Trends in Biotechnology

CellPress

Special Issue: Industrial Biotechnology

Opinion

The Best Model of a Cat Is Several Cats

Claes Gustafsson[1,*] and Jordi Vallverdú[2,*]

Modern biotechnology is emerging at the intersection of engineering, biology, physics, and computer science. As such it carries with it history from several disparate fields of research including a strong tradition in deductive reasoning primarily derived from discovery focused molecular biology and physics. Engineering biological systems is a complex undertaking requiring a broader set of epistemic tools and methods than what is usually applied in today's discovery based research. Inductive reasoning as commonly used in computer science has proven to be a very efficient approach to build knowledge about complex megadimensional datasets, including synthetic biology applications. The authors conclude that the multi-heuristic nature of modern biotechnology makes it an engineering field primed for inductive reasoning to complement the dominating deductive tradition.

The Epistemic Challenge of Modern Biotechnology

Biotechnology has a long history starting with beer and bread production via early genetic engineering for manufacturing of insulin up to current date synthetic biology where complete pathways and genomes can be designed and synthesized *de novo*. Whereas an earlier version of biotechnology relied on discovery based science and (in best case) random search for improved non-natural biological entities, the tools made available over the past decade instead allow for precise design and exact synthesis of a large number of putative solutions to any biological problem. Biotechnology is consequently emerging as an engineering driven science [1]. Where discovery based biotechnology goes searching for naturally existing proteins or genes to combine in interesting new ways, engineering driven biotechnology instead builds novel biology to specifications. The transition from discovery science to engineering driven applications affects biotechnological research in a number of different ways including the challenge of epistemology (the theory of knowledge) [2] (Box 1).

Computational modeling, intensive artificial intelligence (AI) systems, and rigorous statistical procedures are increasingly being placed at the heart of biotechnology together with highly innovative methods, techniques, and ideas from engineering, biology, computer science, physics, and chemistry. In addition to the increased breadth of technologies, there is also a rapidly increasing size and velocity of data being generated to the extent that human data supervision is coming to an end [3]. We are entering an era where scientists can no longer check, verify, replicate, or even analyze the deluge of incoming data. Scientific research is becoming a black box that provides us with new knowledge. As our science paradigm changes, we need to address some of the cornerstones of how we extract knowledge and better understand how we know what we know.

The combination of analytical and synthetic epistemic approaches put us at an interesting position: we create new knowledge and new biological systems, our ideas work, but we are often not able to deconvolute the underlying truth of the mechanistic models supporting our

Trends

Complexity modeling. Biotechnology engineers living systems that are the product of evolution and can be discretized and organized following different organizational layers, each one with its own mechanisms and contexts. Bioengineering practitioners must be aware of the rich and complex relationships among these levels of complexity.

Reproducibility. Most of biotechnology cannot be reproduced for practical, temporal, and economic reasons. Despite this, designed bioengineered entities must work consistently and robustly when released as products.

Multi-heuristics. Several heuristics are implemented in parallel in modern biotechnology. Each one follows a reasoning procedure that provides possible solutions and logical explanations. They are not contradictory, just useful for pragmatic reasons. The challenge is to build engineering knowledge for a biological field.

[1]DNA2.0 Inc., 1140 O'Brien Drive, Menlo Park, CA 94025, USA
[2]Universitat Autònoma de Barcelona, 08193 Bellaterra (BCN), Catalonia, Spain

*Correspondence:
cgustafsson@dna20.com (C. Gustafsson) and
jordi.vallverdu@uab.cat (J. Vallverdú).

Box 1. From a Black Box to Networked Systems: Challenges Facing Biotechnology

Reproducibility. With the exponential increase in scientific publications, datasets, and data sizes comes a limitation in what can be effectively reproduced. Experimental replication by a separate party is difficult and mostly ignored practice. In the few instances where biological experiments are replicated, the results have been less than perfect [37,39]. Experiments can be designed to be internally reproducible via, for example, systematic variance and model building. In reality, the exponential growth of published research in the fields of bioengineering makes it impossible to reproduce all results [40,41].

Supervision. The implementation of intensive AI, ML, and other automated learning systems for the purpose of managing biotechnology labs with increasingly larger datasets will lead to a supervision problem. Given the complexity of biological variable correlations and relationships and the growth of nonlinear, stochastic, and noisy data among complex environments [5], it will be challenging to design a system to handle unforeseen results or contradictory data and be able to expect the unexpected.

Standardization. In many new fields including synthetic biology and systems biology, standardization is a problem for practitioners. The creation of the Synthetic Biology Open Language (SBOL) was motivated by the necessity of being able to represent genetic designs through a standardized vocabulary of schematic glyphs (SBOL Visual) as well as a standardized digital format [42]. Not only is there an urgent need to standardize the genetic elements but also a unification of measurements and protocol automation to reduce what has been described as 'the impact of "cultural art" in laboratory methods' [43].

Validation. Once data is obtained, it must be processed and validated in its context so that relevant knowledge can be extracted [6]. There are numerous ways to triangulate and iterate the obtained data for epistemic purposes, but in most cases here again emerges the classic statistical problem: correlation does not imply causation [44].

Data management. With complex megadimensional structure–function biology data come issues with data management. Automated cloud and emerging statistical analysis includes platforms such as Riffyn [45] and DNA ATLAS [46], or new communities such as DIYBIO [47]. What are the measures for navigating sequence–function space? How to capture and illustrate nodes in space not sampled but only explored by triangulation from experimental data? Challenges also relate to the obtained data, signal distribution, and the somewhat arbitrary defined sequence identity of chemically synthesized genetic materials and the lack of agreement of parameters for measuring the degree of standardization.

Uncertainty. Biology is dynamic and context-dependent in both spatial and temporal dimensions. Signals are often stochastic further diminishing the deductive predictability. Kwok describes five key concerns for building biology networks: undefined sections, unpredictable circuitry, unwieldy complexity, incompatible parts, and crashing variability [48]. Engineering living systems is then an exercise in humility because of all the inherent uncertainties of all the parts in the system.

Ethics. Many are the reviews and opinions on ethics in synthetic biology, ranging from the 'playing God' discussion to patents versus open source, biosecurity, information control, data storage, production permissions, and regulation [49–51].

conclusions [4]. The genetic elements and biobricks used by synthetic biology researchers and increasingly across the board of biotechnology are not static or linearly additive, but dynamic, context-dependent, and interactive [5]. Here can be found the friction point between truly understanding biology in all its mechanistic detail versus engineering biology for novel functional applications [6], with the transversal problem of the role of computational simulations for modeling [7]. This is a growing epistemic struggle as biotechnology is rapidly moving towards larger wet lab datasets while simultaneously embracing *in silico* lab.

Starting with Cohen and Boyer in the mid-1970s, the building of new proteins, pathways, and genomes is gradually becoming commonplace. Bioengineering success stories ranging from the synthesis of the 14 amino acid hormone somatostatin by Genentech in 1977 [8] to the recent 23 gene opioid production pathway engineered by the Smolke lab [9] shows the magnitude of technical advancements over the past few decades. Despite recent success, many biological phenomena ranging from tissue development of organs to manipulating neuronal information are still outside our reach. As the complexity of the systems increases and data accumulates, we will have to bring onboard additional ways to look at results and data and explore different epistemic tools to build knowledge.

Predicting Biology

Biotechnology often relies on an assumed perfect understanding of biological processes to, for example, replace a promoter in a transcriptional unit. Despite the single variable change in an otherwise constant system, it has proven exceedingly difficult even in extremely well controlled systems to precisely predict the phenotypical outcome of a promoter replacement [10]. This is in stark contrast to physics where electrodynamics, optics, magnetic potentials, and electrical circuits all can be perfectly predicted by just four linear partial differential equations (Maxwell's equations). Instead biology is a gooey and redundant complex megadimensional mess of synergy and antagonism, and an abundance of variables that just come along for the ride and correlates with other variables primarily due to evolutionary history, not because of casual effects. Models based on physicochemical first principles rarely suffice to fully explain complex biological functions, at least not at our current level of biological understanding.

The gap between reality and the model of reality is a classic philosophical problem: the infra-determination of causal role theories. You can be using a model to explain your results, it works and satisfies your theoretical and practical requests. . .but it is false. It happened to Copernicus with his heliocentric model, it happened to Newton and his universal model, and to Becher's phlogiston theory. The foundation of biotechnology is based on Crick's central dogma describing the flow of information within a biological system [11]. But the central dogma is incorrect, as evidenced by reverse transcriptase, prions, catalytic RNA, and many additional processes that all violate the core principles of the central dogma [12]. Even the most established of chemical models, H_2O, is oversimplified and incorrect. The reason for water properties such as conductivity, high boiling point, high surface tension, etc., cannot be explained by the H_2O model. Instead, the current scientific model of water requires that each H_2O is hydrogen bonded with up to four other water molecules, and these hydrogen bonds are constantly being broken and reformed every 10^{-12} seconds, creating a large dynamic net of hydrogen bonds [13,14]. This revised model is still just a model, but a much better model at predicting the properties of water. To quote statistician George Box, 'all models are wrong, but some models are useful'.

Similarly in synthetic biology and systems biology, most newly created gene networks are non-functioning due to uncertain initial conditions, stochastic distribution, unaccounted for variables, and disturbances of the extracellular environments [15]. Impressive gene networks have been built *de novo* to specifications but only after many iterations and substantial investment in time and resources. The opposite is also true; numerous are the successful semirandom protein directed evolution experiments focused on altering functional properties where the amino acid substitutions that have the most significant effect on the function were impossible to predict in advance [16].

As a formal conclusion and following the nature of our previous arguments, we can affirm that biological nature is beyond our epistemic real understanding ('understanding' in the sense a mechanical engineer would use the term). We are only able to create idealized maps of the biological territories, but the truth is that the best model of a portion of reality is the same reality, as expressed so nicely by Borges with his brilliant short story '*On Exactitude in Science*' [17]. The model itself is not the truth, just an intentional symbolic oversimplification.

The Philosophies of Truth: Inductive versus Deductive Reasoning

Philosophers have long been discussing how we know what we know and the methods to reach true knowledge (Table 1). Beyond debates about how we capture data and the reliability of senses or instruments, there is a second-level problem: how to extract knowledge from results. Historically, two main and opposite approaches for reasoning have been situated at the core of any knowledge: inductive and deductive reasoning. Inductive methods refer to the process of inferring a general law or principle from the observation of particular instances, whereas a

Table 1. Synthetic Biology and Reasoning

	Reasoning Process		
	Deductive	Inductive	Abductive
Methods and Techniques	Frequentist statistical methods Data testing	Machine learning Bayesian statistics Predictions	Research guesses Modeling and simulation

deductive method is the inference of particular instances by reference to a general law or principle. Hume, Popper, and Carnap worked on inductive ideas while deductive support was defended by Aristotle, Descartes, and Frege, among others. One simplistic example is how we know the dropped coin will fall to the ground. Deductive reasoning says, 'I know the coin will fall to the ground because the two bodies "coin" and "earth" attract each other with a force that is proportional to the product of their masses and inversely proportional to the square of the distance". Inductive reasoning instead says, 'I know the coin will fall to the ground because when I dropped it yesterday and every day in the past it would always fall to the ground'.

And then there is abductive reasoning. Within science in general and the fields of law, computer science, and AI in particular, the concept of abductive reasoning has taken hold [18]. Abduction, or as it is also often called 'inference to the best explanation' or simply 'best guess', is a type of inference that assigns special status to explanatory considerations. The example being: 'The lawn is wet. It probably rained last night'. Other possible explanations exist (sprinklers, hail, or dogs), but rain is the simplest and most efficient explanation. Most philosophers agree that this type of inference is frequently employed, in some form or other, in everyday life and in scientific reasoning [19].

Abductive and inductive reasoning is emerging in systems biology and bioengineering sciences. For example, cellular metabolite fluctuations and response to environmental challenges was predicted and explanatory rules derived from metabolite concentrations combined with existing pathways and kinetic models as input data using a type of inductive logic programming (ILP) called consequence finding induction (CF induction), which integrates abductive and inductive methods. The tool was used to explore all possible metabolic pathway solutions in hyperspace and identifying the most effective and simplest models that could best describe the results [20]. ILP can be defined as the intersection of machine learning (ML) and logic programming to deal with induction in first-order logic. Thus, unlike many other ML techniques, ILP has applicability to discover causal relations and identifying missing biochemical knowledge (Y. Yamamoto, PhD thesis, The Graduate University for Advanced Studies, Hayama, Japan, 2010).

Similarly, abductive logic programming (ALP) has been used to analyze and elucidate relationships from microarray signals to provide general models of how gene interactions cause changes in observable expression levels of genes. The models derived from the ALP process is a formalization of the implicit reasoning that governs biologists designing microarray experiments [21] (I. Papatheodorou, PhD thesis, University of London, 2007).

Any biological process can be understood through deductive or inductive reasoning. At DNA2.0 Inc., we have spent the past decade or so systematically exploring the coding preferences of recombinant genes to maximize heterologous protein expression and identified certain codon biases and other patterns that have causal correlation with high protein expression. We can use deductive reasoning (or more likely abductive reasoning in this case) and explain the codon bias preferences as a consequence of dynamic tRNA amino acid charging differences due to metabolic stress – a model with attractive logic and reasonably well supported by data [22]. Alternatively, we can use inductive reasoning and conclude that certain codons and patterns

have a causal correlation with recombinant protein expression yield [23]. Inductive reasoning provides us with testable, robust, and transferable predictions, whereas deductive reasoning based on the same data has proven less useful.

Where deductive reasoning stumbles, inductive reasoning can be very useful. Statistics and ML derived from inductive reasoning are directly applicable to high complexity systems. Successful examples of applying ML to biological sciences include predicting therapy efficacy from genomic data [24], small molecule drug design through QSAR (quantitative structure–activity relationship) [25], annotation of genetic elements [26], and many other applications recently reviewed [27].

Biology Is Complicated

Engineering biological systems is a complex undertaking. The cell can be engineered on gene, protein, pathway, and genome levels (Figure 1). Each level represents a different abstraction level of the underlying genome sequence and adheres to level-specific assumptions and constraints for how we believe we affect the system, but the different abstraction levels are just figments of our deductive imagination. Changing an A to a C somewhere in the genome for the purpose of introducing an amino acid substitution in a protein may unintentionally affect mRNA structure, RNAi binding sites, chromosome tertiary structure, and a plethora of other biological events. Sometimes the outcome of the engineering supports the prediction model, but is that because we only look for solutions that fit the model (looking for car keys under the streetlight)?

From an ML perspective the genome is just a long string of agnostic ACGT information. The $>10^{12}$ base pairs currently available in GenBank constitute a poorly systematized nucleotide sequence repository that determines all of the known biology, even if we do not understand most of the sequence–function correlations. Inductive ML methods when based on systematically varied data can be used to model the sequence–function correlation and build testable prediction models without any underlying mechanistic understanding. When validated, these inductive models can create novel biological processes encoding properties not previously seen

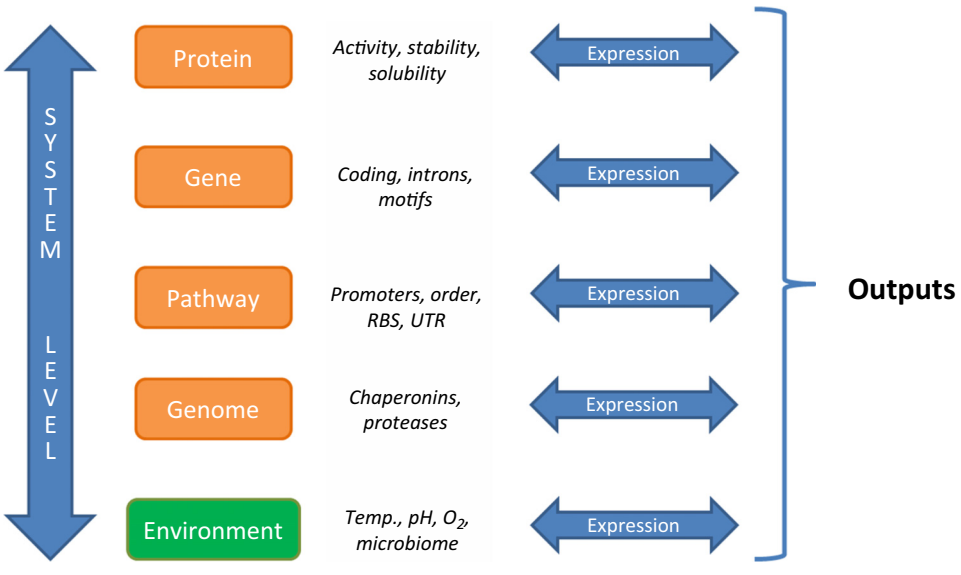

Figure 1. Multisituated Expression and Causal Determination of Biological Mechanisms. Systems biology attempts to explain the causal behavior at different system levels, which are subsequently combined nonlinearly for the whole organism. This multileveled complexity makes it difficult (impossible?) to achieve a complete mechanistic engineering approach to dynamic living entities.

in biology [28–31]. Directed evolution and similar random search-based paradigms do not include a learning step and accordingly are not inductive ML methods.

The deep complexity of living objects prohibits a singular epistemic approach. We are here faced with the fact that a multitude of models and methods may all provide different facets of the underlying truth of biology. Consequently, several epistemic approaches and methods should be considered valid in parallel as they in combination provide a better understanding of biology than each one separately.

Extending Reliable Epistemology to Biotechnology

Even though bioengineering has successfully constructed many amazing non-natural biological entities ranging from small molecules [32] to proteins [33] to pathways [34], the technology is not based on a solid foundation of sequential layers of additive knowledge. Even more so, perhaps we will never be able to explain biology as whole, instead be content with exploring it as the sum and dynamic interaction of several complex models that operate at different levels (see Outstanding Questions). Nevertheless, proper engineering practices make it possible to operate and create successful novel biological entities despite using only models with poor predictive accuracy. In some cases, such as WholeCellSimDB [35], the combination of engineering and understanding combines theoretical and practical results for an integrated outcome. Again, the dynamic, stochastic, and complex behavior of biological systems and subsystems show a nonlinear behavior that is a pragmatic research paradigm.

Our first conclusion is that modern biotechnology represents a crucial challenge for epistemic debates in contemporary sciences. Secondly, biotechnology is a clear example of multi-heuristic research as it often struggles to apply the principles of deductive science to biology. In a series of recent publications, the reproducibility of landmark oncology studies are questioned [36], drug target models are often inaccurate [37], and non-confirmatory findings are difficult to publish [38]. How do we really know what we know? The old problem of knowledge is exacerbated in the field of biotechnology where the building of novel biological systems is based on combining supposedly additive biological elements for predictable phenotypical function.

This led us to a third and last conclusion: is there such a thing as objective absolute knowledge that can be isolated and mapped, or are we just dealing with fuzzy sets of dynamic data that are highlighted and tagged for practical purposes without ever reaching the real truth? The more tools we have in our epistemic toolbox, the better we are equipped to build new biology based on the glimpses of knowledge the data provides us with. Pluralism in reasoning and logic will give us maximal exposure to the truth. To paraphrase the famous mathematician Norbert Wiener – 'the best model of a cat is several cats'.

Outstanding Questions

What does it mean to 'understand' biology?

Contrary to mechanical engineering or physics, biology is the product of billions of years of multidirected evolution with a combination of current and ancestral selective pressures that typically has nothing to do with the properties being engineered by biotechnology practitioners. It is difficult to weed out correlation from causality.

How do we navigate vast biological sequence–function space?

Sequence space is large. Even a small 16 amino acid antibody variable region requires 20^{16} ($\sim 10^{21}$) variants to be sampled exhaustively. This is approximately equal to the grains of sand on earth. We need statistical tools, data management processes, and epistemology different from what is typically used today to address virtual and physical data of that size, scope, and velocity. Many of these tools are already available in adjacent 'big data' fields.

How do we integrate current deductive knowledge with future inductive knowledge?

Will we have different epistemology for different classes of information, or can we use a pluralism of epistemologies in parallel, just like physics can treat light as both a particle and a wave simultaneously.

References

1. Andrianantoandro, E. *et al.* (2006) Synthetic biology: new engineering rules for an emerging discipline. *Mol. Syst. Biol.* 2, 20060028
2. Bensaude Vincent, B. (2013) Discipline-building in synthetic biology. *Stud. Hist. Philos. Biol. Biomed. Sci.* 44, 122–129
3. Casacuberta, D. and Vallverdú, J. (2014) E-Science and the data deluge. *Philos. Psychol.* 27, 126–140
4. Morange, M. (2009) A new revolution? The place of systems biology and synthetic biology in the history of biology. *EMBO Rep.* 10 (Suppl. 1), S50–S53
5. Gramelsberger, G. (2013) The simulation approach in synthetic biology. *Stud. Hist. Philos. Biol. Biomed. Sci.* 44, 150–157
6. Knuuttila, T. and Loettgers, A. (2013) Basic science through engineering? Synthetic modeling and the idea of biology-inspired engineering. *Stud. Hist. Philos. Biol. Biomed. Sci.* 44, 158–169
7. Hunt, C.A. *et al.* (2009) At the biological modeling and simulation frontier. *Pharm. Res.* 26, 2369–2400
8. Itakura, K. *et al.* (1977) Expression in *Escherichia coli* of a chemically synthesized gene for the hormone somatostatin. *Science* 198, 1056–1063
9. Galanie, S. *et al.* (2015) Complete biosynthesis of opioids in yeast. *Science* 349, 1095–1100
10. Kosuri, S. *et al.* (2013) Composability of regulatory sequences controlling transcription and translation in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 110, 14024–14029
11. Crick, F. (1970) Central dogma of molecular biology. *Nature* 227, 561–563
12. Shapiro, J.A. (2009) Revisiting the central dogma in the 21st century. *Ann. N.Y. Acad. Sci.* 1178, 6–28
13. Chang, H. (2012) *Is Water H_2O? – Evidence, Realism and Pluralism.* Springer
14. van der Post, S.T. *et al.* (2015) Strong frequency dependence of vibrational relaxation in bulk and surface water reveals sub-picosecond structural heterogeneity. *Nat. Commun.* 6, 8384

15. Chen, B.S. *et al.* (2009) Robust synthetic biology design: stochastic game theory approach. *Bioinformatics* 25, 1822–1830
16. Bloom, J.D. and Arnold, F.H. (2009) In the light of directed evolution: pathways of adaptive protein evolution. *Proc. Natl. Acad. Sci. U.S.A.* 106 (Suppl. 1), 9995–10000
17. Borges, J.L. (1975) *On Exactitude in Science,* Penguin Books
18. Stadler, F. (ed.) (2004) *Induction and Deduction in the Sciences*, Springer
19. Douven, I. (2011) In *Abduction. Stanford Encyclopedia of Philosophy* (Zalta, E.N., ed.), Stanford University
20. Doncescu, A. *et al.* (2007) Knowledge based discovery in systems biology using CF-induction. In *New Trends in Applied Artificial Intelligence* (Okuno, H.G. and Ali, M., eds), pp. 395–404, Springer
21. Papatheodorou, I. *et al.* (2005) Inference of gene relations from microarray data by abduction. In *Logic Programming and Nonmonotonic Reasoning* (Baralet, C. *et al.*, eds), pp. 389–393, Springer
22. Welch, M. *et al.* (2009) Design parameters to control synthetic gene expression in *Escherichia coli*. *PLoS ONE* 4, e7002
23. Gustafsson, C. *et al.* (2012) Engineering genes for predictable protein expression. *Protein Expr. Purif.* 83, 37–46
24. Verhaak, R.G. *et al.* (2010) Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1 EGFR, and NF1. *Cancer Cell* 17, 98–110
25. Tropsha, A. (2010) QSAR in drug discovery. In *Drug Design: Structure- and Ligand-Based Approaches* (Merz, K.M. *et al.*, eds), pp. 151–164, Cambridge University Press
26. Yip, K.Y. *et al.* (2013) Machine learning and genome annotation: a match meant to be? *Genome Biol.* 14, 205
27. Libbrecht, M.W. and Noble, W.S. (2015) Machine learning applications in genetics and genomics. *Nat. Rev. Genet.* 16, 321–332
28. Govindarajan, S. *et al.* (2015) Mapping of amino acid substitutions conferring herbicide resistance in wheat glutathione transferase. *ACS Synth. Biol.* 4, 221–227
29. Lee, M.E. *et al.* (2013) Expression-level optimization of a multi-enzyme pathway in the absence of a high-throughput assay. *Nucleic Acids Res.* 41, 10668–10678
30. Mutalik, V.K. *et al.* (2013) Quantitative estimation of activity and quality for collections of functional genetic elements. *Nat. Methods* 10, 347–353
31. Jonsson, J. *et al.* (1991) A multivariate representation and analysis of DNA sequence data. *Acta Chem. Scand.* 45, 186–192
32. McDaniel, R. *et al.* (1999) Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel "unnatural" natural products. *Proc. Natl. Acad. Sci. U.S.A.* 96, 1846–1851
33. Ness, J.E. *et al.* (2002) Synthetic shuffling expands functional protein diversity by allowing amino acids to recombine independently. *Nat. Biotechnol.* 20, 1251–1255
34. Paddon, C.J. *et al.* (2013) High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature* 496, 528–532
35. Karr, J.R. *et al.* (2014) WholeCellSimDB: a hybrid relational/HDF database for whole-cell model predictions. *Database (Oxford)* 2014, bau095
36. Begley, C.G. and Ellis, L.M. (2012) Drug development: raise standards for preclinical cancer research. *Nature* 483, 531–533
37. Prinz, F. *et al.* (2011) Believe it or not: how much can we rely on published data on potential drug targets? *Nat. Rev. Drug. Discov.* 10, 712
38. Fanelli, D. (2010) Do pressures to publish increase scientists' bias? An empirical support from US States Data. *PLoS ONE* 5, e10271
39. Open Science Collaboration (2015) PSYCHOLOGY. Estimating the reproducibility of psychological science. *Science* 349, aac4716
40. Gardner, T. (2014) A swan in the making. *Science* 345, 855
41. Landis, S.C. *et al.* (2012) A call for transparent reporting to optimize the predictive value of preclinical research. *Nature* 490, 187–191
42. Galdzicki, M. *et al.* (2014) The Synthetic Biology Open Language (SBOL) provides a community standard for communicating designs in synthetic biology. *Nat. Biotechnol.* 32, 545–550
43. Beal, J. *et al.* (2015) A worldwide baseline for reproducibility in synthetic biology. In *Synthetic Biology: Engineering, Evolution & Design (SEED)*. American Institute of Chemical Engineers
44. Oftedal, G. and Parkkinen, V.P. (2013) Synthetic biology and genetic causation. *Stud. Hist. Philos. Biol. Biomed. Sci.* 44, 208–216
45. Riffyn (2015) *Process Design and Analytics Software for R&D Quality & Collaboration,* Riffyn (http://riffyn.com/)
46. DNA2.0 Inc (2015) *DNAatlas – Your Sequences. Rediscovered,* DNA2.0 Inc. (https://www.dnaatlas.com/)
47. DIYBIO (2015) *An Institution for the Do-It-Yourself Biologist,* DIYBIO (http://diybio.org/)
48. Kwok, R. (2010) Five hard truths for synthetic biology. *Nature* 463, 288–290
49. Smith, K. (2013) Synthetic biology: a utilitarian perspective. *Bioethics* 27, 453–463
50. Heavey, P. (2013) Synthetic biology ethics: a deontological assessment. *Bioethics* 27, 442–452
51. ter Meulen, V. (2014) Time to settle the synthetic controversy. *Nature* 509, 135