no conflict of interest concerning this review.

References

10× Genomics. The Visium Platform. 2019. https://www.10xgenomics.com/products/spatial-gene-expression (12 June 2023, date last accessed).

10× Genomics. The Xenium platform. 2023. https://www.10xgenomics.com/platforms/xenium (12 June 2023, date last accessed).

Agbaje IM, Rogers DA, McVicar CM, McClure N, Atkinson AB, Mallidis C, Lewis SE. Insulin dependant diabetes mellitus: implications for male reproductive function. *Hum Reprod* 2007;**22**:1871–1877.

Alon S, Goodwin DR, Sinha A, Wassie AT, Chen F, Daugharthy ER, Bando Y, Kajita A, Xue AG, Marrett K *et al.* Expansion sequencing: spatially precise in situ transcriptomics in intact biological systems. *Science* 2021;**371**:eaax2656.

Argelaguet R, Clark SJ, Mohammed H, Stapel LC, Krueger C, Kapourani CA, Imaz-Rosshandler I, Lohoff T, Xiang Y, Hanna CW *et al.* Multi-omics profiling of mouse gastrulation at single-cell resolution. *Nature* 2019;**576**:487–491.

Arutyunyan A, Roberts K, Troulé K, Wong FCK, Sheridan MA, Kats I, Garcia-Alonso L, Velten B, Hoo R, Ruiz-Morales ER *et al.* Spatial multiomics map of trophoblast development in early pregnancy. *Nature* 2023;**616**:143–151.

Berg S, Kutra D, Kroeger T, Straehle CN, Kausler BX, Haubold C, Schiegg M, Ales J, Beier T, Rudy M *et al.* ilastik: interactive machine learning for (bio)image analysis. *Nat Methods* 2019;**16**:1226–1232.

Bhat GK, Sea TL, Olatinwo MO, Simorangkir D, Ford GD, Ford BD, Mann DR. Influence of a leptin deficiency on testicular morphology, germ cell apoptosis, and expression levels of apoptosis-related genes in the mouse. *J Androl* 2006;**27**:302–310.

Bourc'his D, Voinnet O. A small-RNA perspective on gametogenesis, fertilization, and early zygotic development. *Science* 2010;**330**:617–622.

Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, Rinn JL. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev* 2011;**25**:1915–1927.

Cabot B, Cabot RA. Chromatin remodeling in mammalian embryos. *Reproduction* 2018;**155**:R147–R158.

Cang Z, Nie Q. Inferring spatial and signaling relationships between cells from single cell transcriptomic data. *Nat Commun* 2020;**11**:2084.

Chen KH, Boettiger AN, Moffitt JR, Wang S, Zhuang X. RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* 2015;**348**:aaa6090.

Chen W, Choi J, Nathans JF, Agarwal V, Martin B, Nichols E, Leith A, Lee C, Shendure J. Multiplex genomic recording of enhancer and signal transduction activity in mammalian cells. bioRxiv 2021a;11.05.467434.

Chen A, Liao S, Cheng M, Ma K, Wu L, Lai Y, Qiu X, Yang J, Xu J, Hao S *et al.* Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell* 2022;**185**:1777–1792 e1721.

Chen H, Liu S, Padula S, Lesman D, Griswold K, Lin A, Zhao T, Marshall JL, Chen F. Efficient, continuous mutagenesis in human cells using a pseudo-random DNA editor. *Nat Biotechnol* 2020;**38**:165–168.

Chen H, Murray E, Sinha A, Laumas A, Li J, Lesman D, Nie X, Hotaling J, Guo J, Cairns BR *et al.* Dissecting mammalian spermatogenesis using spatial transcriptomics. *Cell Rep* 2021b;**37**:109915.

Chen X, Sun YC, Church GM, Lee JH, Zador AM. Efficient in situ barcode sequencing using padlock probe-based BaristaSeq. *Nucleic Acids Res* 2018;**46**:e22.

Chen J, Suo S, Tam PP, Han JJ, Peng G, Jing N. Spatial transcriptomic analysis of cryosectioned tissue samples with Geo-seq. *Nat Protoc* 2017;**12**:566–580.

Cheung TK, Lee CY, Bayer FP, McCoy A, Kuster B, Rose CM. Defining the carrier proteome limit for single-cell proteomics. *Nat Methods* 2021;**18**:76–83.

Cho CS, Xi J, Si Y, Park SR, Hsu JE, Kim M, Jun G, Kang HM, Lee JH. Microscopic examination of spatial transcriptome using Seq-Scope. *Cell* 2021;**184**:3559–3572.e3522.

Choi J, Chen W, Minkina A, Chardon FM, Suiter CC, Regalado SG, Domcke S, Hamazaki N, Lee C, Martin B *et al.* A time-resolved, multi-symbol molecular recorder via sequential genome editing. *Nature* 2022;**608**:98–107.

Datlinger P, Rendeiro AF, Schmidl C, Krausgruber T, Traxler P, Klughammer J, Schuster LC, Kuchler A, Alpar D, Bock C. Pooled CRISPR screening with single-cell transcriptome readout. *Nat Methods* 2017;**14**:297–301.

de Mateo S, Sassone-Corsi P. Regulation of spermatogenesis by small non-coding RNAs: role of the germ granule. *Semin Cell Dev Biol* 2014;**29**:84–92.

Deng Y, Bartosovic M, Kukanja P, Zhang D, Liu Y, Su G, Enninful A, Bai Z, Castelo-Branco G, Fan R. Spatial-CUT&Tag: spatially resolved chromatin modification profiling at the cellular level. *Science* 2022a;**375**:681–686.

Deng Y, Bartosovic M, Ma S, Zhang D, Kukanja P, Xiao Y, Su G, Liu Y, Qin X, Rosoklija GB *et al.* Spatial profiling of chromatin accessibility in mouse and human tissues. *Nature* 2022b;**609**:375–383.

Dixit A, Parnas O, Li B, Chen J, Fulco CP, Jerby-Arnon L, Marjanovic ND, Dionne D, Burks T, Raychowdhury R *et al.* Perturb-Seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell* 2016;**167**:1853–1866.e1817.

Dries R, Chen J, Del Rossi N, Khan MM, Sistig A, Yuan GC. Advances in spatial transcriptomic data analysis. *Genome Res* 2021;**31**:1706–1718.

Dries R, Zhu Q, Dong R, Eng CL, Li H, Liu K, Fu Y, Zhao T, Sarkar A, Bao F *et al.* Giotto: a toolbox for integrative analysis and visualization of spatial expression data. *Genome Biol* 2021;**22**:78.

Emmert-Buck MR, Bonner RF, Smith PD, Chuaqui RF, Zhuang Z, Goldstein SR, Weiss RA, Liotta LA. Laser capture microdissection. *Science* 1996;**274**:998–1001.

Eng CL, Lawson M, Zhu Q, Dries R, Koulena N, Takei Y, Yun J, Cronin C, Karp C, Yuan GC *et al.* Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH. *Nature* 2019;**568**:235–239.

Fang F, Iaquinta PJ, Xia N, Liu L, Diao L, Reijo Pera RA. Transcriptional control of human gametogenesis. *Hum Reprod Update* 2022;**28**:313–345.

Fang X, Ni N, Lydon JP, Ivanov I, Bayless KJ, Rijnkels M, Li Q. Enhancer of Zeste 2 polycomb repressive complex 2 subunit is required for uterine epithelial integrity. *Am J Pathol* 2019;**189**:1212–1225.

Femino AM, Fay FS, Fogarty K, Singer RH. Visualization of single RNA transcripts in situ. *Science* 1998;**280**:585–590.

Ferrero H, Corachán A, Aguilar A, Quiñonero A, Carbajo-García MC, Alamá P, Tejera A, Taboas E, Muñoz E, Pellicer A *et al.* Single-cell RNA sequencing of oocytes from ovarian endometriosis patients reveals a differential transcriptomic profile associated with lower quality. *Hum Reprod* 2019;**34**:1302–1312.

Fonseca MAS, Haro M, Wright KN, Lin X, Abbasi F, Sun J, Hernandez L, Orr NL, Hong J, Choi-Kuaea Y *et al.* Single-cell transcriptomic analysis of endometriosis. *Nat Genet* 2023;**55**:255–267.

Frieda KL, Linton JM, Hormoz S, Choi J, Chow KK, Singer ZS, Budde MW, Elowitz MB, Cai L. Synthetic recording and in situ readout of lineage information in single cells. *Nature* 2017;**541**:107–111.

Fu X, Sun L, Dong R, Chen JY, Silakit R, Condon LF, Lin Y, Lin S, Palmiter RD, Gu L. Polony gels enable amplifiable DNA stamping and spatial transcriptomics of chronic pain. *Cell* 2022;**185**:4621–4633.e17.

Garcia-Alonso L, Handfield L-F, Roberts K, Nikolakopoulou K, Fernando RC, Gardner L, Woodhams B, Arutyunyan A, Polanski K, Hoo R *et al.* Mapping the temporal and spatial dynamics of the human endometrium in vivo and in vitro. *Nat Genet* 2021;**53**:1698–1711.

Garcia-Alonso L, Lorenzi V, Mazzeo CI, Alves-Lopes JP, Roberts K, Sancho-Serra C, Engelbert J, Marečková M, Gruhn WH, Botting RA *et al.* Single-cell roadmap of human gonadal development. *Nature* 2022;**607**:540–547.

Gershon E, Plaks V, Dekel N. Gap junctions in the ovary: expression, localization and function. *Mol Cell Endocrinol* 2008;**282**:18–25.

Goltsev Y, Samusik N, Kennedy-Darling J, Bhate S, Hale M, Vazquez G, Black S, Nolan GP. Deep profiling of mouse splenic architecture with CODEX multiplexed imaging. *Cell* 2018;**174**:968–981.e915.

Green CD, Ma Q, Manske GL, Shami AN, Zheng X, Marini S, Moritz L, Sultan C, Gurczynski SJ, Moore BB *et al.* A comprehensive roadmap of murine spermatogenesis defined by single-cell RNA-Seq. *Dev Cell* 2018;**46**:651–667.e610.

Guilliams M, Bonnardel J, Haest B, Vanderborght B, Wagner C, Remmerie A, Bujko A, Martens L, Thoné T, Browaeys R et al Spatial proteogenomics reveals distinct and evolutionarily conserved hepatic macrophage niches. *Cell* 2022;**185**:379–396.e338.

Guo J, Grow EJ, Mlcochova H, Maher GJ, Lindskog C, Nie X, Guo Y, Takei Y, Yun J, Cai L *et al.* The adult human testis transcriptional cell atlas. *Cell Res* 2018;**28**:1141–1157.

He Y, Tang X, Huang J, Ren J, Zhou H, Chen K, Liu A, Shi H, Lin Z, Li Q *et al.* ClusterMap for multi-scale clustering analysis of spatial gene expression. *Nat Commun* 2021;**12**:5909.

Hess RA, Renato de Franca L. Spermatogenesis and cycle of the seminiferous epithelium. *Adv Exp Med Biol* 2008;**636**:1–15.

Hirz T, Mei S, Sarkar H, Kfoury Y, Wu S, Verhoeven BM, Subtelny AO, Zlatev DV, Wszolek MW, Salari K *et al.* Dissecting the immune suppressive human prostate tumor microenvironment via integrated single-cell and spatial transcriptomic analyses. *Nat Commun* 2023;**14**:663.

Hofmann MC, McBeath E. Sertoli cell-germ cell interactions within the niche: paracrine and juxtacrine molecular communications. *Front Endocrinol* 2022;**13**:897062.

Holler K, Neuschulz A, Drewe-Boß P, Mintcheva J, Spanjaard B, Arsiè R, Ohler U, Landthaler M, Junker JP. Spatio-temporal mRNA tracking in the early zebrafish embryo. *Nat Commun* 2021;**12**:3358.

Hwang WL, Jagadeesh KA, Guo JA, Hoffman HI, Yadollahpour P, Reeves JW, Mohan R, Drokhlyansky E, Van Wittenberghe N, Ashenberg O *et al.* Single-nucleus and spatial transcriptome profiling of pancreatic cancer identifies multicellular dynamics associated with neoadjuvant treatment. *Nat Genet* 2022;**54**:1178–1191.

Jangir RN, Jain GC. Diabetes mellitus induced impairment of male reproductive functions: a review. *Curr Diabetes Rev* 2014;**10**:147–157.

Jemt A, Salmén F, Lundmark A, Mollbrink A, Fernández Navarro J, Ståhl PL, Yucel-Lindberg T, Lundeberg J. An automated approach to prepare tissue-derived spatially barcoded RNA-sequencing libraries. *Sci Rep* 2016;**6**:37137–37137.

Junker JP, Noël ES, Guryev V, Peterson KA, Shah G, Huisken J, McMahon AP, Berezikov E, Bakkers J, van Oudenaarden A. Genome-wide RNA Tomography in the zebrafish embryo. *Cell* 2014;**159**:662–675.

Ke R, Mignardi M, Pacureanu A, Svedlund J, Botling J, Wählby C, Nilsson M. In situ sequencing for RNA analysis in preserved tissue and cells. *Nat Methods* 2013;**10**:857–860.

Kuppe C, Ramirez Flores RO, Li Z, Hayat S, Levinson RT, Liao X, Hannani MT, Tanevski J, Wünnemann F, Nagai JS *et al.* Spatial multi-omic map of human myocardial infarction. *Nature* 2022;**608**:766–777.

Larose H, Shami AN, Abbott H, Manske G, Lei L, Hammoud SS. Gametogenesis: a journey from inception to conception. *Curr Top Dev Biol* 2019;**132**:257–310.

Lee JH, Daugharthy ER, Scheiman J, Kalhor R, Yang JL, Ferrante TC, Terry R, Jeanty SS, Li C, Amamoto R *et al.* Highly multiplexed subcellular RNA sequencing in situ. *Science* 2014;**343**:1360–1363.

Li L, Guo F, Gao Y, Ren Y, Yuan P, Yan L, Li R, Lian Y, Li J, Hu B *et al.* Single-cell multi-omics sequencing of human early embryos. *Nat Cell Biol* 2018;**20**:847–858.

Li L, Li L, Li Q, Liu X, Ma X, Yong J, Gao S, Wu X, Wei Y, Wang X *et al.* Dissecting the epigenomic dynamics of human fetal germ cell development at single-cell resolution. *Cell Res* 2021;**31**:463–477.

Li R, Wang TY, Xu X, Emery O, Yi M, Wu SP, DeMayo FJ. Spatial transcriptomic profiles of mouse uterine microenvironments at pregnancy day 7.5. *Biol Reprod* 2022;**107**:529–545.

Liu Y, Yang M, Deng Y, Su G, Enninful A, Guo CC, Tebaldi T, Zhang D, Kim D, Bai Z *et al.* High-spatial-resolution multi-omics sequencing via deterministic barcoding in tissue. *Cell* 2020;**183**:1665–1681.e1618.

Longo SK, Guo MG, Ji AL, Khavari PA. Integrating single-cell and spatial transcriptomics to elucidate intercellular tissue dynamics. *Nat Rev Genet* 2021;**22**:627–644.

Lu T, Ang CE, Zhuang X. Spatially resolved epigenomic profiling of single cells in complex tissues. *Cell* 2022;**185**:4448–4464.e4417.

Lubeck E, Coskun AF, Zhiyentayev T, Ahmad M, Cai L. Single-cell in situ RNA profiling by sequential hybridization. *Nat Methods* 2014;**11**:360–361.

Maresch CC, Stute DC, Alves MG, Oliveira PF, de Kretser DM, Linn T. Diabetes-induced hyperglycemia impairs male reproductive function: a systematic review. *Hum Reprod Update* 2018;**24**:86–105.

Marston A, Amon A. Meiosis: cell-cycle controls shuffle and deal. *Nat Rev Mol Cell Biol* 2004;**5**:983–997.

Marx V. Method of the year: spatially resolved transcriptomics. *Nat Methods* 2021;**18**:9–14.

Matulonis UA, Sood AK, Fallowfield L, Howitt BE, Sehouli J, Karlan BY. Ovarian cancer. *Nat Rev Dis Primers* 2016;**2**:16061.

Maurya VK, DeMayo FJ, Lydon JP. Illuminating the "black box" of progesterone-dependent embryo implantation using engineered mice. *Front Cell Dev Biol* 2021;**9**:640907.

McGinnis CS, Patterson DM, Winkler J, Conrad DN, Hein MY, Srivastava V, Hu JL, Murrow LM, Weissman JS, Werb Z *et al.* MULTI-seq: sample multiplexing for single-cell RNA sequencing using lipid-tagged indices. *Nat Methods* 2019;**16**:619–626.

McIver SC, Roman SD, Nixon B, McLaughlin EA. miRNA and mammalian male germ cells. *Hum Reprod Update* 2012;**18**:44–59.

McKellar DW, Mantri M, Hinchman MM *et al.* Spatial mapping of the total transcriptome by in situ polyadenylation. *Nat Biotechnol* 2022;**41**:513–520.

Merritt CR, Ong GT, Church SE, Barker K, Danaher P, Geiss G, Hoang M, Jung J, Liang Y, McKay-Fleisch J *et al.* Multiplex digital spatial profiling of proteins and RNA in fixed tissue. *Nat Biotechnol* 2020;**38**:586–599.

Mesa AM, Mao J, Medrano TI, Bivens NJ, Jurkevich A, Tuteja G, Cooke PS, Rosenfeld CS. Spatial transcriptomics analysis of uterine gene expression in enhancer of zeste homolog 2 conditional knockout micedagger. *Biol Reprod* 2021;**105**:1126–1139.

Mirzazadeh R, Andrusivova Z, Larsson L, Newton PT, Galicia LA, Abalo XM, Avijgan M, Kvastad L, Denadai-Souza A, Stakenborg N *et al.* Spatially resolved transcriptomic profiling of degraded and challenging fresh frozen samples. *Nat Commun* 2023;**14**:509.

Mittnenzweig M, Mayshar Y, Cheng S, Ben-Yair R, Hadas R, Rais Y, Chomsky E, Reines N, Uzonyi A, Lumerman L *et al.* A single-embryo, single-cell time-resolved model for mouse gastrulation. *Cell* 2021;**184**:2825–2842.e2822.

Mruk DD, Cheng CY. Sertoli-Sertoli and Sertoli-germ cell interactions and their significance in germ cell movement in the seminiferous epithelium during spermatogenesis. *Endocr Rev* 2004;**25**:747–806.

Mund A, Coscia F, Kriston A, Hollandi R, Kovács F, Brunner AD, Migh E, Schweizer L, Santos A, Bzorek M *et al.* Deep visual proteomics defines single-cell identity and heterogeneity. *Nat Biotechnol* 2022;**40**:1231–1240.

Nanjappa MK, Mesa AM, Medrano TI, Jefferson WN, DeMayo FJ, Williams CJ, Lydon JP, Levin ER, Cooke PS. The histone methyltransferase EZH2 is required for normal uterine development and function in micedagger. *Biol Reprod* 2019;**101**:306–317.

Nanostring. The GeoMx Platform. 2019. https://nanostring.com/products/geomx-digital-spatial-profiler/geomx-dsp-overview/ (12 June 2023, date last accessed).

Nichterwitz S, Benitez JA, Hoogstraaten R, Deng Q, Hedlund E. LCM-Seq: a method for spatial transcriptomic profiling using laser capture microdissection coupled with PolyA-based RNA sequencing. *Methods Mol Biol* 2018;**1649**:95–110.

Nichterwitz S, Chen G, Aguila Benitez J, Yilmaz M, Storvall H, Cao M, Sandberg R, Deng Q, Hedlund E. Laser capture microscopy coupled with Smart-seq2 for precise spatial transcriptomic profiling. *Nat Commun* 2016;**7**:12139.

Ogbeide S, Giannese F, Mincarelli L, Macaulay IC. Into the multiverse: advances in single-cell multiomic profiling. *Trends Genet* 2022;**38**:831–843.

Ou Z, Lin S, Qiu J, Ding W, Ren P, Chen D, Wang J, Tong Y, Wu D, Chen A *et al.* Single-nucleus RNA sequencing and spatial transcriptomics reveal the immunological microenvironment of cervical squamous cell carcinoma. *Adv Sci (Weinh)* 2022;**9**:e2203040.

Pachitariu M, Stringer C. Cellpose 2.0: how to train your own model. *Nat Methods* 2022;**19**:1634–1641.

Pauli A, Rinn JL, Schier AF. Non-coding RNAs as regulators of embryogenesis. *Nat Rev Genet* 2011;**12**:136–149.

Peres LC, Sinha S, Townsend MK, Fridley BL, Karlan BY, Lutgendorf SK, Shinn E, Sood AK, Tworoger SS. Predictors of survival trajectories among women with epithelial ovarian cancer. *Gynecol Oncol* 2020;**156**:459–466.

Phillips BT, Gassei K, Orwig KE. Spermatogonial stem cell regulation and spermatogenesis. *Philos Trans R Soc Lond B Biol Sci* 2010;**365**:1663–1678.

Rao A, Barkley D, França GS, Yanai I. Exploring tissue architecture using spatial transcriptomics. *Nature* 2021;**596**:211–220.

Ratz M, von Berlin L, Larsson L, Martin M, Westholm JO, La Manno G, Lundeberg J, Frisén J. Clonal relations in the mouse brain revealed by single-cell and spatial transcriptomics. *Nat Neurosci* 2022;**25**:285–294.

Ren J, Zhou H, Zeng H, Wang CK, Huang J, Qiu X, Maher K, Lin Z, He Y, Tang X *et al.* Spatiotemporally resolved transcriptomics reveals the subcellular RNA kinetic landscape. *Nat Methods* 2023;**20**:695–705.

Ricci G, Catizone A, Esposito R, Pisanti FA, Vietri MT, Galdieri M. Diabetic rat testes: morphological and functional alterations. *Andrologia* 2009;**41**:361–368.

Richardson B, Lehmann R. Mechanisms guiding primordial germ cell migration: strategies from different organisms. *Nat Rev Mol Cell Biol* 2010;**11**:37–49.

Rodriques SG, Chen LM, Liu S, Zhong ED, Scherrer JR, Boyden ES, Chen F. RNA timestamps identify the age of single molecules in RNA sequencing. *Nat Biotechnol* 2021;**39**:320–325.

Rodriques SG, Stickels RR, Goeva A, Martin CA, Murray E, Vanderburg CR, Welch J, Chen LM, Chen F, Macosko EZ. Slide-seq: a scalable technology for measuring genome-wide expression at high spatial resolution. *Science* 2019;**363**:1463–1467.

Schede HH, Schneider CG, Stergiadou J, Borm LE, Ranjak A, Yamawaki TM, David FPA, Lönnerberg P, Tosches MA, Codeluppi S *et al.* Spatial tissue profiling by imaging-free molecular tomography. *Nat Biotechnol* 2021;**39**:968–977.

Schoeller EL, Schon S, Moley KH. The effects of type 1 diabetes on the hypothalamic, pituitary and testes axis. *Cell Tissue Res* 2012;**349**: 839–847.

Shao X, Li C, Yang H, Lu X, Liao J, Qian J, Wang K, Cheng J, Yang P, Chen H *et al.* Knowledge-graph-based cell-cell communication inference for spatially resolved transcriptomic data with SpaTalk. *Nat Commun* 2022;**13**:4429.

Shipman SL, Nivala J, Macklis JD, Church GM. Molecular recordings by directed CRISPR spacer acquisition. *Science* 2016;**353**:aaf1175.

Srivatsan SR, Regier MC, Barkan E, Franks JM, Packer JS, Grosjean P, Duran M, Saxton S, Ladd JJ, Spielmann M *et al.* Embryo-scale, single-cell spatial transcriptomics. *Science* 2021;**373**:111–117.

Ståhl PL, Salmén F, Vickovic S, Lundmark A, Navarro JF, Magnusson J, Giacomello S, Asp M, Westholm JO, Huss M *et al.* Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* 2016;**353**:78–82.

Stickels RR, Murray E, Kumar P, Li J, Marshall JL, Di Bella DJ, Arlotta P, Macosko EZ, Chen F. Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2. *Nat Biotechnol* 2021; **39**:313–319.

Stur E, Corvigno S, Xu M, Chen K, Tan Y, Lee S, Liu J, Ricco E, Kraushaar D, Castro P *et al.* Spatially resolved transcriptomics of high-grade serous ovarian carcinoma. *iScience* 2022;**25**:103923.

Tian L, Chen F, Macosko EZ. The expanding vistas of spatial transcriptomics. *Nat Biotechnol* 2023;**41**:773–782.

Trevino LS, Wang Q, Walker CL. Phosphorylation of epigenetic "readers, writers and erasers": implications for developmental reprogramming and the epigenetic basis for health and disease. *Prog Biophys Mol Biol* 2015;**118**:8–13.

Vitak SA, Torkenczy KA, Rosenkrantz JL, Fields AJ, Christiansen L, Wong MH, Carbone L, Steemers FJ, Adey A. Sequencing thousands of single-cell genomes with combinatorial indexing. *Nat Methods* 2017;**14**:302–308.

Wang X, Allen WE, Wright MA, Sylwestrak EL, Samusik N, Vesuna S, Evans K, Liu C, Ramakrishnan C, Liu J *et al.* Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* 2018a;**361**:eaat5691.

Wang M, Liu X, Chang G, Chen Y, An G, Yan L, Gao S, Xu Y, Cui Y, Dong J *et al.* Single-cell RNA sequencing analysis reveals sequential cell fate transition during human spermatogenesis. *Cell Stem Cell* 2018b;**23**:599–614.e594.

Wang G, Moffitt JR, Zhuang X. Multiplexed imaging of high-density libraries of RNAs with MERFISH and expansion microscopy. *Sci Rep* 2018c;**8**:4847.

Wang W, Vilella F, Alama P, Moreno I, Mignardi M, Isakova A, Pan W, Simon C, Quake SR. Single-cell transcriptomic atlas of the human endometrium during the menstrual cycle. *Nat Med* 2020;**26**: 1644–1653.

Wang Z, Xu X, Li JL, Palmer C, Maric D, Dean J. Sertoli cell-only phenotype and scRNA-seq define PRAMEF12 as a factor essential for spermatogenesis in mice. *Nat Commun* 2019;**10**:5196.

Wang Y, Yuan P, Yan Z, Yang M, Huo Y, Nie Y, Zhu X, Qiao J, Yan L. Single-cell multiomics sequencing reveals the functional regulatory landscape of early embryos. *Nat Commun* 2021;**12**:1247.

Wassarman PM. Channels of communication in the ovary. *Nat Cell Biol* 2002;**4**:s7–s9.

Wu S, Yan M, Ge R, Cheng CY. Crosstalk between sertoli and germ cells in male fertility. *Trends Mol Med* 2020;**26**:215–231.

Xia C, Fan J, Emanuel G, Hao J, Zhuang X. Spatial transcriptome profiling by MERFISH reveals subcellular RNA compartmentalization and cell cycle-dependent gene expression. *Proc Natl Acad Sci U S A* 2019;**116**:19490–19499.

Yan R, Gu C, You D, Huang Z, Qian J, Yang Q, Cheng X, Zhang L, Wang H, Wang P *et al.* Decoding dynamic epigenetic landscapes in human oocytes using single-cell multi-omics sequencing. *Cell Stem Cell* 2021;**28**:1641–1656.e1647.

Yosefzon Y, David C, Tsukerman A, Pnueli L, Qiao S, Boehm U, Melamed P. An epigenetic switch repressing Tet1 in gonadotropes activates the reproductive axis. *Proc Natl Acad Sci U S A* 2017;**114**: 10131–10136.

Zeng Z, Li Y, Li Y, Luo Y. Statistical and machine learning methods for spatially resolved transcriptomics data analysis. *Genome Biol* 2022;**23**:83.

Zhang D, Deng Y, Kukanja P, Bartosovic M, Ma S, Su G, Liu Y, Xiao Y, Rosoklija GB, Dwork AJ *et al.* Spatial epigenome–transcriptome co-profiling of mammalian tissues. *Nature* 2023;**616**:113–122.

Zhang K, Hocker JD, Miller M, Hou X, Chiou J, Poirion OB, Qiu Y, Li YE, Gaulton KJ, Wang A *et al.* A single-cell atlas of chromatin accessibility in the human genome. *Cell* 2021;**184**:5985–6001.e5919.

Zhao L, Yao C, Xing X, Jing T, Li P, Zhu Z, Yang C, Zhai J, Tian R, Chen H *et al.* Single-cell analysis of developing and azoospermia human testicles reveals central role of Sertoli cells. *Nat Commun* 2020;**11**:5683.

Zhu C, Preissl S, Ren B. Single-cell multimodal omics: the power of many. *Nat Methods* 2020;**17**:11–14.

Abbreviations list

AMMy	Anti-mesometrial myometrium
Asb4	Ankyrin repeat and SOCS box-containing 4
C7	Complement C7
camk2g1	Calcium/calmodulin-dependent protein kinase II gamma 1
CCC	Cell–cell communication
CDHR3	Cadherin-related family member 3
CFD	Complement factor D
CRISPR	Clustered regularly interspaced short palindromic repeats
CSCC	Cervical squamous cell carcinoma
Cxcl14	Chemokine (C-X-C motif) ligand 14
dazl	Deleted in azoospermia-like
Dio2	Deiodinase, iodothyronine, type II
ECM1	Extracellular matrix protein 1
ER	Excellent responders
ETV5	ETS variant transcription factor 5
EVTs	Extravillous trophoblast cells
Ezh2	Enhancer of zeste homolog 2
FAM183A	Family with sequence similarity 183 member A
FAP	Fibroblast activation protein alpha
FMI	Fetal–maternal interface
FOXJ1	Forkhead box J1
Fut9	Fucosyltransferase 9
GT	Glandular trophoblast cells
H2-Ab1	Histocompatibility 2, class II antigen A, beta 1
Habp4	Hyaluronan binding protein 4
HGSC	High-grade serous ovarian carcinoma
Hopx	HOP homeobox

Igfbp5	Insulin-like growth factor-binding protein 5
Il1b	Interleukin 1 beta
ISH	*In situ* hybridization
ISS	*In situ* sequencing
KO	Knockout
KRT10	Keratin 10
L1TD1	LINE1-type transposase domain containing 1
LCM	Laser capture microdissection
Lcn2	Lipocalin 2
LGR5	Leucine-rich repeat-containing G protein-coupled receptor 5
Malat1	Metastasis-associated lung adenocarcinoma transcript 1
MD	Mesometrial decidua
MERFISH	Multiplexed error-robust fluorescence *in situ* hybridization
MMP11	Matrix metallopeptidase 11
MMy	Mesometrial myometrium
Mt3	Metallothionein 3
myCAFs	Cancer-associated myofibroblasts
NGS	Next-generation sequencing
NOTCH2	Notch receptor 2
NR2F2	Nuclear receptor subfamily 2 group F member 2
Pdyn	Prodynorphin
PDZ	Primary decidual zone
PIFO	Primary cilia formation
PIWIL4	Piwi-like RNA-mediated gene silencing 4
PR	Poor responder
RCA	Rolling circle amplification
SCGB2A2	Secretoglobin family 2A member 2
scRNA-seq	Single-cell RNA sequencing
SDZ	Secondary decidual zone
seqFISH	Sequential single-molecule fluorescence *in situ* hybridization
Slc5a7	Solute carrier family 5, member 7
Smcp	Sperm mitochondria-associated cysteine-rich protein
smFISH	Single-molecule fluorescence *in situ* hybridization
SOX9	SRY-box transcription factor 9
ST	Spatial transcriptomics
STARmap	Spatially resolved transcript amplicon readout mapping
Stereo-seq	Spatial enhanced resolution omics-sequencing
Sult1d1	Sulfotransferase family 1D, member 1
TDZ	Transition decidual zone
TG	Trophoblast giant cells
TPPP3	Tubulin polymerization promoting protein family member 3
udStr	Undifferentiated stroma
Vsnl1	Visinin-like 1
VSZ	Vascular sinus zone
WNT7A	Wnt family member 7A
WT	Wild type