BioEssays
www.bioessays-journal.com

Black Swans of CRISPR: Stochasticity and Complexity of Genetic Regulation

Kang Hao Cheong, Jin Ming Koh, and Michael C. Jones*

1. Factoring in Known Unknowns

The social and philosophical sciences need to be involved in a deeper way than before in determining how and when to utilize biotechnologies.[1] The modern rise of biomedical technology has reached an explosive pace: headlines are made on an almost daily basis, typically surrounding sensational developments that are both technically promising and profoundly alarming—this includes gene manipulation techniques showcased in the elimination of an HIV susceptibility locus in human embryos, that were born in China; the cloning of gene-edited monkeys to aid in human genetic and psychiatric research, as reported in *National Science Review*; and a recent article in *Nature* describing precise CRISPR/Cas gene editing "blueprints" for human therapeutics potentially free from off-target errors.[2]

Aside from human gene editing, headlines also augur potentially massive applications in manipulating plant breeding, the control of insect pests, and the suppression of human disease vectors, amongst whose side effects could conceivably be the alteration of ecosystems in unforeseeable ways. The world stands at the brink of unlocking the immense potential of genetic biotechnology—but it is precisely at this juncture that caution becomes imperative. We quote a warning from Freeman Dyson,[3] that "in the near future, we will be in possession of genetic engineering technology which allows us to move genes precisely and massively from one species to another. Careless or commercially driven use of this technology could make the concept of species meaningless, mixing up populations and mating systems so that much of the individuality of species would be lost."

It appears, in fact, that recent waves of ethical, philosophical, and technical controversies surrounding genetic engineering have overtaken the pace of traditional scientific correspondence, and spilled into the realm of popular science in Twitter battles between reputable scientists and their followers; in these public arenas, rampant allegations of "bandwagon" publication bias driven by vested interests in profitable applications clash wildly against ideological opposition from skeptics and watchdogs. For instance, probability theorist and ethicist N. N. Taleb recently retweeted a series of five tweets on the same day as complexity theorist Joe Norman (February 5 2019, https://twitter.com/normonics/status/1092827181263265792)[4] (see **Figure 1**), warning that "the ability to construct arbitrary sequences and consequent behaviors at the microscopic levels actually EXACERBATES [sic] our uncertainty at macroscopic scales, e.g., in ecosystems... Thus, what is sold as precision is the opposite!"

Pithy as a tweet may be, we wish to expand on the warning of the scale-dependent paradox of "precision." The fact remains that we do not fully understand the risks of gene editing on in vivo human tissue even when it is on-target (precise), and we know even less about the long-term effects of precisely edited genomes on animal lineages, including that of humans. Nor can we confidently claim comprehension of the interactive effects between multiple artificially edited loci within developing organisms, much less propagative and evolutionary implications for the higher-order ecosystems of which they are part. Tampering with mortality modes, physiological traits, or life cycles may disrupt intricate eco-evolutionary "optimizational" measures[5] critical to the viability and persistence of natural life—such consequences could take several generations to assess.

2. Machines Running on Stochastic Fuel

The grave danger is that we can develop a flawlessly precise on-target gene editing technology,[2] yet have a flawed, imprecise understanding of what perfectly edited genes will do when they are embedded, like all genes, in complex, pleiotropic gene regulatory networks (GRNs). It must be noted that although the structure of GRNs is specified by DNA sequences, their developmental or phenotypic actions (outputs) are, by nature of the coupled molecular pathways involved, emergent phenomena that cannot be predicted solely from the underlying genetic material. It is the task of GRNs to filter and tune the inherent stochasticity of gene expression into adaptive, viable outcomes at all levels of organismal growth, from cell differentiation to macroscopic events, such as metamorphosis and polyphenic development; but the complexity of vertebrate development presents a mathematically intractable network of interacting mechanics that eludes comprehensive description and treatment with current methods. Such a dense structure of networks can effectively be seen as a massive macroscopic GRN (mGRN) governing life processes at the organismal level, programmed into manifestation by the genome.

Deciphering the simplest organizing principles of stochastic GRNs begins with relatively low-dimensional systems, such as basic bipotential cell fate decisions. An example is the control of differentiation of bipotential neuromesodermal progenitors

Asst. Prof. K. H. Cheong, J. M. Koh
Science and Math Cluster
Singapore University of Technology and Design
8 Somapah Road
Singapore 487372, Singapore
E-mail: kanghao_cheong@sutd.edu.sg

M. C. Jones
PO Box 432
Ashland, Columbia, MO 65010, USA

Also see the video abstract here https://youtu.be/Fpm4ax7AUyY.

DOI: 10.1002/bies.201900032

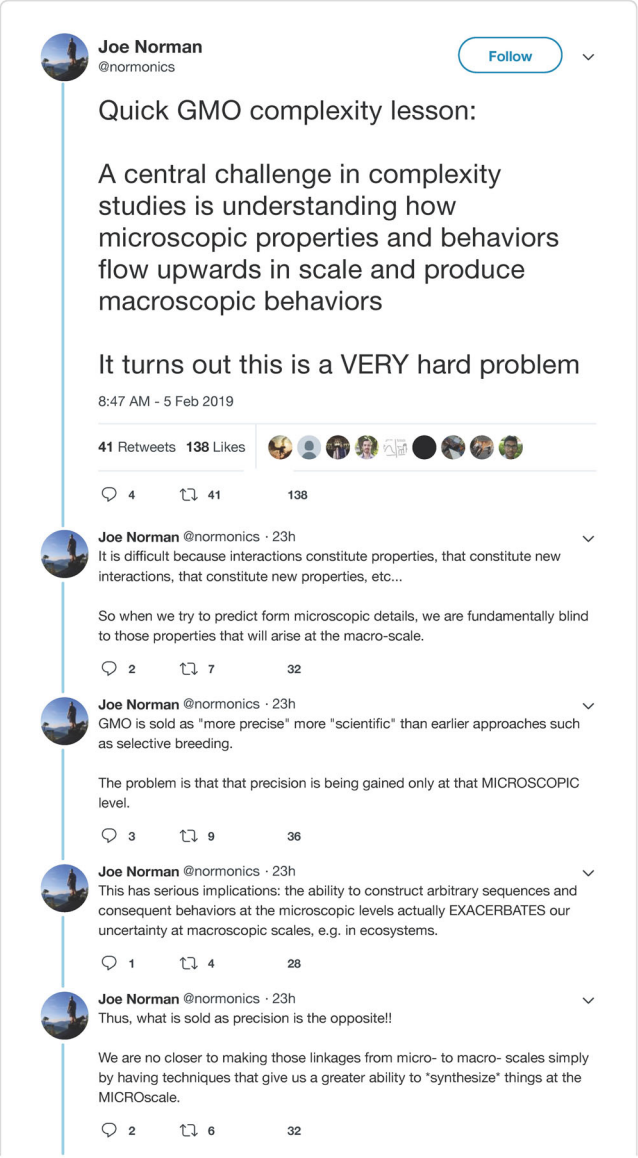

Figure 1. Series of tweets from Joe Norman, warning of the complexity of genetic editing.

(NMPs) that fuel vertebrate embryo elongation through the generation of spinal cord and trunk mesoderm tissue.[6] In this system, the stochasticity of gene expression is entrained within a gene transcriptional network that is feedback-regulated by extracellular signals between proliferating cells, enabling control over cell fate decisions and the balancing of size and composition of developing tissues arising from the bipotential NMPs. In effect, a feedback-modulated noisy bistable switch allows the balanced generation of two cell types in unequal but adaptive proportions from a bipotential progenitor, thereby facilitating orderly elongation of the embryo.[5] Inherent, tunable bi- or polystability, or switch-like behavior between attractors in the state space landscapes of noisy gene networks, is now increasingly recognized as a widespread, if not ubiquitous, property of GRNs across scales.[7]

The variety of possible feedback and feedforward loops within the cellular mGRN (in the form of GRN motifs and sub-circuits, for instance) that could evolve to adaptively shape the distribution of noise is limited by the imagination; but in general, the more complex the organism and the more variable its environmental niche, the greater the extent to which development depends on intricate and sophisticated pathways for tuning the ubiquitous force of biochemical noise toward biological viability. Such tuning may expectedly comprise some combination of amplifying, damping, or splitting of stochastic processes in pathways; the ultimate goals of these processes may be controlling anything from the average number of molecules of neurotransmitter released from a synaptic bouton in response to an action potential, to the macroscopic vagaries of heterochrony in the evolution of metamorphosis and other life history traits. The function of the mGRN is thus understood, but the complexity of its internal mechanisms is stupendous: for instance, phenomena of signal transduction cascades and network crosstalk, and the mixed dynamics of chromatin remodeling, transcription factor binding, cis-regulatory sequences, alternative mRNA splicing—amongst myriad others—will all act in parallel.[8]

This perspective confers a picture of the entire apparatus of differential gene expression regulation (the mGRN) within cells as a noise-filtering and tuning device that allows an adaptive spectrum of graded, discrete and multi-stable responses to stimuli, potentially also of stochastic nature as in environmental conditions. The recent first-ever genome-wide evolutionary and systemic study of transcriptional noise[9] indeed suggests that selection at the pathway (or network) level is widespread in driving the evolution of gene expression at the single-gene scale. Stochastic noise in transcriptional pathways, due to molecular binding and diffusion processes, generates a cascade leading to the synthesis of a protein from its encoding gene; the assembly of these pathways into regulatory networks of increasing scale ultimately produces the mGRN, mapping to the totality of the expressed phenotype. In multicellular organisms, the constructive control of gene expression noise is an emergent organismal-level property, or a covariance structure, of the entire population of cells making up the organism throughout the full trajectory of development (life history) as the organism interacts with its environment (see the paper by Noble and Noble[10] and references therein). That is, the mGRN is not a deterministic developmental machine but rather one that "runs on stochastic processes and finds ways to channel stochasticity."[11]

Thus, the tide of studies culminates in an unfortunate deduction, namely that gene expression is a complex product of noise distributed across inherently stochastic regulatory networks, and therefore almost impossible to describe or to predict fully. Many different combinations of noise in upstream elements of regulation could sum to the same effect, just as a stochastic change in a single element can lead to an array of different downstream effects; and tracing the sources of noise and the propagation of their effects is made even more difficult due to the presence of feedback mechanics. This also implies that mutations, even at the level of single genes, can alter network-level noise control with unpredictable consequences for phenotypic expression. Such alterations may happen not only in the affected cell but potentially across other cells of the

same lineage or in interacting lineages, courtesy of higher-order regulatory components spanning intercellular scales. As alluded to in the emergent nature of the mGRN schematic and its covariance structural characteristics, these consequences may be observably manifest over an immensely wide spectrum of scales, spanning cellular, tissue, organ, organismal, and potentially populational levels—possibly over long timescales.

This is especially concerning, because a large proportion of genetic loci discovered so far to be associated with risk variance in complex traits or diseases belong to the thousands of non-protein-coding genes that influence expression in ways we are far from understanding. Although the effects of noise across scales and their pleiotropic, multistationary effects are impossible for us to calculate ad hoc, the process of evolution has already performed the computations to some degree with the trudging abacus of natural selection—otherwise biological life would not have persisted. It is therefore very reasonably assumed that the genetic structures we see today are already remarkably optimized and stabilized. Negligently thinking that we can improve on the current status quo could be hubris with disastrous consequences. The cascading effects of edited genetics at the microscopic level will not necessarily stop at the mesoscopic-organismal level; they can be transmitted further to affect macroscopic demographics and the response of populations to selection. The problems are further entangled if there are multiple species bearing edits across different genes in interlocking ecosystems, where macroscopic dynamics are elevated to be as important as microscopic ones.

3. Caution in the Face of Uncertainty

The aforementioned concerns may seem to be far-fetched, but if history teaches us one thing, it is that the future reality always demands an accommodation of our expectations—and not the other way around. One can imagine, for instance, a near-future driven by profits, hope, hype, and heaps of research investment capital, in which it becomes possible to present parents with a buffet of genes to edit embryonically. This "pick-and-choose store" would offer statistically "safer" alleles based on genome-wide association study results of complex diseases, such as schizophrenia, for which over a hundred "significant" loci of small effect have already been claimed at present.[12] The problems of noise control in our own epistatic and pleiotropic interaction landscape may only be discovered after the irreversible arc of development is complete—humans become their own guinea pigs in an irreversible experiment.

With our growing appreciation of the role of noise in gene regulation and the nonlinear dynamics between single genes, genotype, development, and evolution, it is imperative to provide a cautionary perspective on CRISPR and other rapidly advancing biotechnologies. While CRISPR in medicine is undoubtedly worth pursuing, our inability to predict its black swans carries heavy risks. No development in biotechnology more richly deserves stringent application of Taleb et al.'s[13] "precautionary principle," to both its research and its presentation to the public.

Acknowledgements

This work was supported by Singapore University of Technology and Design under Start-up Research Grant (SRG SCI 2019 142).

Conflict of Interest

The authors declare no conflict of interest.

Keywords

biotechnology, CRISPR, gene regulatory networks, genetic editing, noise, nonlinear dynamics, stochasticity

Received: February 14, 2019
Revised: February 21, 2019
Published online:

[1] A. Moore, *BioEssays* **2019**, *41*(*1*), 1800242.
[2] P. Akcakaya, M. L. Bobbin, J. A. Guo, J. Malagon-Lopez, K. Clement, S. P. Garcia, M. D. Fellows, M. J. Porritt, M. A. Firth, A. Carreras, T. Baccega, F. Seeliger, M. Bjursell, S. Q. Tsai, N. T. Nguyen, R. Nitsch, L. M. Mayr, L. Pinello, M. Bohlooly-Y, M. J. Aryee, M. Maresca, J. K. Joung, *Nature* **2018**, *561*, 416.
[3] F. Dyson, **2019**, https://www.edge.org/conversation/freeman_dyson-biological-and-cultural-evolution (accessed: February 2019).
[4] Tweet thread by J. Norman, https://twitter.com/normonics/status/1092827181263265792 (accessed: February 2019).
[5] K. H. Cheong, J. M. Koh, M. C. Jones, *Proc. Natl. Acad. Sci.* **2018**, *115*, E5258.
[6] M. Gouti, J. Delile, D. Stamataki, F. J. Wymeersch, Y. Huang, J. Kleinjung, V. Wilson, J. Briscoe, *Dev. Cell* **2017**, *41*, 243 e7.
[7] P. S. Stumpf, R. C. G. Smith, M. Lenz, A. Schuppert, F.-J. Müller, A. Babtie, T. E. Chan, M. P. H. Stumpf, C. P. Please, S. D. Howison, F. Arai, B. D. MacArthur, *Cell Syst* **2017**, *5*, 268 e7.
[8] G. Innocentini, A. Hodgkinson, O. Radulescu, *Frontiers in Physics* **2018**, *6*, 46.
[9] G. V. Barroso, N. Puzovic, J. Y. Dutheil, *Genetics* **2018**, *208*, 173.
[10] R. Noble, D. Noble, *Chaos An Interdiscip. J. Nonlinear Sci* **2018**, *28*, 106309.
[11] E. V. Koonin, *Phys. Life Rev.* **2013**, *10*, 341.
[12] Schizophrenia Working Group of the Psychiatric Genomics Consortium, *Nature* **2014**, *511*(*7510*), 4210028-0836.
[13] N. N. Taleb, R. Read, R. Douady, J. Norman, Y. Bar-Yam, arXiv.org. **2014**. (arXiv:1410.5787).

