The challenges for synthetic biology in eukaryotes

Francesca Ceroni[1,2] and Tom Ellis[2,3,4]

1. Department of Chemical Engineering, Imperial College London, London, UK
2. Imperial College Centre for Synthetic Biology, Imperial College London, London, UK
3. Department of Bioengineering, Imperial College London, London, UK
4. Visiting Faculty, Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, USA

Correspondence should be addressed to t.ellis@imperial.ac.uk

Standfirst (50 words)

Synthetic biology is maturing into a true engineering discipline for model microbes, but remains far from straightforward for most eukaryotes. Here we outline the key challenges facing those trying to engineer biology across eukaryota and suggest areas of focus that will aid future progress.

Main text

Current stars of synthetic biology

Synthetic biology has expanded from a field of a few groups focused on engineering small genetic systems, to a major part of bioengineering[1], where researchers undertake international genome projects[2] and launch high value companies. Most of the foundational work that can be justifiably termed bioengineering has been done in bacteria, especially in *Escherichia coli*, which can now count, calculate, sense, see, respond to and do a plethora of interesting and valuable things. This has been driven by *E. coli* being well-studied and exceptionally amenable to genetic engineering. Yet compared to the possibilities offered by most organisms, *E. coli* is frankly quite boring.

For eukaryote synthetic biology, the pace is being set by the yeast *Saccharomyces cerevisiae*, which like *E. coli*, is well characterized, grows quickly and is easy to stably transform. As such it is well-suited for projects that require advanced genetic engineering such as for metabolic biosynthesis or for genome redesign[2]. But *S. cerevisiae* is not really representative of eukaryotes, especially when engineering DNA: it only has a minimal level of chromatin, it lacks RNAi, has no alternative splicing and rarely undergoes any type of cell differentiation. It is effectively a ‘eukaryote lite’, and along with

E. coli, is effectively the starter kit for biological engineers. Most eukaryotic cells we study are much more challenging to engineer, with more layers of regulation and the differentiation and specialism that comes from multicellularity. Yet it is precisely the specialist nature of most eukaryotic cells that makes them attractive for synthetic biologists working towards applications beyond biosynthesis. When we also consider that we and most of what we eat are eukaryotes, it becomes obvious why we seek to advance synthetic biology in these organisms for energy, agricultural and medical applications.

Diversity and multicellularity

So what are the main challenges holding back the engineering of eukaryotes from filamentous fungi through to worms, mosses, flies, trees, fish, bananas, humans, diatoms and slugs? For a start it is the two things that attract us to them the most: their diversity and their multicellularity. Their fascinating diversity means that research efforts rarely focus on a single model organism or cell line and so can't represent whole kingdoms, phyla or classes. We know a little about genetically manipulating many different organisms, while doing complex synthetic biology requires deeper knowledge and new tools. Perhaps predictable engineering for eukaryotes could be accelerated if everyone agreed to work only on a handful of model organisms such as mouse, fruit flies and thale cress. This would certainly improve fundamental capabilities for these, but it is arguable whether it would be widely beneficial, as it is unlikely to be straightforward to transfer knowledge and tools from model organisms to their more readily-applied relatives, such as humans, mosquitos and crops.

Multicellularity is an exciting property to exploit, but is challenging, especially for mammalian synthetic biology where many different cell types could be engineered. For example, do we know enough about each cell type to be able to engineer a genetic system that will function similarly in all the cells of an organism, or, alternatively only operate in a specific subset? Our understanding of tissue-specific tools, such as transcription factors, is currently limiting. To make headway we typically perform foundational work in quick-to-grow and easy-to-transform cell lines. Yet these cells and their growth conditions are often not very representative of cells in their origin organisms, as they are typically tumor-derived. This also means they are more prone to genomic rearrangements, mutations and epigenetic changes which all make robust genome engineering much more of a challenge. The increased availability of induced pluripotent stem cells will hopefully help solve some of these issues and should provide a new focus point, particularly for expanding human synthetic biology and for applications like tissue and organoid generation and utilization.

The challenges of eukaryote DNA

Part of the reason why cell lines are popular is because it is difficult to introduce DNA constructs into most primary eukaryotic cells. Indeed, it would be fair to say that nothing holds back eukaryote

synthetic biology more than the poor transformation rates of most cells and organisms. Factor-in their slow growth, such that it often takes months to stably integrate a multigene system into the genome of any eukaryote, and yet doing this is essential to avoid the short-term performance and high variability of transient transfections. For many eukaryote viruses are emerging as the workhorses for reliable DNA delivery. Future efforts to engineer modular viruses that efficiently deliver large DNA constructs and can be reprogrammed to target many hosts would greatly benefit synthetic biology.

Placing DNA accurately into the cell's genome is also a challenge. CRISPR-based tools and cell lines 'pre-engineered' to contain *landing pads* that promote site-specific integration into reliable chromosomal regions are now aiding this. Artificial chromosomes provide a promising alternative that can be used to host multiple synthetic constructs without disrupting the native genome. Although human artificial chromosomes (HACs) and their equivalents have been around for decades, synthetic biology efforts should now spur the construction of a new generation made not just to be reliable as a long-term vector but also designed for inherent portability into other organisms[3]. Community efforts to establish modular 'neochromosomes' for different branches of eukaryota will be of great value, and hopefully will accelerate the use of standardized DNA modules and tools.

Modular programming at the DNA level is well-established in synthetic biology, but the further layers of chromatin, alternative splicing and post-translation modifications adds many different mechanisms for gene regulation and enforcing feedback beyond the basic transcriptional regulation programs used in bacteria and yeast. Our current approach for the design of genetic circuits is largely based on the synthesis of bare plasmids where the DNA is written to encode the functions. But what if in addition to DNA we could write chromatin in modules? Genes and pathways could be packaged into chromatin-like structures that could be activated or silenced using just small changes in a few chromatin remodeling factors. This would allow for more constructs to be co-regulated and efforts in this area could exploit the recent development of synthetic and orthogonal chromatin remodeling factors[4].

Opportunities for progress

Orthogonality - separating the engineered systems from those of their host cell - could also be pursued by taking advantage of subcellular compartmentalisation, a natural feature of eukaryotes. We can take inspiration from mitochondria and chloroplasts, which operate their own genetic programs, and from organelles like peroxisomes that perform isolated metabolism. Synthetic organelles could be designed to run gene expression programs, protein-protein interactions or metabolic pathways in ways that intentionally minimise the interactions with the host biology. This would greatly simplify the many considerations currently required when engineering within a cell's highly-connected systems.

In bacteria, the challenges of engineering with consideration to the host is being aided by the increased use of transcriptomics and whole-cell modelling approaches, which are improving the characterisation of synthetic constructs and enabling prediction of how they perform dynamically in a growing host cell. This should also be a viable future strategy for most eukaryotes, given the explosion in 'omics' data now available for many organisms, especially for the main mammalian models. Ultimately, the more knowledge we have on the biology of the host and how this is altered by synthetic engineering, then the faster progress will be for the field. The ultimate goal of being able to design and build a full custom synthetic genome for a mammalian cell[5] will require significant technological advances, but also much more understanding about eukaryotic biology in general increased communication between those working to study cells and those trying to build them.

Acknowledgements

For valubale input we thank Karen Polizzi, Karmella Haynes, Devang Mehta, Michael Elowitz and all of those who voiced opinions on this subject via Twitter. FC is supported by the Imperial College London Junior Research Fellowship Scheme.

Competing interests

The authors declare no competing interests.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Cameron, D.E. *et al.* A brief history of synthetic biology. *Nature Reviews Microbiology* **12**, 381-390 (2014).

2. Richardson, S.M. *et al.* Design of a synthetic yeast genome. *Science* **355**, 1040-1044 (2017).

3. Martella, A. *et al.* Mammalian Synthetic Biology: Time for Big MACs. *ACS Synthetic Biology* **5**, 1040-1049 (2016).

4. Neyer, D.B. *et al*. Regulation of cancer epigenomes with a histone-binding synthetic transcription factor. *NPJ Genomic Medicine* **2**, 1 (2017)

5. Boeke, J.D. *et al*. The Genome Project–Write. *Science* **353**, 126-127 (2016).