SYNTHETIC BIOLOGY

Synthetic Biology, 2021, **6(1)**, 1–5
DOI: https://doi.org/10.1093/synbio/ysab015
Advance access publication date: 25 June 2021
Perspectives

Building Biofoundry India: challenges and path forward

Binay Panda* and Pawan K. Dhar

Biofoundry India, School of Biotechnology, Jawaharlal Nehru University, New Delhi, India
*Corresponding author: E-mail: binaypanda@jnu.ac.in

Abstract

Biofoundry is a place where biomanufacturing meets automation. The highly modular structure of a biofoundry helps accelerate the design–build–test–learn workflow to deliver products fast and in a streamlined fashion. In this perspective, we describe our efforts to build Biofoundry India, where we see the facility add a substantial value in supporting research, innovation and entrepreneurship. We describe three key areas of our focus, harnessing the potential of non-expressing parts of the sequenced genomes, using deep learning in pathway reconstruction and synthesising enzymes and metabolites. Toward the end, we describe specific challenges in building such facility in India and the path to mitigate some of those working with the other biofoundries worldwide.

Key words: biofoundry; dark matter of the genome; pathway reconstruction; deep learning

1. Introduction

India is among the handful of nations that realized the potential of biotechnology very early when the country established the Department of Biotechnology within the Ministry of Science and Technology in 1986 to support research and development activities in the sector (1). Since then, the sector has emerged as a key driver for economic growth in the country. Much of it is due to the expansion of scientific institutions, workforce and facilities and the emergence of the 'millennial' and 'zillennial' entrepreneurial class. Taking the bio-economy to the next level will require a much higher level of funding in research and development as a percentage of gross national product. This will pave ways to establishing multiple large facilities for shared access among the academic and industry partners, developing a culture to share an equipment among the institutions and scientists and creating sustainable ecosystems for entrepreneurial-driven growth in the sector. The science of constructing biomolecular components has given rise to synthetic biology or biological engineering, which has emerged as a key vertical within the biotechnology sector to help achieve this goal.

As we build Biofoundry India (BI), our focus will be to get fully integrated into a global alliance of other biofoundries, such as the Global Biofoundries Alliance (GBA) (2, 3), participate in international consortia like the Genome Project-write (4), coordinate with other large facilities to develop and share best practices and promote collaborations. Here, we describe the key areas of focus of BI and some of the challenges that we see in the coming years and provide strategies on how we can overcome those. We also describe how shared facilities like BI can help the country when needed, like during the coronavirus disease 2019 (COVID-19) pandemic, to design and deploy large-scale diagnostics and design vaccine candidates.

2. Focus areas

2.1 The dark matter of genome

Traditionally, scientists have focused on the protein and RNA coding regions of the genomes. The non-expressing sequences, the 'dark matter' of the genomes (Figure 1A), are the least understood regions of the genome and thought to be regulatory in their function. The key question that one of us (P.D.) has been interested in asking is that how did nature determine the insertion points for start and stop signals in such a vast stretch in the genomes of various organisms? Did she sample all the possibilities, retaining good results (e.g. protein and RNA coding genes), retiring bad outcomes (e.g. pseudogenes) and leaving some of the regions unexpressed (the dark matter), perhaps to be employed in exigencies? The significant evolutionary conservation of genes indicates that nature did not pay much attention on making brand new genes. Instead, she largely used the available molecular inventory to create a wide variety of organisms. The assumption that nature did not test every possible genomic combination for protein synthesis in a specific organism makes the vast unexpressed genome available for innovation.

Welcome to Functional genomics 2.0 (FG2.0), the field of genomics that deals with understanding the function of non-expressing DNA by artificially expressing the sequences. The grand challenge of FG2.0 is to catalog, computationally model and experimentally characterize every non-expressing DNA sequence. The non-expressing DNA inventory comprises

Submitted: 12 May 2021; **Received (in revised form):** 27 May 2021; **Accepted:** 24 June 2021

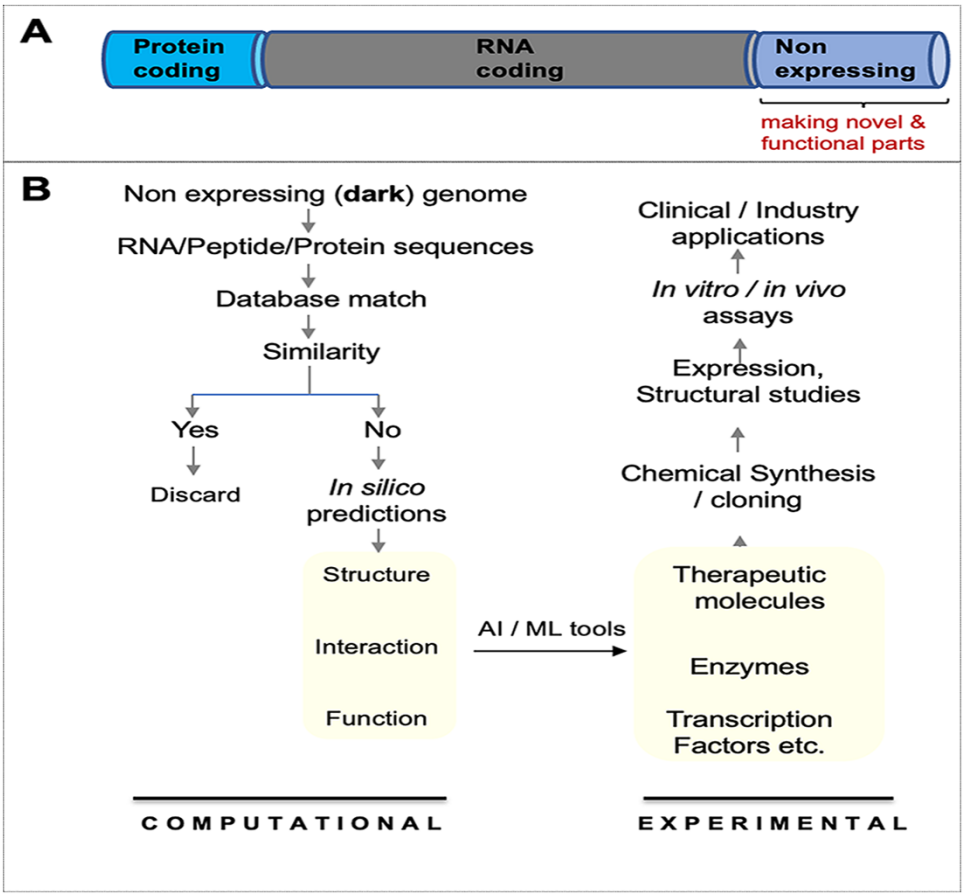

Figure 1. The dark matter of the genome (A, not to scale) and the general approach of making novel biological parts toward various applications (B).

intergenic sequences, introns, pseudogenes, antisense DNA and reverse coding DNA. Furthermore, we also plan to make proteins from repetitive DNA sequences and long non-coding RNA sequences. In 2009, one of our groups provided the proof of the concept of making functional proteins from *Escherichia coli* intergenic sequences (5). This was followed by the feasibility of making functional molecules from such sequences that were evolutionarily unused for RNA and protein expression (6–8). Some of these molecules have pharmaceutical importance, as shown recently with the anti-leishmanial property of a novel first-in-class tRNA-encoded peptide (9). BI will focus on automating the workflow and making processes to explore and exploit this further in sequenced genomes. Additionally, BI will try attracting investment from the industry to make enzymes and especially therapeutic molecules for specific indications, similar to what we have recently demonstrated (9).

2.2 Software and artificial intelligence

One of BI's significant areas of focus is to create predictive software based on deep learning and automate its use alongside existing computational tools to implement the various phases of design–build–test–learn (DBTL) cycle. For example, we are interested in implementing such integration and automation in our proposed research on the dark matter of various genomes (Figure 1B). Additionally, BI plans to create, deploy and provide access to software and systems for sample collection, workflow automation and reproducibility, and data production, storage, distribution and security with federated identity.

Artificial intelligence (AI), as a technology, helps improve biological engineering efficiency, as shown recently with the development of an automated recommendation tool (10). Within this key area of focus for BI, we are interested in developing predictive tools for pathway reconstruction. One of the authors' (BP) lab has sequenced, analyzed and interpreted the genome and organ-specific transcriptomes of a plant well-known for several pharmaceutically and agronomically important secondary metabolites and potential therapeutic targets (11–13). Under this area of focus within BI, building platforms' emphasis will be on integrating systems' biology-based approach in pathway reconstruction using deep learning. We plan to use different levels of data (Figure 2A, high-throughput multi-ome; fluxes and balances in chemical reactions; structures of nucleic acids and proteins; interactions between proteins; stoichiometry in chemical reactions and organ morphology imaging) and pre-processing steps (Figure 2B, curation by pruning, organizing and integrating data from different levels; annotation by discovering genes and their positions; inferring metabolic system dynamics by identifying reaction and enzyme classes) in a deep learning-based prediction system (Figure 2C, neural networks like multi-layer perceptrons for non-image data and convolutional neural networks for image data; generate robust predictive models using train, validate and test processes). To start with, BI will focus on building processes toward minimal pathway construction that will only involve key secondary metabolite production steps. In a proof-of-concept study using convolutional neural networks, we have shown plant part images to predict azadirachtin content class (14). It is often difficult to present a

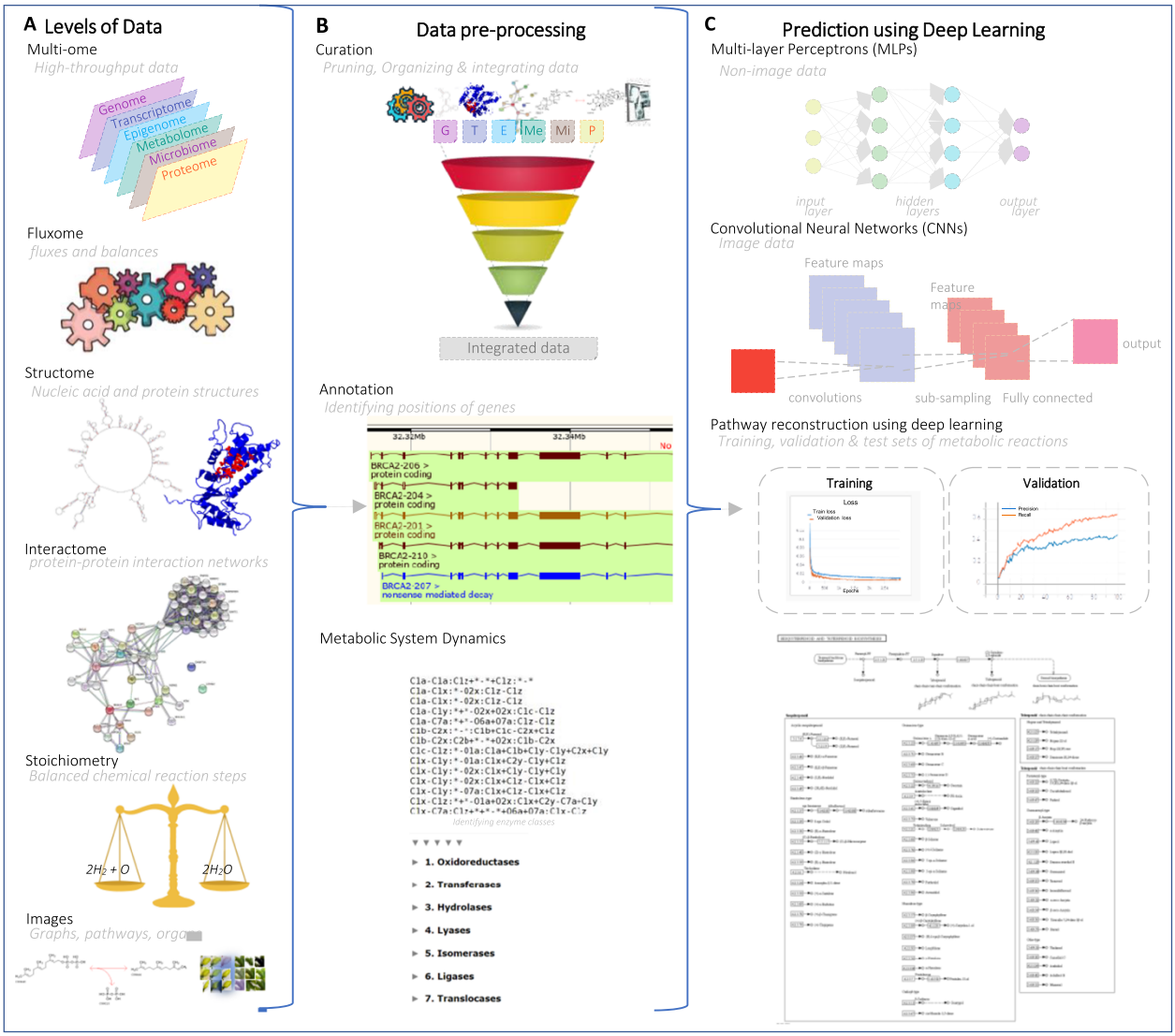

Figure 2. Integrated systems biology data analyses and pathway reconstruction using deep learning.

complete picture of the pathways of interest in organisms with incomplete or partial information on gene expression, annotation, biochemical reactions and interaction networks. In such cases, BI will explore using pathway knowledge and the gene expression information from closely related organisms to simulate the most likely pathways with the missing constituents using AI.

2.3 Bioprocess engineering

The third area of BI focus is designing and developing bioprocesses in industrially established hosts (bacterial, yeast and mammalian cells). Here, the emphasis will be on designing hosts with modified genes toward bioprocess development with improved flux. This will enable the biomanufacturing of commercially relevant biomolecules at higher capacity. However, the overexpression of recombinant proteins may cause metabolic burden by draining the host cell's resources necessary for its own growth and maintenance (15). Typically, the cell responds to this metabolic imbalance by triggering cellular stress response leading to down-regulation of the expression of novel recombinant proteins. In BI, we plan to alleviate the effect of the cellular stress response by identifying the key regulatory molecules responsible for the cellular stress and use this information to systemically engineer the host cell. Some of the critical bioprocess engineering challenges we plan to address in BI are as follows: (a) designing a high-throughput screening strategy to select appropriate host strains, (b) creating synergistic combinations of knock-ins and knock-outs toward optimal over-expression of recombinant proteins in a specific panel of strains and (c) building scale-up models using lab-scale bioreactors. To enable customized and rapid bioprocess development, BI shall use instruments like High Performance Liquid Chromatography (HPLC), Liquid Chromatography–Mass Spectrometry (LC-MS), Fluorescence-Activated Cell Sorting (FACS), spectrometer, recombinant product development facility of Good Laboratory Practice (GLP) standards, bacterial and mammalian cell bioreactors with Fourier-Transform Infrared Spectroscopy (FTIR) analysis, AKTA-Prime, AKTA-Explorer Fast Protein Liquid Chromatography (FPLC) protein purification systems, microscopes and automated liquid handling systems.

3. Challenges

Apart from adequate financial and technical resources, establishing and running the biofoundry will require integrating physical and digital infrastructures and taking care of the organizational

and operational issues (16). In India, past efforts toward making such shared facilities to serve multiple stakeholders beyond the institutions in which the facilities exist have largely failed. This is partly due to the legacy of science practice in India, where the growth of science has been mainly through the efforts of individual laboratories. The presence of a competitive market force driven by private entities providing cost-effective services and the lack of a sustainable business model are other reasons for the past failures. Therefore, adopting a large resource base in a biofoundry will require re-educating the science administrators, granting agencies, institutions and individual scientists. As science becomes more international and large consortia-driven projects become the norm for producing data, tools and methodologies, Indian scientists need to adjust to this new normal quickly and learn to use shared facilities like biofoundries for their individual scientific needs. A booming economy with a large research and development spending in the coming years and India's desire to be a leader in technological innovation will further boost this cause.

If the biofoundry cannot provide cost-effective service, outside users will seldom use it. However, cost alone cannot be the only factor driving its use. BI plans to provide cost-effective services coupled with hard-to-find technical and scientific expertise by opening up collaborative turn-key projects for the industry where it will leverage the physical and digital infrastructure of the biofoundry and the scientific skills and reputation of the academic investigators. BI plans to develop several models for client engagement within the institutional framework and following national guidelines with a long-term goal to having fully operational independence with its own staff to run and execute collaborative projects while prioritizing equipment access among internal institutional faculty and staff. In the Indian research milieu, a 'fee for service' model may not work due to the higher than usual expectation and customers' mindset. Therefore, shared facilities like biofoundries cannot be seen as places for revenue generation only but as a facility funded mainly by government research grants and utilized for achieving shared goals. This, along with investigators'-driven research grants where a part of the grant can be earmarked for new equipment purchase and maintenance to partially support the biofoundry will help toward the long-term sustenance of the facility. Third, India provides distinct advantages in terms of cost, skillset and time to product delivery on specific verticals, like software development and testing and development of optimized data analysis workflows and tools. BI plans to engage with the members of the larger GBA community and provide the necessary skills for the delivery of software and information technology-driven tools and data analysis platforms for global use. Lastly, BI can be the global powerhouse in maintaining, supporting and upgrading software and data analysis tools in all biofoundries. This will require a greater cooperation among all the members of the GBA community and by integrating BI within the larger GBA community to share knowledge, best practices and client services. We believe that the GBA as a community of researchers and facilities can engage better with the international bodies and granting agencies to get large scientific projects and deploy time-sensitive solutions, such as developing high-throughput diagnostic assays for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) detection, globally.

One of the biggest challenges in the Indian scientific grant system is to fund personnel in the biofoundry on a long-term basis and retaining them. Lack of institutional mechanism to pay for the permanent technical positions within the public university system further exacerbates the situation. This can be dealt partially by the funding from BI's collaborative research projects and by providing a defined career path within the facility. In addition, BI plans to support personnel using resources from the individual PI's scientific grants.

4. Future direction

Biofoundries are places for distributed manufacturing of biological parts to accelerate, automate and streamline research and innovation. Although envisioned to help the fields of synthetic biology/biological engineering, the infrastructure within a biofoundry can be quickly repurposed to solve other scientific problems. Like any laboratory with extensive equipment for automation, biofoundries, in addition to requiring substantial initial investment, need a business model for sustenance. This is critical to maintaining instruments, ensuring long-term employment of personnel engaged in running the facility and responsible for data generation and management, and most importantly, making the processes and products compliant with national and international regulations.

The COVID-19 pandemic has caused significant loss of lives and livelihoods worldwide, including India that witnessed a devastating second wave of the pandemic. Never before, the scientific community was challenged to come up with solutions for the masses with minimal time and resources. The pandemic has taught us to repurpose the existing scientific infrastructure and workforce for solving pressing issues, like developing high-throughput diagnostic assays for SARS-CoV-2 as shown recently (17, 18) and designing safe and efficacious vaccine candidates. Vaccine production, especially during a pandemic, will benefit from the systematic workflow approach of biological engineering (19). Biofoundries with appropriate automation and equipment are especially well suited for vaccine designing work that does not require whole cells (20). The future will require the scientific community to quickly use the high-throughput capabilities of the biofoundries and use the DBTL workflow to design, construct and scale for easy and quick deployment of solutions for multiple scientific problems.

Acknowledgments

We thank Paul Freemont and Neeraja M Krishnan for critically reading the manuscript and NMK for helping us with Figure 2.

Data Availability statement

No data was generated and used for the paper.

Conflict of interest statement

The authors declare no conflict of interest.

Funding

The work presented in this article was not supported by any external funding.

References

1. Natesh,S. and Bhan,M.K. (2009) Biotechnology sector in India: strengths, limitations, remedies and outlook. *Curr. Sci. (India)*, **97**, 13.
2. Hillson,N., Caddick,M., Cai,Y., Carrasco,J.A., Chang,M.W., Curach,N.C., Bell,D.J., Le Feuvre,R., Friedman,D.C., Fu,X. *et al.*

(2019) Building a global alliance of biofoundries. *Nat. Commun.*, **10**, 2040.
3. GBA. *Global Biofoundries Alliance*. https://biofoundries.org/ (18 April 2021, date last accessed).
4. Boeke,J.D., Church,G., Hessel,A. *et al.* (2016) Genome engineering. The genome project-write. *Science*, **353**, 126–127.
5. Dhar,P.K., Thwin,C.S., Tun,K., Tsumoto,Y., Maurer-Stroh,S., Eisenhaber,F., Surana,U. (2009) Synthesizing non-natural parts from natural genomic template. *J. Biol. Eng.*, **3**, 2.
6. Shidhi,P.R., Suravajhala,P., Nayeema,A. (2015) Making novel proteins from pseudogenes. *Bioinformatics*, **31**, 33–39.
7. Raj,N., Helen,A., Manoj,N., Harish,G., Thomas,V., Singh,S., Sehrawat,S., Seth,S., Nair,A.S., Grover,A. *et al.* (2015) In silico study of peptide inhibitors against BACE 1. *Syst. Synth. Biol.*, **9**, 67–72.
8. Joshi,M., Kundapura,S.V., Poovaiah,T. (2013) Discovering novel anti-malarial peptides from the not-coding genome - a working hypothesis. *Curr. Synth. Sys. Biol.*, **1**, 1–4.
9. Chakrabarti,A., Kaushik,M., Khan,J. (2020) Discovery of tREP-18, a novel class of tRNA encoded peptide with potent leishmanicidal activity. *bioRxiv*, 10.04.325217.
10. Radivojevic,T., Costello,Z., Workman,K., Garcia Martin,H. (2020) A machine learning Automated Recommendation Tool for synthetic biology. *Nat. Commun.*, **11**, 4879.
11. Krishnan,N.M., Jain,P., Gupta,S., Hariharan,A.K., Panda,B. (2016) An improved genome assembly of Azadirachta indica A. Juss. *G3 (Bethesda)*, **6**, 1835–1840.
12. Krishnan,N.M., Katoh,H., Palve,V., Pareek,M., Sato,R., Ishikawa,S., Panda,B. (2019) Functional genomics screen with pooled shRNA library and gene expression profiling with extracts of Azadirachta indica identify potential pathways for therapeutic targets in head and neck squamous cell carcinoma. *PeerJ*, **7**, e6464.
13. Krishnan,N.M., Pattnaik,S., Jain,P., Gaur,P., Choudhary,R., Vaidyanathan,S., Deepak,S., Hariharan,A.K., Krishna,P.G., Nair,J. *et al.* (2012) A draft of the genome and four transcriptomes of a medicinal and pesticidal angiosperm Azadirachta indica. *BMC Genomics*, **13**, 464.
14. Krishnan,N.M. and Panda,B. (2019) Prediction of a plant intracellular metabolite content class using image-based deep learning. *bioRxiv*, 488783.
15. Gill,R.T., Valdes,J.J. and Bentley,W.E. (2000) A comparative study of global stress gene regulation in response to overexpression of recombinant proteins in *Escherichia coli*. *Metab. Eng.*, **2**, 178–189.
16. Holowko,M.B., Frow,E.K., Reid,J.C., Rourke,M., Vickers,C.E. (2021) Building a biofoundry. *Synth. Biol. (Oxf)*, **6**, ysaa026.
17. Crone,M.A., Priestman,M., Ciechonska,M., Jensen,K., Sharp,D.J., Anand,A., Randell,P., Storch,M., Freemont,P.S. (2020) A role for biofoundries in rapid development and validation of automated SARS-CoV-2 clinical diagnostics. *Nat. Commun.*, **11**, 4464.
18. Crone,M.A., Priestman,M., Ciechonska,M., Jensen,K., Sharp,D.J., Anand,A., Randell,P., Storch,M., Freemont,P.S. (2020) Author correction: a role for biofoundries in rapid development and validation of automated SARS-CoV-2 clinical diagnostics. *Nat. Commun.*, **11**, 4793.
19. Roundtable on Genomics and Precision Health; Forum on Drug Discovery D, and Translation; Board on Health Sciences Policy; Health and Medicine Division; National Academies of Sciences, Engineering, and Medicine. (2016) Deriving drug discovery value from large-scale genetic bioresources: proceedings of a workshop. The National Academies Collection: Reports funded by National Institutes of Health, Washington, DC, 2 November. PMID: 27631038.
20. Kitney,R.I., Bell,J. and Philp,J. (2021) Build a sustainable vaccines industry with synthetic biology. *Trends Biotechnol*, **37**, pii:S0167-7799(0120)30331-30330.