Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Systems Biology

Measurements drive progress in directed evolution for precise engineering of biological systems

Drew S. Tack, Eugenia F. Romantseva, Peter D. Tonner, Abe Pressman, Jayan Rammohan and Elizabeth A. Strychalski

Abstract

Precise engineering of biological systems requires quantitative, high-throughput measurements, exemplified by progress in directed evolution. New approaches allow high-throughput measurements of phenotypes and their corresponding genotypes. When integrated into directed evolution, these quantitative approaches enable the precise engineering of biological function. At the same time, the increasingly routine availability of large, high-quality data sets supports the integration of machine learning with directed evolution. Together, these advances herald striking capabilities for engineering biology.

Addresses

National Institute of Standards and Technology, Gaithersburg, MD, 20898, USA

Corresponding author: Strychalski, Elizabeth A (elizabeth.strychalski@nist.gov)

Current Opinion in Systems Biology 2020, **23**:32–37

This review comes from a themed issue on **Synthetic Biology (2020)**

Edited by **Matteo Barberis** and **Tom Ellis**

For complete overview of the section, please refer the article collection - Synthetic Biology (2020)

Available online 20 September 2020

https://doi.org/10.1016/j.coisb.2020.09.004

2452-3100/Published by Elsevier Ltd.

Keywords

Engineering biology, Directed evolution, Automation, Machine learning, Metrology, Fitness landscapes.

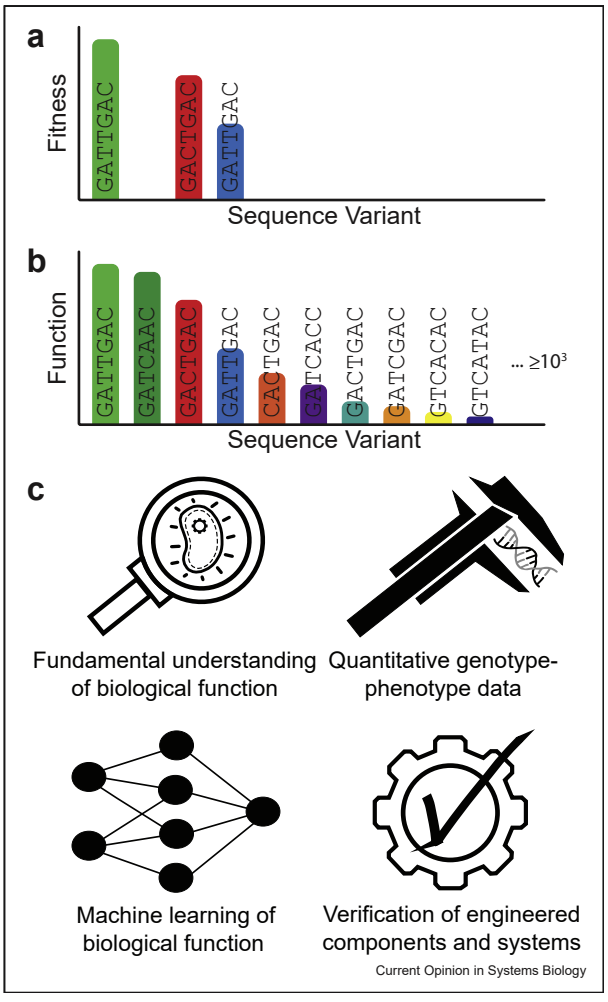
Introduction

Measurements are foundational to engineering biology. The design–build–test–learn workflow at the heart of engineering biology relies on meaningful measurements to quantify, design, predict, and verify the performance of biological components and systems [1,2]. Measurements validate the success of an engineered function, find the limits of performance, and provide a basis for reproducibility and interoperability [3–5]. Many current and proposed applications of engineering biology, such as living therapeutics [6,7], bioremediation [8], and materials manufacturing [9], depend on functions with precisely defined performance [10]. Engineering such specified functions requires measurements to quantitatively characterize the performance of biological parts and systems.

Directed evolution is often the most effective approach to engineering biology (Figure 1) [11,12]. Today, directed evolution uses artificial selections to optimize biological functions, for example, increasing enzymatic activity, binding affinity, or fluorescence signal of an existing biological construct. Directed evolution begins by building libraries of sequence variants, frequently through mutagenesis of a natural gene or organism. Artificial selections then enrich sequence variants with the desired function, often by connecting the target function to the growth rate of a cell, although many other methods are routinely used, such as fluorescence-activated cell sorting [13]. After selection, tens — or, more ambitiously, hundreds — of surviving sequence variants are quantitatively characterized and evaluated against the engineering goal. While directed evolution indirectly optimizes a biological function by connecting the desired function to organismal fitness, engineering biological functions with precise specifications will require high-throughput measurements (HTMs) that provide quantitative information on the performance of many individual sequence variants.

HTMs have proven their utility over the past twenty years, for example, with the widespread and routine implementation of flow cytometry to quantify the phenotypes of cells. More recent innovations in sequencing render HTMs of phenotypes and their associated genotypes accessible. These innovative approaches can advance precise engineering of biological function. In particular, deep sequencing has been adapted to directly quantify the performance of 10^6 protein variants [14–16]. This allows thorough characterization of evolutionary paths, more predictive design of biomolecules, and general consideration of a more complete suite of engineering options from a given biological system than previously accessible. Arriving at an optimal design then becomes a much more tractable task or achievable at all. Common examples of HTMs for this purpose include

Figure 1



High-throughput measurements (HTMs) integrate into engineering biology workflows to enable precise engineering of biological function: **(a)** Directed evolution selects a limited subset of sequence variants with desirable properties from a library of diverse sequence variants. After an artificial selection, only a few sequence variants are evaluated, leaving most sequence variants uncharacterized. **(b)** HTMs enhance directed evolution by measuring each member of a library for its ability to meet the engineering goal. **(c)** The comprehensive information resulting from HTMs offers many benefits toward precise engineering of biological function.

deep sequencing of sorted populations following fluorescence-activated cell sorting (sort-seq [17,18]), measuring changes to the relative abundance of DNA sequences during growth (bar-seq [19–22], and deep mutational scanning [23–25]). Multiple HTMs may also be combined to relate phenotypic performance to the genotypic context of the cellular host. For example, RNA-seq combined with ribo-seq quantifies both transcriptional and translational kinetics, which uncovered previously unknown mechanisms of cellular burden [26]. Given the challenges imposed by biological complexity, HTMs are uniquely suited to capturing the heterogeneity of a biomolecular or cellular population. The resulting data relating genotypes to phenotypes can inform workflows to achieve increasingly sophisticated engineered function [15,27].

HTMs enhance directed evolution for precise engineering of biological systems

Integrating HTMs into directed evolution gives unprecedented insights into the process for engineering biological systems, as well as actionable information to optimize both the workflow and engineered product. HTMs can characterize the genotype and phenotype of every sequence variant within a library to overcome common limitations, such as the exponential distribution in sequence diversity that typically occurs during library construction [28], and enable new capabilities, such as identifying the genotypes of poorly performing sequence variants. Typically, the diversity in a library is assayed only minimally, for example, by sequencing a few sequence variants at random before a selection. These minimal measurements fail to illuminate the diversity within the library and only guarantee that the most grievous potential biases, such as significant imbalances of nucleotides, were avoided during library construction. Instead, measuring the size and diversity of the entire library before selection can identify biases that occur during library construction, and these biases can be accounted for downstream in the directed evolution workflow. In this way, HTMs ensure that the required genetic and phenotypic diversity exists within a library at the outset of a directed evolution experiment. Monitoring changes in the diversity throughout experimental evolution yields the relative enrichment, and thus the functional success, of each sequence variant in a population, including exceptionally rare genotypes often lost during bottlenecking [15,29]. The wealth of information collected from HTMs of an entire library can inform more optimal conditions for further selections, shorten the path to reach the engineering target, and increase the likelihood of success. More broadly, directed evolution with HTMs will increase our fundamental understanding of biology and expand the scope and applicability of future efforts in engineering biology.

HTMs enable the precise engineering of biological systems by quantitatively characterizing the function of biomolecules *in vitro* and decoupled from cellular fitness. An effective engineering biology workflow alters the biophysical properties of a biomolecule, for example, by delivering a specified reaction rate, binding affinity, or metabolic flux. However, conventional methods to assay these properties are slow and laborious [30,31], often requiring the purification of individual biomolecules, and allow for the processing of only a few samples. HTMs speed the engineering workflow considerably by directly and quantitatively assaying

sample function *in vitro*. Early practical examples include repurposed flow cells from deep sequencing platforms to evaluate nucleic acid binding [32], while more recent demonstrations applied a similar approach to quantify kinetic properties [33,34]. By directly characterizing the kinetics of many functional biomolecules, rather than their associated selection fitness, HTMs circumvent the need to tie the function of a biomolecule to the fitness of an organism or some other reproductive advantage [32,35—37]. In this way, HTMs enable new capabilities in the engineering workflow, as well as the ability to effectively engineer a broader range of targets and associated functions.

An illustrative demonstration of enzyme engineering and directed evolution with high-throughput measurements identified enzymes that were both highly active and selective toward a target substrate [38]. Typical selections for this conversion are inefficient, often selecting for highly active enzymes with poor substrate specificity. By using HTMs to evaluate every enzyme in a library for activity toward both the original substrate and the new substrate, the researchers could simply choose the optimal balance between these two design parameters [38]. The activity and specificity of thousands of different enzymes were characterized, and further evaluation of a few chosen enzymes confirmed the effectiveness of this approach. In another recent example, HTMs quantified the function of nearly 10^6 allosteric transcription factors [15]. This allowed the researchers to choose transcription factors targeting three different functional parameters with quantitative specifications, representing unprecedented control over allosteric function [15]. Importantly, HTMs enabled precise engineering of biomolecules in both studies, by achieving multiple engineering design goals simultaneously.

HTMs allow integration of machine learning into directed evolution

HTMs increase the utility and practicality of machine learning (ML) approaches to engineering biology. HTMs provide information crucial to construct predictive models of biological function. Simple, linear ML models trained on small, low-throughput data sets produce biased predictions that underestimate the effect of genetic substitutions on the phenotype, limiting the potential for the rational design of new function [39]. In contrast, ML models trained on HTMs can efficiently explore vast spaces for genetic design and model generation [40]. The large data sets resulting from HTMs are also required to train deep neural networks (DNNs) [41], which readily capture nonlinear effects of multiple genetic substitutions and enable the rational design of biological parts on a scale orders of magnitude larger than previously possible [42]. In addition, while directed evolution gains information on high-performance sequence variants, HTMs capture quantitative information on all sequence variants, even those that perform poorly. Overall, HTMs can constitute comprehensive, unbiased training sets that sample a usefully large space relating genotypes to phenotypes, and ML methods, such as DNNs, trained on these data promise to better predict engineered function.

ML techniques can also increase the value of HTMs in directed evolution. When searching a large space of sequence variants, ML approaches can learn lower dimensional representations of a library (e.g. embedding models) to achieve a practical reduction in the design space [43,44]. These approaches effectively display complex biological data to drive intuition-guided experiments and inform library design for directed evolution. These approaches also provide meaningful insights into the underlying biochemical mechanisms driving evolution by representing similarity between function through distance in the embedding space. This has particular value when there are multiple design objectives or when new or novel categories of phenotypes are desired or unexpectedly arise [15]. DNNs often provide only black-box predictions, offering no insight into the underlying biophysical processes. Understanding and interpreting these predictions mechanistically from a biomolecular perspective, however, will become increasingly valuable to engineering efforts. Demonstrated approaches for interpretable DNN models include embedding DNN models inside biophysical models [45] and using post-hoc approaches to identify relevant motifs that contribute heavily to a model's predictions [46]. These mechanistic insights then augment the ability to design new functions for engineered biological systems. The growing use of HTMs, together with the development of new ML approaches, is drastically improving our ability to engineer biology; we expect massive and synergistic gains in the rate at which new functions are engineered.

Conclusions and outlook

HTMs are well matched to our current understanding of inherently complex biological systems and so are well poised to advance engineering biology with directed evolution. Still, a need persists for measurement assurance. Standard materials and methods to evaluate measurement performance of new techniques against established methods would facilitate the development of novel tools relevant to HTMs. Standards for reporting methods and data along with published results [47], as well as the development of data repositories suited to storing and curating HTMs of genotypes and phenotypes [48], can aid comparability and interoperability. These open platforms support confidence in reported results and decrease redundant workflows by informing

other engineering goals. The widespread publication of HTMs may also reveal novel phenotypes beyond the expected capabilities of the biological parts being engineered, increasing the utility of those engineering targets [15].

HTMs are revolutionizing the utility of directed evolution, enabling a new rigor suited to the precise engineering of biology required for applications outside the laboratory. The increase in quantitation in engineering biology dovetails with industrial best practices, where traceable measurements are routinely implemented for quality and control for the burgeoning bioeconomy. The growing use of laboratory automation further complements HTMs and the precise engineering of biological systems, as evidenced by investment in biofoundries coupled to pipelines for large-scale data collection [49]. Although balancing even two design goals remains novel for engineering biology, this infrastructure will advance capabilities in engineering biology toward a better match with the routine achievements of other engineering fields. Rather than approaching an engineering task with rigid and narrow expectations, we become better equipped through measurement to learn from biology. Surveying the landscape of biological function through HTMs may then instill a renewed sense of wonder and possibility, perhaps ultimately freeing us from the limitations of building only what we can imagine.

Disclaimer

Certain commercial equipment, instruments, or materials are identified to adequately specify experimental procedures. Such identification neither implies recommendation nor endorsement by the National Institute of Standards and Technology nor that the equipment, instruments, or materials identified are necessarily the best for the purpose.

Conflict of interest statement

Nothing declared.

Acknowledgements

This work was supported by the National Research Council Postdoctoral Research Associateships to PDT and AP. The authors thank David Ross and Sasha Levy for constructive review.

References

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Beal J, Haddock-Angelli T, Farny N, Rettberg R: **Time to get serious about measurement in synthetic biology**. *Trends Biotechnol* 2018, **36**:869–871.
Discusses the importance of measurements in transitioning biology to an engineering discipline.

2. Endy D: **Foundations for engineering biology**. *Nature* 2005, **438**:449–453.
Provides a seminal overview of engineering biology as an emerging field, highlighting the importance of "standardization" of biological part performance, "decoupling" DNA synthesis from living organisms, and "abstraction" for defining how parts can be combined into complex devices.

3. • Tack D, Tonner P, Musteata E, Paralanov V, Ross D: *Improved stability of an engineered function using adapted bacterial strains*. bioRxiv; 2020.
Examines the performance limitations of an engineered function using HTMs to monitor activity at the single-cell level. Discovered performance varied between bacterial strains, but the performance in a poorly performing strain could be improved by adapting the strain to the growth environment before encoding the engineered function.

4. Nielsen AAK, Der BS, Shin J, Vaidyanathan P, Paralanov V, Strychalski EA, Ross D, Densmore D, Voigt CA: **Genetic circuit design automation**. *Science* 2016, **352**:aac7341.
Demonstrates the design and development of a robust programming toolkit for living cells, with an attempt to predict the performance of complex programmed functions and the discrepancies between predicted and observed results.

5. Sarkar S, Tack D, Ross D: **Sparse estimation of mutual information landscapes quantifies information transmission through cellular biochemical reaction networks**. *Communications Biology* 2020, **3**:1–8.
Describes a method to measure the performance of biological systems in mutual information, a population-based metric with broad utility and comparability.

6. Landry BP, Tabor JJ: **Engineering diagnostic and therapeutic gut bacteria**. *Microbiol Spectr* 2017, **5**.
Reviews efforts towards engineering gut microbes for diagnostic and therapeutic purposes, including considerations of the reliability of engineered constructs and metrics to evaluate engineered cells for this purpose.

7. Riglar DT, Silver PA: **Engineering bacteria for diagnostic and therapeutic applications**. *Nat Rev Microbiol* 2018, **16**:214–225.
Reviews technologies and approaches to engineering living cells for use as diagnostic and therapeutic devices.

8. Dvořák P, Nikel PI, Damborský J, de Lorenzo V: **Bioremediation 3.0: engineering pollutant-removing bacteria in the times of systemic biology**. *Biotechnol Adv* 2017, **35**:845–866.
Reviews engineering biology efforts and approaches to bioremediation, focusing on pathway engineering to catabolize pollutants.

9. Cao Y, Feng Y, Ryser MD, Zhu K, Herschlag G, Cao C, Marusak K, Zauscher S, You L: **Programmable assembly of pressure sensors using pattern-forming bacteria**. *Nat Biotechnol* 2017, **35**:1087–1093.
Exemplifies materials synthesis, by engineering bacteria to structurally form a pressure-sensitive material.

10. Charbonneau MR, Isabella VM, Li N, Kurtz CB: **Developing a new class of engineered live bacterial therapeutics to treat human diseases**. *Nat Commun* 2020, **11**:1–11.
Considers engineered biological systems for use as living therapeutics, both practically and from a regulatory perspective, including the need for better metrics to evaluate both the biological systems themselves and their engineered performance.

11. Zeymer C, Hilvert D: **Directed evolution of protein catalysts**. *Annu Rev Biochem* 2018, **87**:131–157.
Reviews strategies for the directed evolution of enzymes, including library design and construction, as well as artificial selections to identify desirable function.

12. Giese B, Koenigstein S, Wigger H, Schmidt JC, von Gleich A: **Rational engineering principles in synthetic biology: a framework for quantitative analysis and an initial assessment**. *Biological Theory* 2013, **8**:324–333.
Reviews engineering biology as a field focused on approaches to engineer with predictable results from rationally designed sequence variants.

13. Sanchez MI, Ting AY: **Directed evolution improves the catalytic efficiency of TEV protease**. *Nat Methods* 2020, **17**:167–174.
Exemplifies directed evolution of enzymatic function integrating FACS and resulting in proteases with increased catalytic activity.

14. Starita LM, Pruneda JN, Lo RS, Fowler DM, Kim HJ, Hiatt JB, Shendure J, Brzovic PS, Fields S, Klevit RE: **Activity-enhancing**

mutations in an E3 ubiquitin ligase identified by high-throughput mutagenesis. In *Proceedings of the national academy of the sciences*, vol. 110; 2013. E1263–E1272.
Implements HTMs to measure the enzymatic associated with nearly 10^6 different sequence variants.

15. •• Tack DS, Tonner PD, Pressman A, Olson ND, Levy SF, Romantseva EF, Alperovich N, Vasilyeva O, Ross D: *The genotype-phenotype landscape of an allosteric protein*. bioRxiv; 2020.
Describes the use of HTMs to comprehensively characterize nearly 10^5 allosteric transcription factors. This enabled the precise engineering of genetic sensors with quantitatively targeted functional parameters. Additionally, comprehensive evaluation of many transcription factors enabled the discovery of transcription factors with novel functions.

16. Sarkisyan KS, Bolotin DA, Meer MV, Usmanova DR, Mishin AS, Sharonov GV, Ivankov DN, Bozhanova NG, Baranov MS, Soylemez O, *et al.*: **Local fitness landscape of the green fluorescent protein**. *Nature* 2016, **533**:397–401.
Describes the HTM of protein function, by using DNA-barcoding to explore variations along the full length of a protein, and thus capture a basic distribution of epistasis.

17. Peterman N, Levine E: **Sort-seq under the hood: implications of design choices on large-scale characterization of sequence-function relations**. *BMC Genom* 2016, **17**:206.
Considers design and implementation of Sort-Seq in engineering biology workflows, including library design and modeling.

18. Rohlhill J, Sandoval NR, Papoutsakis ET: **Sort-seq approach to engineering a formaldehyde-inducible promoter for dynamically regulated Escherichia coli growth on methanol**. *ACS Synth Biol* 2017, **6**:1584–1595.
Implements Sort-Seq in bacterial promoter engineering. Resulting genotype-phenotype data was used to design promoters with predictable activity.

19. Smith AM, Heisler LE, Mellor J, Kaper F, Thompson MJ, Chee M, Roth FP, Giaever G, Nislow C: **Quantitative phenotyping via deep barcode sequencing**. *Genome Res* 2009, **19**:1836–1842.
Describes Bar-Seq, a method implementing DNA barcodes to track the performance of cells and biomolecules using deep sequencing.

20. Schlecht U, Liu Z, Blundell JR, StOnge RP, Levy SF: **A scalable double-barcode sequencing platform for characterization of dynamic protein-protein interactions**. *Nat Commun* 2017, **8**:15586.
Describes a novel method to combinatorially assemble genetic diversity to search for and evaluate protein–protein interactions in living cells.

21. Levy SF, Blundell JR, Venkataram S, Petrov DA, Fisher DS, Sherlock G: **Quantitative evolutionary dynamics using high-resolution lineage tracking**. *Nature* 2015, **519**:181–186.
Describes an approach to quantitatively evaluate cellular fitness and evolutionary dynamics, by tracking genome-encoded DNA barcodes associated with known genotypes.

22. • Roy KR, Smith JD, Vonesch SC, Lin G, Tu CS, Lederer AR, Chu A, Suresh S, Nguyen M, Horecka J, *et al.*: **Multiplexed precision genome editing with trackable genomic barcodes in yeast**. *Nat Biotechnol* 2018, **36**:512–520.
Describes a method to make multiple changes in a yeast genome, while recording the changes that were made by simultaneously assembling a DNA barcode, which then facilitates the evaluation of cellular fitness associated with specific genomic changes.

23. Faber MS, Whitehead TA: **Data-driven engineering of protein therapeutics**. *Curr Opin Biotechnol* 2019, **60**:104–110.
Reviews the role of data, often from HTMs, in advancing the development of protein therapeutics, focusing on deep mutational scanning in enzyme and antibody engineering.

24. Klesmith JR, Bacik J-P, Wrenbeck EE, Michalczyk R, Whitehead TA: **Trade-offs between enzyme fitness and solubility illuminated by deep mutational scanning**. In *Proceedings of the national academy of the sciences*, vol. 114; 2017: 2265–2270.
Investigates enzyme activity and solubility with deep mutational scanning to conclude that the two are interconnected and cannot be manipulated completely independently.

25. Wrenbeck EE, Faber MS, Whitehead TA: **Deep sequencing methods for protein engineering and design**. *Curr Opin Struct Biol* 2017, **45**:36–44.
Reviews deep sequencing approaches to engineering biology.

26. • Gorochowski TE, Chelysheva I, Eriksen M, Nair P, Pedersen S, Ignatova Z: **Absolute quantification of translational regulation and burden using combined sequencing approaches**. *Mol Syst Biol* 2019, **15**, e8719.
Combines RNA-seq and Ribo-seq for simultaneous quantification of transcriptional and translational flux. The combination of two HTMs revealed ribosome sequestration triggers numerous mechanisms of cellular burden.

27. Blanco C, Janzen E, Pressman A, Saha R, Chen IA: **Molecular fitness landscapes from high-coverage sequence profiling**. *Annu Rev Biophys* 2019, **48**:1–18.
Comprehensively reviews fitness landscape measurements and HTMs, including deep mutational scanning, HTS analysis of kinetic function, changing environments, and theoretical and unconventional fitness landscapes.

28. Neylon C: **Chemical and biochemical strategies for the randomization of protein encoding DNA sequences: library construction methods for directed evolution**. *Nucleic Acids Res* 2004, **32**:1448–1459.
Describes methods to construct sequence variant libraries for use in directed evolution, and challenges in these methods.

29. Venkataram S, Dunn B, Li Y, Agarwala A, Chang J, Ebel ER, Geiler-Samerotte K, Hérissant L, Blundell JR, Levy SF, *et al.*: **Development of a comprehensive genotype-to-fitness map of adaptation-driving mutations in yeast**. *Cell* 2016, **166**:1585–1596.e22.
Describes a precise HTM of genome mutations and their associated fitness effect for a yeast strain and classifying mutations that increase fitness in the experimental conditions.

30. Pollard TD: **A guide to simple and informative binding assays**. *Mol Biol Cell* 2010, **21**:4061–4067.
Describes classical methods to measure binding affinities of biomolecules.

31. Bisswanger H: **Enzyme assays**. *Perspectives in* Science 2014, **1**:41–55.
Describes classical methods to evaluate enzymatic activity and characterize the kinetic properties of biomolecules.

32. Nutiu R, Friedman RC, Luo S, Khrebtukova I, Silva D, Li R, Zhang L, Schroth GP, Burge CB: **Direct measurement of DNA affinity landscapes on a high-throughput sequencing instrument**. *Nat Biotechnol* 2011, **29**:659–664.
Exemplifies HTM of DNA-binding affinity on a high throughput sequencing instrument.

33. •• Pressman AD, Liu Z, Janzen E, Blanco C, Müller UF, Joyce GF, Pascal R, Chen IA: **Mapping a systematic ribozyme fitness landscape reveals a frustrated evolutionary network for self-aminoacylating RNA**. *J Am Chem Sci* 2019, **141**:6213–6223.
Exemplifies HTMs for the relative and quantitative characterization of catalytic activity. HTMs were used to select and characterize ribozymes with autocatalytic activity from a library of oligonucleotides, revealing several distinct evolutionary solutions. Additionally, the HTM of activity quantitatively characterized kinetic properties of nearly 107 sequence variants.

34. Jalali-Yazdi F, Lai LH, Takahashi TT, Roberts RW: **High-throughput measurement of binding kinetics by mRNA display and next-generation sequencing**. *Angew Chem Int Ed* 2016, **55**:4007–4010.
Exemplifies the high throughput quantification of binding kinetics between a peptide and 20,000 RNA sequences.

35. Adams RM, Mora T, Walczak AM, Kinney JB. *Measuring the sequence-affinity landscape of antibodies with massively parallel titration curves*, vol. 5. e*Life*; 2016.
Demonstrates the use of multiple HTMs to quantitatively characterize the binding properties of an antibody.

36. Buenrostro JD, Araya CL, Chircus LM, Layton CJ, Chang HY, Snyder MP, Greenleaf WJ: **Quantitative analysis of RNA-protein interactions on a massively parallel array reveals biophysical and evolutionary landscapes**. *Nat Biotechnol* 2014, **32**:562–568.

Exemplifies repurposing a high throughput sequencing instrument to study binding and dissociation of interacting biomolecules.

37. Guenther U-P, Yandek LE, Niland CN, Campbell FE, Anderson D, Anderson VE, Harris ME, Jankowsky E: **Hidden specificity in an apparently nonspecific RNA-binding protein**. *Nature* 2013, **502**:385–388.
Discovers subtle protein properties, detectable only when characterized with HTMs, to study binding to thousands of different RNA sequences.

38. Zhang MS, Brunner SF, Huguenin-Dezot N, Liang AD, Schmied WH, Rogerson DT, Chin JW: **Biosynthesis and genetic encoding of phosphothreonine through parallel selection and deep sequencing**. *Nat Methods* 2017, **14**:729–736.
Demonstrates the use of parallel HTMs to simultaneously explore multiple properties of a biomolecule, including specificity and activity.

39. Otwinowski J, Plotkin JB: **Inferring fitness landscapes by regression produces biased estimates of epistasis**. *Proc Natl Acad Sci* 2014, **111**:E2301–E2309.
Establishes inherent bias of classical linear models as applied to modeling fitness landscapes, motivating the need for more robust, nonlinear and higher-throughput approaches.

40. Yang KK, Wu Z, Arnold FH: **Machine-learning-guided directed evolution for protein engineering**. *Nat Methods* 2019, **16**:687–694.
Reviews the potential of ML methods for enhancing directed evolution by covering a number of case-studies in which ML methods lead to more efficient directed evolution.

41. Eraslan G, Avsec Ž, Gagneur J, Theis FJ: **Deep learning: new computational modelling techniques for genomics**. *Nat Rev Genet* 2019, **20**:389–403.
Reviews the potential of deep learning in genomics, including details of numerous approaches relevant to the field and practical considerations for their deployment.

42. • Kotopka BJ, Smolke CD: **Model-driven generation of artificial yeast promoters**. *Nat Commun* 2020, **11**:2113.
Constructs a convolutional neural network model on promoter expression measured from over 106 sequences. Applies this model for rational design of new biological parts, demonstrating the computational design of biology.

43. • Alley EC, Khimulya G, Biswas S, AlQuraishi M, Church GM: **Unified rational protein engineering with sequence-based deep representation learning**. *Nat Methods* 2019, **16**:1315–1322.
Develops a unified representation of protein sequences in a continuous latent space by training a recurrent DNN on 24 million protein sequences. Demonstrates the utility of large datasets available to the community.

44. • Yang KK, Wu Z, Bedbrook CN, Arnold FH: **Learned protein embeddings for machine learning**. *Bioinformatics* 2018, **34**:2642–2648.
Describes a k-mer based method for embedding protein sequences and validates usefulness in downstream regression tasks.

45. Tareen A, Kinney JB: *Biophysical models of cis-regulation as interpretable neural networks*. arXiv; 2019.
Develops a biophysical model with parameters predicted from underlying neural network components to maximize predictive accuracy and interpretability.

46. Greenside P, Shimko T, Fordyce P, Kundaje A: **Discovering epistatic feature interactions from neural network models of regulatory DNA sequences**. *Bioinformatics* 2018, **34**:i629–i637.
Develops an approach to interpret deep neural network models in genomics that quantifies the epistasis in sequence variants.

47. Hecht A, Filliben J, Munro S, Salit M: **A minimum information standard for reproducing bench-scale bacterial cell growth and productivity**. *Communications Biology* 2018, **1**:219.
Fractional factorial design of experiments is used to identify the most important factors for repeatable and reproducible growth of engineered organisms.

48. •• Esposito D, Weile J, Shendure J, Starita LM, Papenfuss AT, Roth FP, Fowler DM, Rubin AF: **MaveDB: an open-source platform to distribute and interpret data from multiplexed assays of variant effect**. *Genome Biol* 2019, **20**:223.
Describes a database for HTMs of sequence variants. This platform will greatly facilitate the distribution and application of HTMs, including the evaluation against other engineering goals, and the exploration and discovery of novel biological function.

49. Hillson N, Caddick M, Cai Y, Carrasco JA, Chang MW, Curach NC, Bell DJ, Le Feuvre R, Friedman DC, Fu X, *et al.*: **Building a global alliance of biofoundries**. *Nat Commun* 2019, **10**:2040.
Introduces the Global Biofoundry Alliance, a group of laboratories and institutes with automation capabilities focused on collaboration and communication to address societally impactful challenges by engineering biology.