Written submission from Wellcome Sanger Institute (ENB0021)

Introduction

1. The Wellcome Sanger Institute is a world leader in genomics research. We apply and explore genomic technologies at scale to advance understanding of biology and improve health. Making discoveries not easily made elsewhere, our research delivers insights across health, disease, evolution and pathogen biology. We are open and collaborative; our data, results, tools, technologies and training are freely shared across the globe to advance science.

2. The Wellcome Sanger Institute has invested in a new programme of research, Generative and Synthetic Genomics, which seeks to develop underlying technologies and open datasets for engineering biology.

3. The Wellcome Sanger Institute is based on the Wellcome Genome Campus: a world-leading hub for genomes and biodata research. The campus is also home to EMBL-EBI, Connecting Science, Genomics England and several spin-out and start-up companies.

Key messages

4. Decades of forward-thinking investment by Wellcome and the UK Government has contributed to a highly advanced and internationally competitive genomics sector in the UK. The UK has the potential to develop an equally advanced engineering biology sector, subject to strategic, ambitious and coordinated investment in people, technology development, and data generation.

5. Government should prioritise the fundamentals of engineering biology to make it faster, cheaper and more accessible, with a view to enable a broader range of ambitious and useful applications in the future. To do this, government should outline an ambitious mission-driven strategy focussed on technology development and massive biological data generation to focus minds, promote investment, and attract talent from around the world. In particular, government should fund the scientific community to build massive open biological datasets quickly and efficiently to enable the development of open predictive machine learning models, which in turn will enable the multi-faceted engineering of biology.

6. Government could invest in a dedicated national engineering biology institute that can harness academic excellence, develop long-term technical and operational expertise, and attract the best leaders from around the world to relocate to the UK.

7. Computational labs developing and applying AI models should be brought together with experimental labs designing assays and generating data. A national engineering biology institute could be co-located with computer science research institutes and companies to facilitate cross-discipline collaboration.

8. Engineering Biology raises profound ethical questions that should be addressed with wider society through meaningful two-way dialogue that will influence scientific decision-making processes and support public trust. The Government could suppport researchers to embed ethical considerations and principles of Responsible Research and Innovation (RRI) in the scientific process by convening different stakeholders and by providing funding for these activities.

9. To bridge the engineering biology skills gap and make the UK a world-leader in engineering biology, the Government needs to consider how to better attract international talent.

<u>Question 1: What are the UK's key strengths in the area of engineering biology?</u>

10. Decades of forward-thinking investment by Wellcome and the UK Government has contributed to a highly advanced and internationally competitive genomics sector in the UK. For example, the UK is home to cutting-edge DNA sequencing facilities and companies, leading research institutes, and rich biodata resources (e.g. Genomics England, Oxford Nanopore, Wellcome Sanger Institute, EMBL's European Bioinformatics Institute, UK Biobank). The UK also boasts a thriving biotechnology landscape comprising a diverse array of small and innovative genomics start-up companies[1]. Much of what is considered "engineering biology" has genomics at its core. **Therefore, the UK has the potential to develop an equally advanced engineering biology sector, subject to strategic, ambitious and coordinated investment in people, technology development, and data generation.**

[1] BIA, Genomics Nation 2023, December 2023, https://www.bioindustry.org/policy/strategic-technologies/genomics/genomics-nation-2023.html

11. The Wellcome Sanger Institute has invested in a new programme of research, Generative and Synthetic Genomics, which seeks to develop underlying technologies and open datasets for engineering biology[2]. The goal of generative genomics is to build huge open genomic datasets that can be used to train open machine-learning models to predict and program biology. These models will predict the consequences of genetic variation on the properties of biological molecules and systems, and these predictions will allow researchers to design new biological molecules and systems from scratch. The goal of synthetic genomics is to develop methods and technologies to synthesise, edit and engineer large genomes at scale. In simple terms, generative genomics addresses the question of "what to build", and synthetic genomics addresses "how to build it".

12. Alongside UK and international research projects that have focussed on writing genomes of organisms[3,4], a key development in the field has been the expansion of the genetic code with new chemical and physical properties, led by scientists at the MRC Laboratory of Molecular Biology[5].

13. The US and China have established ambitious flagship engineering biology projects that have secured significant investment and attracted world-leading scientists[6,7]. The UK scientific community, supported by government, should act urgently to develop a similar large-scale and ambitious initiative in engineering biology (i.e. generation of massive open biological datasets, design of predictive machine learning models and development of technology to write genomes) that matches the ambitions of scientific endeavours such as the Human Genome Project and the sequencing of genomes in the UK Biobank – with embedded ethical design. This ambition is needed to focus minds, promote investment, and attract talent from around the world.

14. DNA synthesis, DNA sequencing and machine learning play a critical role in engineering biology, and both the US and China

---

[2] https://www.sanger.ac.uk/news_item/predicting-and-engineering-biology-in-new-research-programme/

[3] Bourzac K., Engineered yeast breaks new record: a genome with over 50% synthetic DNA, Nature 623, 4692023, November 2023, https://doi.org/10.1038/d41586-023-03495-4

[4] Fredens J. et al., Total synthesis of Escherichia coli with a recoded genome, Nature 569, 514–518, May 2019, https://doi.org/10.1038/s41586-019-1192-5

[5] Chin J. W., Expanding and reprogramming the genetic code, Nature 550, 53–60, October 2017, https://doi.org/10.1038/nature24031

[6] https://engineeringbiologycenter.org/

[7] https://en.genomics.cn/en-organization-3243.html

outcompete the UK in these areas. For example, China's Institute of Super Cell is part of the BGI group, which performs DNA sequencing at a larger scale than the Wellcome Sanger Institute, and Chinese research institutes publish more highly-cited papers on AI than even the US[8]. The world's largest DNA synthesis companies are predominantly based in the US (e.g. Twist Bioscience, Integrated DNA Technologies, Agilent) or China (e.g. BGI), though these can only produce short lengths of DNA using chemical synthesis. When working at large-scale, UK research organisations procure technology and services from large overseas companies to ensure robust support infrastructure and quality assurance. This exposes UK research institutes and companies to exchange rate issues, and makes building relationships with providers slower and more difficult. More broadly, it constrains the UK's ambitions to develop a world-leading and holistic engineering biology ecosystem. Promisingly, many companies commercialising next-generation synthesis methods are UK-based, including NunaBio, Evonetix and Camena Bioscience. Next-generation synthesis methods such as enzymatic synthesis and thermal control allow the production of more useful longer pieces of DNA[9].

<u>Question 2: What are the key applications of engineering biology?</u>

15. The primary aim of generative genomics and synthetic genomics is to enable scientists to explore causal relationships between genome sequence and biological function. This approach will maximally leverage the knowledge developed from genomics research and data to better understand and improve human health. For example, with technologies that allow the synthesis of large genomes, scientists will be able to better study and understand complex large-scale genetic rearrangements that lead to cancer. Compared with current genome editing techniques, genome design and synthesis would allow scientists to make genomic changes at great scale and density and create biological systems with properties that cannot be accessed by natural evolution.

16. Other short-term applications include the engineering of cells to produce useful molecules, such as therapeutic proteins. Protein design is expected to accelerate and expand in scope over the

---

[8] https://www.business-standard.com/article/international/china-leads-research-in-advanced-technologies-like-ai-us-analysts-report-123011801146_1.html

[9] Seydel C., DNA writing technologies moving toward synthetic genomes, Nature Biotechnology 41, 1504–1509, October 2023, https://doi.org/10.1038/s41587-023-02006-0

coming years through the use of increasingly sophisticated machine learning algorithms trained on massive biological datasets.

17. In the medium-term, cell engineering could be used for more than producing molecules. Cells will increasingly be engineered into useful experimental models or cellular therapies. This approach is already being used clinically in CAR T-cell therapy, and advances in engineering biology could broaden the applicability of cellular therapies to treat a range of cancers.

18. In the long-term, whole genome synthesis could be used to engineer fully bespoke and tuneable cellular therapies. Other long-term applications of these technologies might include transplantation of tissues with virus-resistant scaffolds, repair of genomes with large rearrangements, and creation of cells of extinct or novel species. However, these are highly speculative and may never be realised.

19. Rather than focus too heavily on applications, government should prioritise the fundamentals of engineering biology (i.e. development of underlying technology and generation of massive open biological datasets) to make it faster, cheaper and more accessible, with a view to enable a broader range of ambitious and useful applications in the future.

<u>Question 3: How can government policy support the development of engineering biology?</u>

20. The key underlying technologies that enable the applications of engineering biology have arisen from two scientific and technical revolutions of the past decades:

a. The revolution in DNA sequencing, synthesis and editing technologies. These technologies enable biological engineers to carry out experiments that explore perturbations of proteins, pathways and cells. Efforts are ongoing to develop technology for the synthesis of large sections of DNA (up to the chromosome or even genome level) and the transfer of large sections of DNA between cells. This will allow the engineering of large-scale structural variation in genomes in addition to minor perturbations.

b. The revolution in machine learning. Predictive and generative machine learning models, if trained on biological data of sufficient scale and diversity, can predict the consequences of genetic variation and design biological molecules or systems from genomic sequence.

21. Key to the advancement of engineering biology is the generation of massive open biological datasets. For example, the scientific community has built the Protein Data Bank (PDB), a rich open database of 3D protein structures, which was used by Google Deepmind to develop the machine learning model Alphafold to predict protein structure from sequence[10]. The engineering biology community needs equivalent open datasets for other areas of biology, such as enzymatic activities, protein stabilities, and data describing larger scale cell pathways. However, in general, it can be difficult to receive funding for technology development and data generation projects from UKRI. **The Government should fund the scientific community to build these datasets quickly and efficiently to enable the multi-faceted engineering of biology.**

22. Importantly, these massive datasets should be shared openly and widely among the scientific community to accelerate the development and democratisation of engineering biology technology, and to ensure that predictive machine learning models trained on these datasets are transparent and open to scrutiny. In addition, the architecture and parameters of these predictive machine learning models should also be openly shared to enable reusability and transparency.

23. The Government's National Vision for Engineering Biology[11] sets out support for the underpinning technology of engineering biology alongside supporting industrial applications across sectors. This whole system approach is very welcome, as our impression is that government has thus far been overly focussed on applications of current engineering biology technology rather than unleashing its full potential by improving upon the underlying technology.

24. In the vision, government acknowledges that the UK is a leader in engineering biology due to "forward thinking investment … over the last decade". Government funding of Synthetic Biology Research Centres in 2012[12] and the Engineering Biology Mission

---

[10] https://alphafold.ebi.ac.uk/

[11] https://www.gov.uk/government/publications/national-vision-for-engineering-biology/national-vision-for-engineering-biology

Hubs and Mission Awards this year[13] was certainly welcome. However, our perspective is that government investment has been relatively small-scale, with funds distributed across the higher education sector in a piecemeal way. Although there is promise to scale up investment to £2 billion over 10 years, there is little detail on how this would be directed. **The Government, possibly through ARIA, should outline an ambitious mission-driven strategy focussed on technology development and massive open biological data generation to focus minds, promote investment, and attract talent from around the world.**

25. We can learn lessons from how the Human Genome Project – which was largely a technology development and data generation endeavour with a clear and ambitious goal from the outset – attracted significant funding, and has since generated substantial health and economic benefit[14]. Similarly, investment in the development of technology and open biological datasets that underlie engineering biology would likely see the sector flourish over the coming decades. It has been estimated that biological materials and engineering biology could eventually produce 60% of the physical inputs to the global economy[15].

26. Rather than supporting a distributed engineering biology academic network across the UK, as is the status quo, government should consider a centralised model that could concentrate expertise and achieve scale, and form the foundations of an engineering biology sector that can compete with the US and China. For example, **government could invest in a dedicated national engineering biology institute that can harness academic excellence, develop long-term technical and operational expertise, and attract the best leaders from around the world to relocate to the UK**. This would not preclude engineering biology specialisms from developing across the UK – simply that these would exist alongside a centralised coordinated effort to achieve the ambitious aims of large-scale data generation and technology development. A national institute could comprise a fleet of experimental automated platforms for measurement and testing of engineered proteins and cells at scale. Such platforms are not

---

[12] https://www.ukri.org/what-we-do/browse-our-areas-of-investment-and-support/synthetic-biology-for-growth/

[13] https://www.ukri.org/news/new-100m-fund-will-unlock-the-potential-of-engineering-biology/

[14] https://www.genome.gov/27544383/calculating-the-economic-impact-of-the-human-genome-project

[15] https://www.mckinsey.com/industries/life-sciences/our-insights/the-bio-revolution-innovations-transforming-economies-societies-and-our-lives

currently routinely used in academic laboratories, and access could be made available to researchers across the UK.

27. Engineering biology is a cross-discipline field that relies on skills in computer science, engineering, and biological science. **Computational labs developing and applying AI models should be brought together with experimental labs designing assays and generating data. For example, a national engineering biology institute could be co-located with computer science research institutes and companies to facilitate cross-discipline collaboration.** In particular, collaboration with big tech companies such as Microsoft, Google AI and Meta, who have unrivalled capacity and expertise in machine learning but less experience in biological data, could be a key step in accelerating progress in large scale biodata generation and creation of machine learning models.

28. In addition, UK universities should explore ways to make undergraduate courses more interdisciplinary. For example, undergraduate biology courses should involve training in computer science and machine learning, and engineering and computer science courses should expose students to the biological applications of their subjects.

29. Pleasingly, principles of Responsible Research and Innovation (RRI) are referenced throughout the national vision for engineering biology. Although most of the focus relates to a national security and risk context, there are occasional references to other aspects of RRI, such as the consideration of societal and ethical factors in scientific decision making. RRI depends on two-way dialogue on research and technology development between the scientific community and the public, and so it is encouraging to see that the vision outlines a commitment to an open public dialogue on engineering biology.

<u>Question 4: How can the UK maximise the economic potential of developments in engineering biology?</u>

30. Biotech start-ups based in the UK often find it difficult to scale into large multinational companies as they often encounter a significant bottleneck at Series A funding and beyond. This leaves the field open for large firms, often headquartered in the US, to acquire these start-ups and integrate their technology into their

portfolios. A typical example is Illumina sequencing – the most dominant sequencing technology in the world – which was developed by UK-based Solexa before being bought and scaled by US-based Illumina. The Government should consider various approaches to bolster scale-up investment in UK start-ups, including initiatives that leverage UK markets to stimulate venture capital such as the Chancellor's Mansion House Reforms.

31. Engineering biology is a multi-disciplinary field where relevant stakeholders might belong to completely separate ecosystems e.g. biopharma vs agritech vs consumer goods. Engagement, networking and collaboration among experts, companies and investors with different skills and focuses is crucial for growth of the sector. Government can play a crucial role in convening investors and companies across sectors and facilitate cross-discipline working.

32. The Government could also support translation of academic research through the mixing of academic institutions with start-ups and larger firms to facilitate knowledge transfer and innovation. This could constitute a consortium of academic and industry partners funding pre-competitive data generation and machine learning model building for generative and synthetic genomics through a model similar to Open Targets[16], a public-private partnership for systematic identification and prioritisation of drug targets. The Government could also fund academic researchers to train in entrepreneurship, through programmes similar to the Wellcome Genome Campus' startup school – a six month ten-session programme that enables participants to explore scientific ideas stemming from their research through an entrepreneurial lens, with the opportunity to be mentored by members of the Cambridge life sciences community[17].

Question 5: What are the risks posed to society by engineering biology?

33. Many of the risks of misuse of engineering biology technology are similar to the existing risks arising from developments of other genomic technologies. Understanding the technologies and data that underlie engineering biology and developing predictive models in the UK will be important mitigations against these risks. For example, checkpoints could be put in place at the stage of DNA synthesis to ensure DNA synthesis companies, supported by

[16] https://www.opentargets.org/
[17] https://www.wellcomegenomecampus.org/scienceandinnovation/startup-school/

government, screen orders using machine learning models that can predict the likely intended function of the DNA sequence. Nevertheless, the risk of misuse will be challenging to mitigate completely, as checkpoints might be circumvented through ordering sections of DNA from multiple DNA synthesis companies and combining the products of each request. However, when assessing the balance between risk and benefit, it should be recognised that bioterrorists or hostile states could achieve the same ends by using naturally occurring pathogens without using engineering biology.

34. The Wellcome Sanger Institute, and the UK as a whole, has a strong culture of open access research due to the many benefits this offers, such as fostering transparency and improving public trust, increasing the reusability of research and the speed at which it can be translated, and enabling increased collaboration across disciplines. Therefore, predictive machine learning models for engineering biology, and the massive biological datasets that underlie them, should be made openly available. Although open sharing of data and models may increase the risk of misuse, this must be balanced against these huge benefits of open science. In addition, ensuring a culture of openness and transparency in the architecture, parameters, and underlying datasets of predictive models will allow some mitigation of risk by enabling scrutiny of models by the engineering biology community and governments (e.g. the UK AI Safety Institute).

35. A more immediate concern is the risk of misaligning engineering biology research and development goals with societal values, and the subsequent risks of reduced trust in scientific institutions and downstream technologies. **Engineering Biology raises profound ethical questions that should be addressed with wider society through meaningful two-way dialogue that will influence scientific decision-making processes and support public trust.** When engaging the public, researchers should avoid focus on long-term speculative applications, but instead identify practical ways in which public groups can influence decision-making throughout research and technology development in an iterative manner. This approach enables research processes and outcomes to be more inclusive and address a wider variety of ethical and societal issues.

36. **The Government could support researchers to embed ethical considerations and principles of RRI in the scientific process by convening different stakeholders and providing**

funding for these activities. Research grants could stipulate that expertise in RRI is embedded within engineering biology teams and institutes to facilitate incorporation of RRI principles within the research and development processes. In addition, UK universities could further embed regulatory and ethical aspects of research in STEM undergraduate and postgraduate courses.

Question 6: How should engineering biology be regulated?

37. Much discovery research considered to be engineering biology is not covered by existing regulation, and does not need to be. However, the work does raise novel ethical questions and challenges, which researchers should be addressing in partnership with regulators, government, legal experts, ethicists and public groups, as the research develops, with government playing a coordinating role.

38. There are adjacent research and technology domains with governance strategies that researchers in engineering biology could take inspiration from. The International Society for Stem Cell Research (ISSCR) Guidelines for Stem Cell Research and Clinical Translation[18] employs a voluntary global governance framework that utilises a scientific self-governance approach to advise on the use of human embryonic stem cells and on clinical translation of stem cell-based medical interventions. The WHO global framework for governance of human genome editing[19] provides advice and recommendations on institutional, national, regional and global governance mechanisms. A systematic exploration of related technologies and associated governance approaches could be an effective way to explore governance of engineering biology.

39. To ensure equitable benefit sharing from the outputs of engineering biology technology, it will be important to improve diversity in the underlying datasets of machine learning models and the representation of researchers involved in developing the technology. In addition, government should support engagement with international publics and indigenous communities to enable international awareness and input of expertise and pave the way towards globally inclusive governance.

[18] https://www.isscr.org/guidelines

[19] WHO Expert Advisory Committee on Developing Global Standards for Governance and Oversight of Human Genome Editing, Human Genome Editing: A framework for governance, July 2021, https://www.who.int/publications/i/item/9789240030060

Question 7: What are the possible barriers and limitations to good and effective use of engineering biology?

40. Biology is incredibly complicated, comprised of many components, and is incredibly difficult to predict and engineer. Thus far, engineering biology efforts have been too small-scale, piecemeal and application-focussed to allow full exploration of genetic variation. Engineering biology has potential to explore the vast uncharted sequence space of biological molecules and systems but only if the scientific community is supported in building massive open biological datasets quickly and efficiently (see Q3).

41. Engineering biology is a cross-discipline field that relies on people with skills across computer science, engineering, and biological science. A nationwide shortages in these skills has led to significant challenges in recruitment for both academia and industry[20]. Exacerbating this, postdoctoral scientists are usually employed on unstable short fixed-term contracts of typically 2-5 years. It is becoming increasingly difficult to recruit and retain postdoctoral scientists, particularly those who have cross-discipline expertise in wet lab and computational skills necessary for engineering biology as they are often desirable candidates for higher-paying secure roles in industry and other sectors.

42. Supporting development of home-grown talent is necessary to sustainably address the skills gap. Machine learning and data science is being introduced in schools and further education colleges, but teachers often do not feel they have the necessary skills or confidence to teach these topics.

43. Another part of the solution to bridging this skills gap is to make the UK a more attractive destination for scientists from across the globe, as referenced in the national vision for engineering biology. There are considerable obstacles to this, including expensive visa costs exacerbated by the recent increase to the Immigration Health Surcharge[21]. Uncertainty from the announcement of increased salary requirements for skilled worker and family visas[22,23] also risks discouraging international scientists

[20] BIA, Genomics Nation 2022, July 2022, https://www.bioindustry.org/static/8d69c95e-dfa3-4470-b9a3f230800f7026/BIA-Genomics-Nation-2022.pdf
[21] https://commonslibrary.parliament.uk/research-briefings/cbp-9859/
[22] https://www.gov.uk/government/news/home-secretary-unveils-plan-to-cut-net-migration
[23] https://www.science.org/content/article/u-k-visa-changes-imperil-recruitment-scientific-talent-policy-

from working in the UK. **If the Government is serious about bridging the engineering biology skills gap and making the UK a world-leader in engineering biology, it needs to consider how to better attract international talent.**

07 May 2024

experts-warn