GEN Biotechnology
Volume 2, Number 6, 2023

DOI: 10.1089/genbio.2023.29125.awo

GEN Biotechnology

VIEWS & NEWS

Defining the Rules of Biology in the Age of Genome Writing

Adrian Woolfson*

In a series of new papers published in* Cell *and* Cell Genomics*, an international consortium has generated synthetic versions of all 16 yeast chromosomes as well as 1 synthetic neochromosome. Half of these synthetic chromosomes were incorporated into cells that can survive and reproduce without any significant growth deficit. The synthetic yeast genome project (Sc2.0) lays the foundation for a Hipparchian map of biology.

Synthetic organisms generated on a personal computer using rational design rules.
(Image generated by the author using DALL-E, inspired by a DALL-E image produced by Andrew Hessel and released on "X".)

Replay Holdings, Inc., San Diego, California, USA.

*Address correspondence to: *Adrian Woolfson, Replay Holdings, Inc., 5555 Oberlin Drive, San Diego, CA 92121, USA,* E-mail: awoolfson@replay.bio

Around 300 years prior to the publication of Claudius Ptolemy's Algamest,[1] in the second century AD, the Greek astronomer Hipparchus assembled the first ever compendium of the night sky.[2] This mathematical map of the cosmos detailed, for the first time, the position of nearly every star and heavenly body visible to the naked eye. The Hipparchian map is of key importance to the history of science, as it marks the beginning of the "mathematization" of nature.

Through fusing Babylonian celestial modeling with Greek geometry, Hipparchus transformed astronomy from a descriptive science into a predictive one. Despite the compelling precedent, as yet there are no universal biological laws reminiscent, for example, of Newton's law of motion or Carnot's law of thermodynamics. Although replete with data, biology lacks a unified predictive framework with which to comprehend it. This situation, however, may be about to change.

The Synthetic Yeast Genome Project (Sc2.0)

In a landmark achievement marking the inception of a new generative age of genome design and writing, the international Sc2.0 consortium initiated in 2010 by Jef Boeke at NYU Langone Health, Patrick Cai at the University of Manchester, and colleagues in China, Singapore, and Australia, announced that they had finished the complete "bottom up" synthesis, assembly, and debugging of all 16 of the human-designed budding yeast *Saccharomyces cerevisiae* chromosomes.[3–12]

In addition to the 16 synthetic rewritten chromosome analogs whose design is based on their naturally occurring counterparts, the team also manufactured a completely artificial 17th neochromosome, for which there is no precedent in nature.[8] To date, 7.5 of the synthetic chromosomes, representing around 54% of the 12-Mb genome of yeast, have been assembled into a single artificial organism.

The importance of this achievement in the context of the history of life on Earth cannot be overstated. It represents the closing of a circle that enables a product of human culture, the ability to design and write genomes, to feed back into the genetic structures of organisms, and reconfigure them in a way that goes far beyond anything possible in nature (Fig. 1). Following 3.6 billion years of incremental building upon life's ancient genetic history, it is finally possible to entertain the possibility of systematically redesigning existing species, resurrecting extinct ones, and constructing entirely new ones from first principles. It establishes the foundation of a new science of "generative biology."

Each of the 16 natural eukaryotic *S. cerevisiae* chromosomes was redesigned *in silico* using an open-source tool called BioStudio, according to a prescribed set of iteratively applied design principles.[13] These may broadly be grouped into streamlining, defragmentation/refactoring, and the introduction of novel features. The streamlining design features included (1) removal of transposons, and (2) nonessential introns, as well as (3) telomere standardization. Defragmentation/refactoring was achieved by (4) relocating all tRNA genes onto a 17th artificial neochromosome to increase genomic stability. Novel introduced features included (5) an on-demand inducible evolution system known as synthetic chromosome recombination and modification by LoxPsym-mediated evolution (SCRaMbLE). Following Cre-recombinase induction combinatorial chromosomal rearrangements generate tremendous genetic diversity. Others were (6) the swapping of TAG codons for TAA to provide the potential to incorporate noncanonical amino acids into proteins, (7) the introduction of a PCRtag watermark system, and (8) installing a synthetic

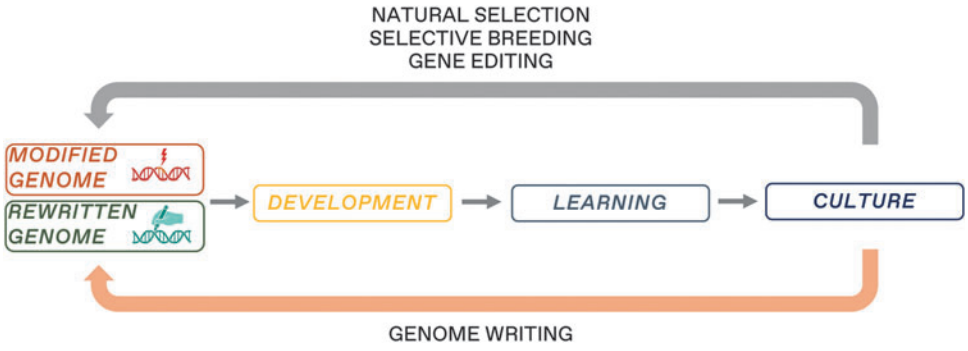

FIG. 1. Augmented feedback into genome sequences resulting from genome folding enabled design and genome writing technologies.

The modification of genomes has, since life's origin, been driven by stochastic mutation events caused by copying errors, ultraviolet radiation, and other mechanisms coupled with evolution by natural selection and historical contingency. These canonical natural methods of mutagenesis were subsequently, in our own species, supplemented by natural breeding methods for altering the frequency of genes in populations. Somatic and germline recombinant DNA methods and gene editing technologies such as CRISPR, base editing, prime editing, zinc finger proteins, and TALENs have more recently allowed the sequences of genomes to be modified in a direct and bespoke manner. Such methods are not, however, readily capable of extensive redesign, which has been made possible using genome writing technologies coupled with emerging new methods of genome design.

nutrient-dependent biocontainment system (not yet done, but there are several auxotrophies in the strain used, so it is perhaps debatable whether this feature is necessary). It is notable that the basic genome writing method, SwAP-In that was used to build the whole genome predated the use of CRISPR, and indeed the synthetic chromosomes were generated without the use of CRISPR.

Early Biological Insights from Sc2.0

The key objective of the Sc2.0 project was to generate a reprogrammable yeast with a redesigned genome that recapitulates the wild-type phenotype and strain fitness with high fidelity. The principal measure of this is the presence or absence of growth defects known as "bugs." These may arise as a result either of chromosomal sequence modifications—especially to noncoding regions—or following amalgamation of synthetic chromosomes, which may generate combinatorial bugs. They are largely a consequence of design defects resulting from our current incomplete knowledge of genome function. The team demonstrated that bugs may be efficiently hunted down and corrected through both rational engineering, and adaptive laboratory evolution. As such, the researchers have essentially realized their objective.

The principal result is the astonishing plasticity of the *S. cerevisiae* genome, which turns out to be remarkably resilient to fundamental restructuring and pervasive nucleotide sequence modifications. Of the approximate 250,000 sequence changes made, just 0.1% led to a fitness deficit "bug." This, and other ongoing lessons learned through an iterative Design-Build-Test-Learn process, such as the fact that centromeres are more complicated than we thought, will help establish the ground rules for restructuring and reinventing the genomes of more complex species. The success of the Sc2.0 project, in some ways, also provides a validation of the computational metaphor in biology,[14] which at least in this case, provides the basis of a compelling and coherent operative model. The successful generation of a neochromosome has established the basis of a new science of *de novo in silico* chromosome design and construction.

A striking example of this was the hierarchical engineering of 275 tRNA genes into eight modules in which all the tRNAs are encoded on the same strand. These modules were, furthermore, designed to avoid collisions between actively transcribing RNA polymerase and oncoming replication forks.

Unresolved questions waiting to be determined include: whether all introns and the accompanying spliceosomal machinery can be removed from the genome, whether new species can be generated by genome scrambling, whether new strains can be made to grow faster than existing ones, if translation can be reprogramed using the tRNA neochromosome, and the long-term behavior of the rRNA neochromosome.

While biological systems lack the orthogonality of human-made machines[15] and likely operate through "fuzzy" logic[16] rather than the exclusively Boolean-type processes characteristic of existing synthetic biology approaches, and generate much of their order through the self-organizing processes that emerge as a result of physics and chemistry, they should nevertheless be computable. As such, they must be rule-based, and those rules should be discoverable. Ultimately, it seems reasonable

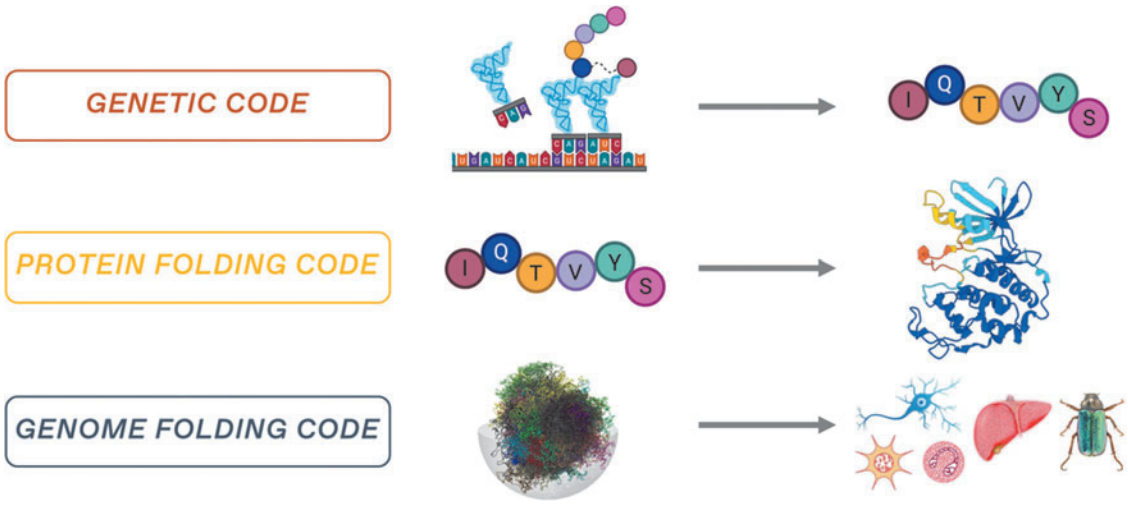

FIG. 2. The three-dimensional genome is envisaged as containing the key information necessary to compute specific differentiated cell types, tissues, organs, and potentially also the morphology, and emergent behaviors of complex organisms.

This information, comprising the integral of all transcriptional regulatory activity in a genome, defines a putative "genome folding code" reminiscent of the genetic code that specifies how linear nucleotide sequences are translated into primary amino acid sequences, and the protein folding code that may be indirectly defined through the use of AI models, such as AlphaFold, that enable many protein structures to be reliably predicted.[17] It is likely that genome folding models derived from primary nucleotide sequences are likely to further benefit from AI models that have been trained using chromosome conformation capture and super-resolution microscopy data. AI, artificial intelligence.

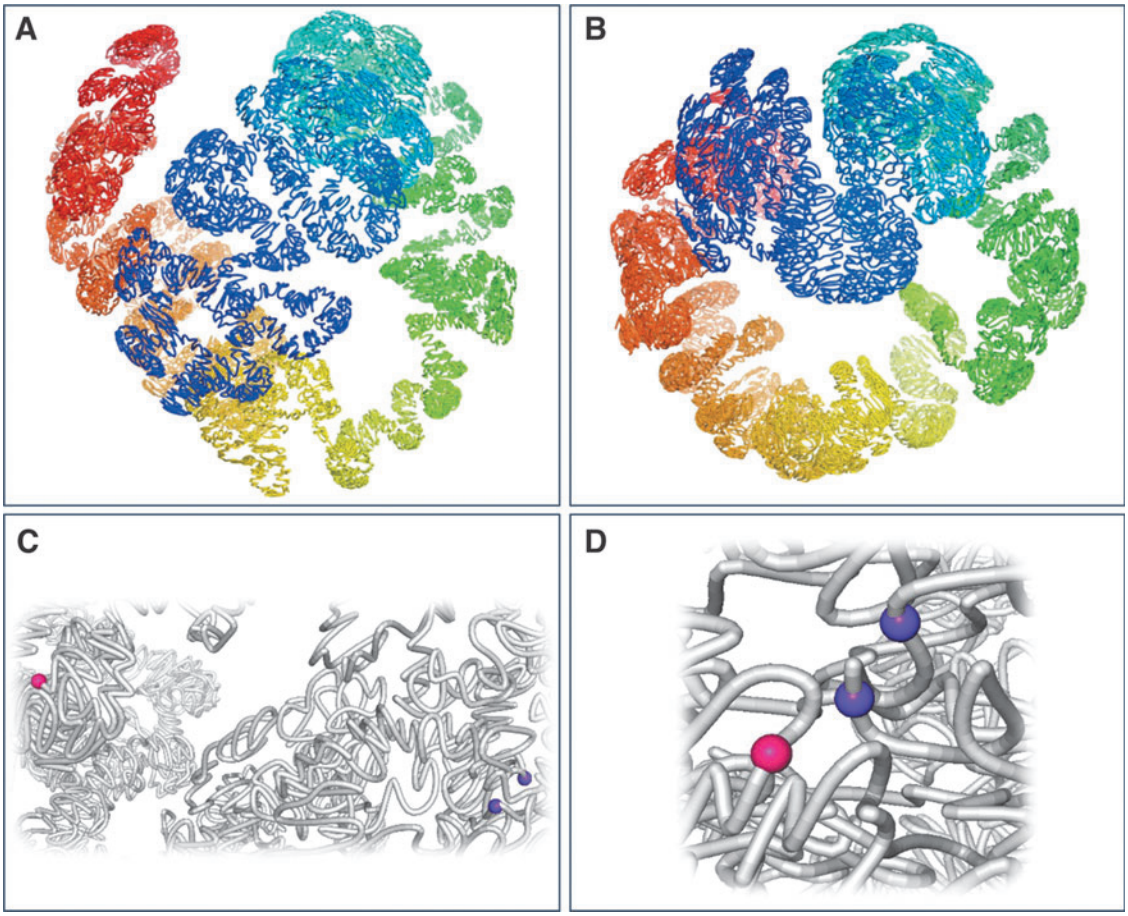

FIG. 3. AI-assisted folding of human chromosome 1 (250 Mb) depicting inferred 3D contact maps in undifferentiated and differentiated cell types.

(A) The chromosome fold in a hESC, showing open chromatin, which is conducive to pluripotency.

(B) The chromosome fold in a differentiated cell where chromatin is regionally condensed.

(C) In hESCs a non-coding Alzheimer's disease-related SNP (magenta) identified by GWAS and two genes (purple) are spatially distant and do not interact.

(D) In a differentiated cell, the same SNP and genes are in close spatial proximity, despite being ~10 Mb apart (a distance almost equivalent to the size of the entire yeast genome) in linear sequence (Bo Jing unpublished data). GWAS, genome-wide association studies; hESC, human embryonic stem cell; SNP, single nucleotide polymorphism.

to hypothesize that the "program" determining complex organismal structure resides in the 3D structure of genomes, namely their "fold." If there are laws of predictive biological design to be discovered, they are, most likely, located there.

A key finding of the Sc2.0 project is that the Hi-C contact maps derived from synthetic chromosomes have much smoother contact patterns than those of native chromosomes. This enhanced 'mappability', likely resulting from the removal of repetitive elements, provides a rationale for the streamlining of human chromosomes to better characterize their folding behavior. It also establishes a paradigm for investigating how the trajectory of chromosomes in nuclear space impacts the core features of living organisms.

The apparent resilience of the yeast genome to substantive changes in chromosomal nucleotide sequence and spatial orientation, including fusing the 16 chromosomes into two giant ones and modifications that literally turn the genome 'inside out',[18–20] raises the question of whether a unicellular eukaryotic organisms like yeast, comprises a suitable model for studying chromatin spatial organization in higher order eukaryotes, in particular humans. While Jacques Monod argued for the generality of biological principles, stating that what it true for the *E. coli* is true for the elephant, it is also the case that a difference in amount may amount to a difference in kind.

As the complexity of genetic regulation increased in phylogeny, the importance of genome spatial configuration likely became more prominent. It seems likely, therefore, that human and other high order genomes will be more sensitive to the types of manipulations to

which the yeast genome was subjected by the Sc2.0 consortium, and it remains to be seen whether the behavior of human chromatin recapitulates the lack of sensitivity to 3D genome reconfiguration observed in yeast.

The complexity of multicellular form and function is likely predicated on the ability to systematically control 3D spatial orientation in a more complex and nuanced manner. In mammals, for example, chromatin is packaged into higher order structures such as topological association domains (TADs), as compared with yeast that lacks well-defined TAD-like intra-nuclear chromatin territories. The potential relevance of global 3D chromatin structure in human genome function in contrast with the predominant role of local structure in yeast, is supported by data implicating its role in both pathological alterations in, for example, acute myeloid leukemia and in natural development processes, such as the facilitation of enhancer specificity in limb morphogenesis.[21,22]

In the same way that AI can predict protein structure without understanding the laws by which this actually happens, it should by analogy be possible to infer core features of complex organismal structure from the study of genome folds (Fig. 2). As is the case with AI-based protein structure prediction, you don't necessarily need to understand the details of the physical and chemical nature of the self-organizing principles of life, to infer and utilize general construction rules to build it.[17]

Understanding the "folding code" of the human genome is likely to be immensely consequential and has the potential to open up countless opportunities to improve human health, longevity, and well-being (Fig. 3). It will also have diagnostic and prognostic utility in disease settings and may predict potential responders to therapeutic interventions. It is possible to anticipate a new type of therapeutic medical science focused on the repair of aberrant genome folds.

Defining a Hipparchian Map of Species and a Manifesto for Life

When Hipparchus looked up at the night sky, he was able to discern just a minute fraction of the stars and planets in the known Universe. Beyond this lay an impenetrable, and unknowable heart of darkness. With the launch of the James Webb Space Telescope (JWST) in 2021,[23] the extent of the visible Universe dramatically expanded. Previously invisible regions of sky were now seen teeming with sparkling galaxies and breathtakingly expansive seas of multicolored galaxies. If, by analogy, the putative generative laws of biology could be defined they would form a type of biological JWST, and make it possible, like Hipparchus, to produce a map detailing a morphospace comprising not just all existing and extinct species, but the incomprehensively vast and uncharted universe of all possible species.[24,25] At a more granular level, and as the resolution increases, it should also be possible to define, for example, the space of all possible cell types, including 'unused' cell types that have never previously been identified or realized. The achievements of the Sc2.0 project are the earliest tentative steps in the creation of this Hipparchian map.

This cartographic project, dependent on a comprehension of life's generative grammar, will eventually impact almost every human endeavor and requires an extensive and inclusive societal debate to define a manifesto for life. This will help calibrate the relationship between the natural and artificial, emphasize the importance of preserving biodiversity, facilitate the establishment of an appropriate governance and regulatory framework, and ensure that this new frontier is explored in a responsible and ethical manner that benefits all humanity.

Acknowledgments

Thanks to Bo Jing for preparing the AI-assisted folding images of human chromosome 1 shown in Figure 3. These previously unpublished images represent to our knowledge the first attempt to use AI to fold a human chromosome from DNA sequence only, made possible by recent developments in GPU technology. Thanks also to Ashley Buckle for assistance compiling Figures 1 and 2 and Andrew Hessel for inspiring the hero image.

Adrian Woolfson *is the Book Review Editor of GEN Biotechnology and Executive Chairman, President, and Co-Founder of Replay Holdings, Inc., a genome engineering biotechnology company developing TCR-NK cell therapies, HSV-based gene therapies, programmable organoids, multiplex transcriptional regulators, AI-assisted genome folding methods, and genome writing technologies. He is a regular contributor to the "Opinions" section of the* Wall Street Journal.

References

1. Toomer GJ. Ptolemy's Almagest. Princeton University Press: Princeton, NJ; 1998.
2. Marchant J. First known map of night sky found hidden in Medieval parchment. Nature 2022;(610):613–614; doi: 10.1038/d41586-022-03296-1
3. Zhao Y, Coelho C, Hughes AL, et al. Debugging and consolidating multiple synthetic chromosomes reveals combinatorial genetic interactions. Cell 2023;186(24):5220–5236.e16; doi: 10.1016/j.cell.2023.09.025
4. Lauer S, Luo J, Lazar-Stefanita L, et al. Context-dependent neocentromere activity in synthetic yeast chromosome VIII. Cell Genomics 2023;3(11):100437; doi: 10.1016/j.xgen.2023.100437
5. McCulloch LH, Sambasivam V, Hughes AL, et al. Consequences of a telomerase-related fitness defect and chromosome substitution technology in yeast synIX strains. Cell Genomics 2023;3(11):100419; doi: 10.1016/j.xgen.2023.100419
6. Shen Y, Gao F, Wang Y, et al. Dissecting aneuploidy phenotypes by constructing Sc2.0 chromosome VII and SCRaMbLEing synthetic disomic yeast. Cell Genomics 2023;3(11):100364; doi: 10.1016/j.xgen.2023.100364
7. Blount BA, Lu X, Driessen MRM, et al. Synthetic yeast chromosome XI design provides a testbed for the study of extrachromosomal circular DNA dynamics. Cell Genomics 2023;3(11):100418; doi: 10.1016/j.xgen.2023.100418
8. Schindler D, Walker RSK, Jiang S, et al. Design, construction, and functional characterization of a tRNA neochromosome in yeast. Cell 2023;186(24):5237–5253.e22; doi: 10.1016/j.cell.2023.10.015
9. Williams TC, Kroukamp H, Xu X, et al. Parallel laboratory evolution and rational debugging reveal genomic plasticity to S. cerevisiae synthetic chromosome XIV defects. Cell Genomics 2023;3(11):100379; doi: 10.1016/j.xgen.2023.100379
10. Foo JL, Kitano S, Susanto AV, et al. Establishing chromosomal design-build-test-learn through a synthetic chromosome and its combinatorial reconfiguration. Cell Genomics 2023;3(11):100435; doi: 10.1016/j.xgen.2023.100435
11. Luo J, Vale-Silva LA, Raghavan AR, et al. Synthetic chromosome fusion: Effects on mitotic and meiotic genome structure and function. Cell Genomics 2023;3(11):100439; doi: 10.1016/j.xgen.2023.100439
12. Zhang W, Lazar-Stefanita L, Yamashita H, et al. Manipulating the 3D organization of the largest synthetic yeast chromosome. Mol Cell 2023; In Press; doi: 10.1016/j.molcel.2023.10.015

13. Richardson SM, Mitchell LA, Stracquadanio G, et al. Design of a synthetic yeast genome. Science (80-) 2017;355(6329):1040–1044; doi: 10.1126/science.aaf4557
14. Nicholson DJ. Is the cell really a machine? J Theor Biol 2019;477:108–126; doi: 10.1016/j.jtbi.2019.06.002
15. Porcar M, Peretó J. Nature versus design: Synthetic biology or how to build a biological nonmachine. Integr Biol 2016;8(4):451–455; doi: 10.1039/C5IB00239G
16. Liu H, Jeffery CJ. Moonlighting proteins in the fuzzy logic of cellular metabolism. Molecules 2020; 25(15):3440; doi: 10.3390/molecules25153440
17. David A, Islam S, Tankhilevich E, et al. The AlphaFold Database of protein structures: A biologist's guide. J Mol Biol 2022;434(2): 167336; doi: 10.1016/j.jmb.2021.167336
18. Fudenberg G, Getz G, Meyerson M, et al. High order chromatin architecture shapes the landscape of chromosomal alterations in cancer. Nat Biotechnol 2011;29:1109–1113; doi: 10.1038/nbt.2049
19. Hnisz D, Weintraub AS, Day DS, et al. Activation of proto-oncogenes by disruption of chromosome neighborhoods. Science 2016;351:1454–1458; doi: 10.1126/science.aad9024
20. Lupiáñez DG, Kraft K, Heinrich V, et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell 2015;161: 1012–1025; doi: 10.1016/j.cell.2015.04.004
21. Xu J, Song F, Lyu H, et al. Subtype-specific 3D genome alteration in acute myeloid leukaemia. Nature 2022;611(7935):387–398; doi: 10.1038/s41586-022-05365-x
22. Kragesteen BK, Spielmann M, Paliou C, et al. Dynamic 3D chromatin architecture contributes to enhancer specificity and limb morphogenesis. Nat Genet 2018;50(10):1463–1473; doi: 10.1038/s41588-018-0221-x
23. Stofan E. Hubble's successor, at last. Science (80-) 2021;374(6572):1175; doi: 10.1126/science.abn3947
24. Ollé-Vila A, Duran-Nebreda S, Conde-Pueyo N, et al. A morphospace for synthetic organs and organoids: the possible and the actual. Integr Biol (Camb) 2016;8(4):485–503; doi: 10.1039/c5ib00324e
25. Budd GE. Morphospace. Curr Biol 2021;31(19): R1181–R1185; doi: 10.1016/j.cub.2021.08.040