CELL & GENE THERAPY INSIGHTS

VECTOR CHARACTERIZATION & VALIDATION

SPOTLIGHT

EXPERT INSIGHT

Challenges and opportunities of machine-guided capsid engineering for gene therapy

Eric D Kelsic & George M Church

Advances in DNA delivery will be crucial to the enablement of new gene therapies. Pioneering efforts in developing recombinant Adeno-Associated Virus (AAV) technology for gene delivery have led to a recent wave of treatments for genetic diseases with great unmet need. Most current therapies use the capsids of AAV isolated from Adenovirus stocks or primary human and primate tissues, that were then discovered to have favorable tropism, immunogenicity and manufacturability properties. However, despite much work invested in the engineering of new capsids for enhanced delivery, many delivery targets remain out of reach. In our view, high-throughput capsid engineering could be done more effectively by combining three advanced technologies in a closed-loop manner: i. next-gen library synthesis, ii. next-gen sequencing, and iii. machine learning. This approach enables a machine-guided data-driven workflow in which the search for improved capsids is dramatically accelerated relative to traditional open-loop methods. In this report, we review the technological advances that are pushing the field of AAV capsid engineering toward machine-guided methods, describe and explore the promise of this new approach, and discuss anticipated challenges. In the near future, machine-guided approaches will revolutionize our ability to design safe, targeted, delivery tools for the treatment of genetic conditions.

Submitted for Peer Review: Feb 26 2019 ▶ **Published**: May 16 2019

Cell & Gene Therapy Insights 2019; 5(4), 523–536

DOI: 10.18609/cgti.2019.058

While we currently understand the molecular basis for thousands of genetic conditions, most patients with such diseases lack any satisfactory treatment options. Gene therapy proposes to rescue mutated and dysfunctional genes by adding a new functional copy, or by repairing a genetic defect through editing of the genome. When it works, gene therapy can therefore be a one-dose cure for such conditions. *In vivo* gene therapies comprise a delivery component and a payload component, the genetic payload being delivered to a patient's cells via direct administration into the body. Excellent reviews have been written summarizing advances in payload translation and development [1,2]. In this report we focus on the delivery component, in particular on viral capsids from the Adeno-Associated Virus (AAV), as a promising application for new approaches to machine-guided protein engineering. Optimizing AAV capsids as delivery vectors will have great translational impact (Figure 1). Defining the goal precisely, an ideal *in vivo* delivery vector would be highly capable at delivering its payload to precisely the cells that need it and no others. It would work robustly independent of the immune history of each patient (no pre-existing immunity), and be well tolerated by the immune system, provoking no immune reaction that would eliminate the therapy or that might prevent re-administration of another dose or a second therapy. It would be cheap and easy to manufacture. And finally, it would be helpful to increase the payload size of AAV beyond 4.7kb, perhaps closer to the 5-6kb genomes that are observed in other parvoviruses with similar capsid structures. It is fortunate that AAV capsids accomplish some of these aims, but obviously they were not naturally evolved for this therapeutic task. While today's best AAV capsids are useful tools that have recently led to dramatic life-changing [3] and life-saving treatments [4], today's engineered capsids are still highly similar in sequence to natural isolates and no engineered capsids have reached late stage clinical trials. In our view, this is because screening limitations, the complexity of AAV capsids and uncertainty in basic AAV biology have prevented engineering efforts from generating enhanced capsids that fully meet the needs of today's researchers.

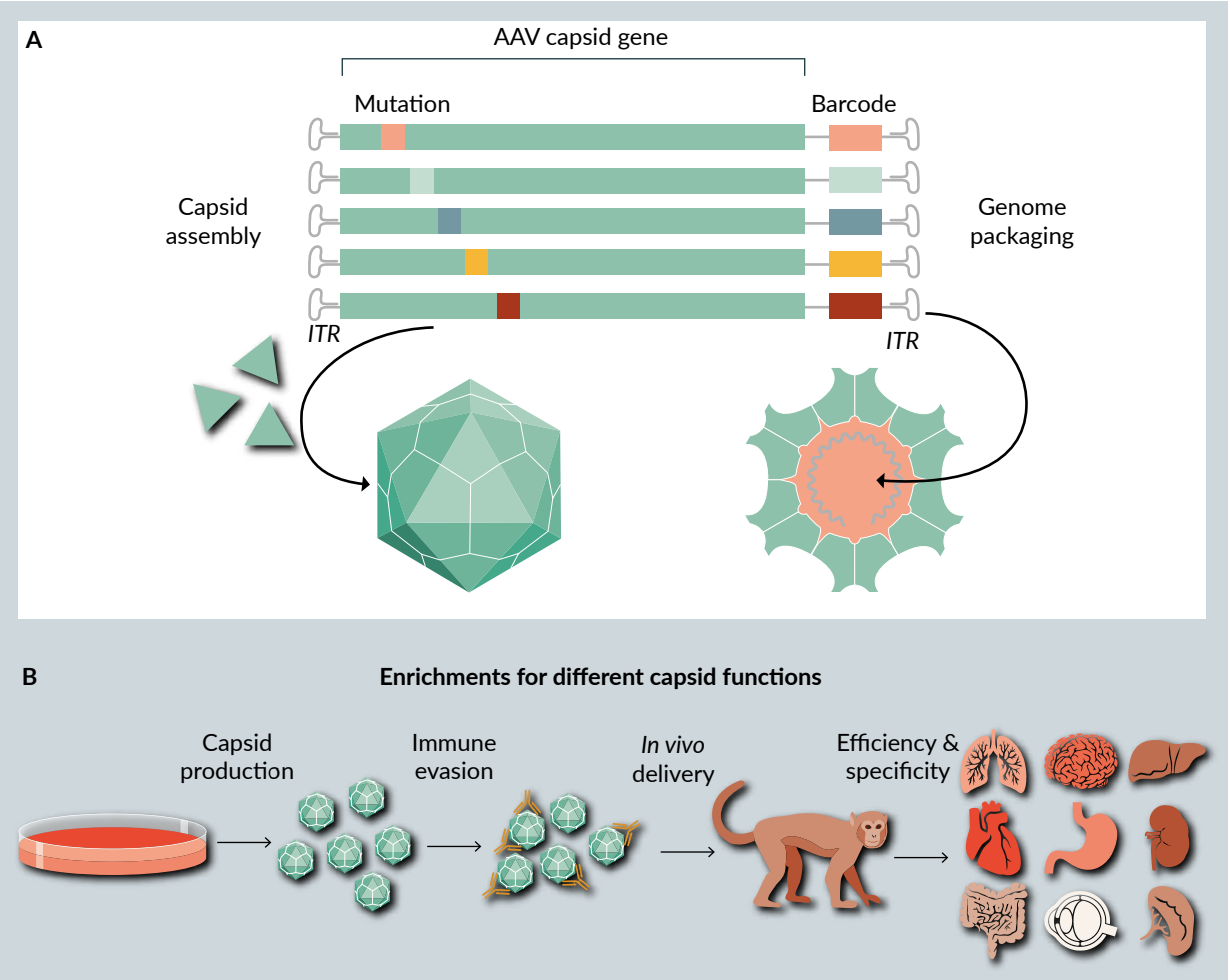
TOWARD MACHINE-GUIDED DESIGN THROUGH CLOSED-LOOP DATA-DRIVEN WORKFLOWS

Early efforts to engineer AAV capsids used random methods to generate diverse libraries, but over time the field has become more data-driven. As with any protein, a fundamental challenge for AAV engineering is that the size of sequence space is extremely large, growing exponentially with the number of mutations introduced. Furthermore, most protein sequence variants are non-functional [5,6]. Based on initial observations from NGS sequencing of barcoded AAV capsid libraries [7], and our own observations [8], we estimate that when a single position of the capsid is modified to a random amino acid that less than one of five mutants will be viable at forming a capsid. This simple benchmark illustrates the challenge of building diverse libraries. If less than 1/5

DOI: 10.18609/cgti.2019.058

FIGURE 1

High-throughput measurement of functional properties of AAV capsids for *in vivo* gene delivery.



A) Measuring capsid function from capsid sequence. The AAV capsid is composed of 60 monomers forming an icosahedral shell which surrounds the ssDNA viral genome. The AAV genome also contains flanking Inverted Terminal Repeat (ITR) sequences that act as packaging signals. Under appropriate conditions capsids produced in library format can be made to carry their own genome. The sequence of each genome can then be read in high-throughput from the mutation itself or from a linked DNA barcode.
B) The library of capsid genes can be enriched for functional variants using assays that select on the capsid protein: viral production, immune evasion, and *in vivo* delivery efficiency and specificity for target tissues.

sequences with a single mutation are viable, then assuming rare epistatic rescue events, less than 1/25 of double mutants and 1/125 of triple mutants will be viable, etc. The conclusion is that as purely random libraries become more diverse that the quality of these libraries decreases exponentially.

This tradeoff between diversity and quality is critical to library design. Over time, new technologies and new data have improved the quality of capsid engineering libraries, with a general trend toward incorporating more and more data into the design process. Closed-loop workflows represent the final stage of this trend. See [9,10] for comprehensive reviews of AAV capsid engineering. Here we highlight a few illustrative works to provide context for what has recently become possible thanks to new technologies.

Capsid engineering efforts to date have usually begun with natural AAV sequences. Starting at a viable location in sequence space, random mutation methods like error-prone PCR can explore the space of adjacent variants [11]. However, due to the exponential drop in viability away from the wild-type (WT), these libraries are restricted to a narrow region of sequence space, and improved variants are then by definition very similar in sequence to WT. These approaches invite the question: are dramatic improvements possible when engineered variants are so similar in sequence to natural capsids? In our view, this is unlikely to occur: solving the needs of the field, especially for applications like avoiding neutralization by antibodies that exist due to prior infection by a natural capsid, will require development of AAV capsids that are dramatically different in sequence from any natural serotype prevalent in human populations. The goal of making more changes to the capsid while maintaining viability has motivated alternative library construction approaches. Toward this aim, early libraries used short peptide insertions at locations on the external projections of the capsid [12,13]. Since external positions are more tolerant of insertions, the viability of these libraries is higher than for error-prone libraries. This incorporation of 3D structural data into library design can be seen as a first step towards a data-driven approach. Another popular method is capsid shuffling [14,15], in which a number of natural genomes are broken into smaller pieces and reassembled into chimeric capsids. By using diverse natural capsids as source material, the desire to make large changes in sequence is balanced by the important requirement of maintaining viability. From a data-driven perspective, decisions on which natural capsids to include and how to shuffle their sequences are informed by the functional properties of those serotypes as well as sequence and structural information.

Driven by advances in DNA synthesis capabilities, library design strategies in recent years have become more sophisticated by incorporating more phylogenetic and structural data. Phylogenetic information has been incorporated into library design in multiple ways. Some libraries focused on shuffling with breakpoints set around the variable regions in evolutionary alignments, so as to access the diversity of these locations while maintaining functional sequences in conserved regions [16]. Ancestral reconstructions enabled the identification of new AAV capsids that are very different in sequence from present serotypes, but that are still functional [17]. Domain grafting represents an alternative approach, in which the sequence and 3D structural differences between two serotypes are compared. This highly informed approach has enabled the construction serotypes with hybrid properties even without large library diversity [18]. Structural data from AAV capsids bound to monoclonal antibodies has also been used to focus libraries on positions where these antibodies bind the capsid surface, generating viable capsids that can better avoid neutralizing serum [19]. While many libraries have focused on random search either through degenerate libraries or combinatorial assembly, methods for high-throughput direct synthesis of short DNA oligos, also called “oligo pools”, offer

DOI: 10.18609/cgti.2019.058

a compelling alternative to random synthesis. Instead of randomly generating diversity, 100,000 or more oligos at 200–300nt in length can be printed in array format and then used as input material for molecular cloning. Early applications of oligo pools included low-cost DNA assembly into full length genes [20] and high-throughput screening of diverse genetic constructs [21,22]. Oligo pools were recently applied to peptide insertion libraries of AAV capsids, wherein rather than using random insertions, the library used short peptides scanning along a set of proteins with known bioactivity – motivated by the desire to enrich for AAV capsids capable of neuronal retrograde transport relative to a purely random approach [23]. This capability of having full control over library composition is particularly exciting, since incorporating more information in the design should yield richer, more diverse libraries while maintaining high viability.

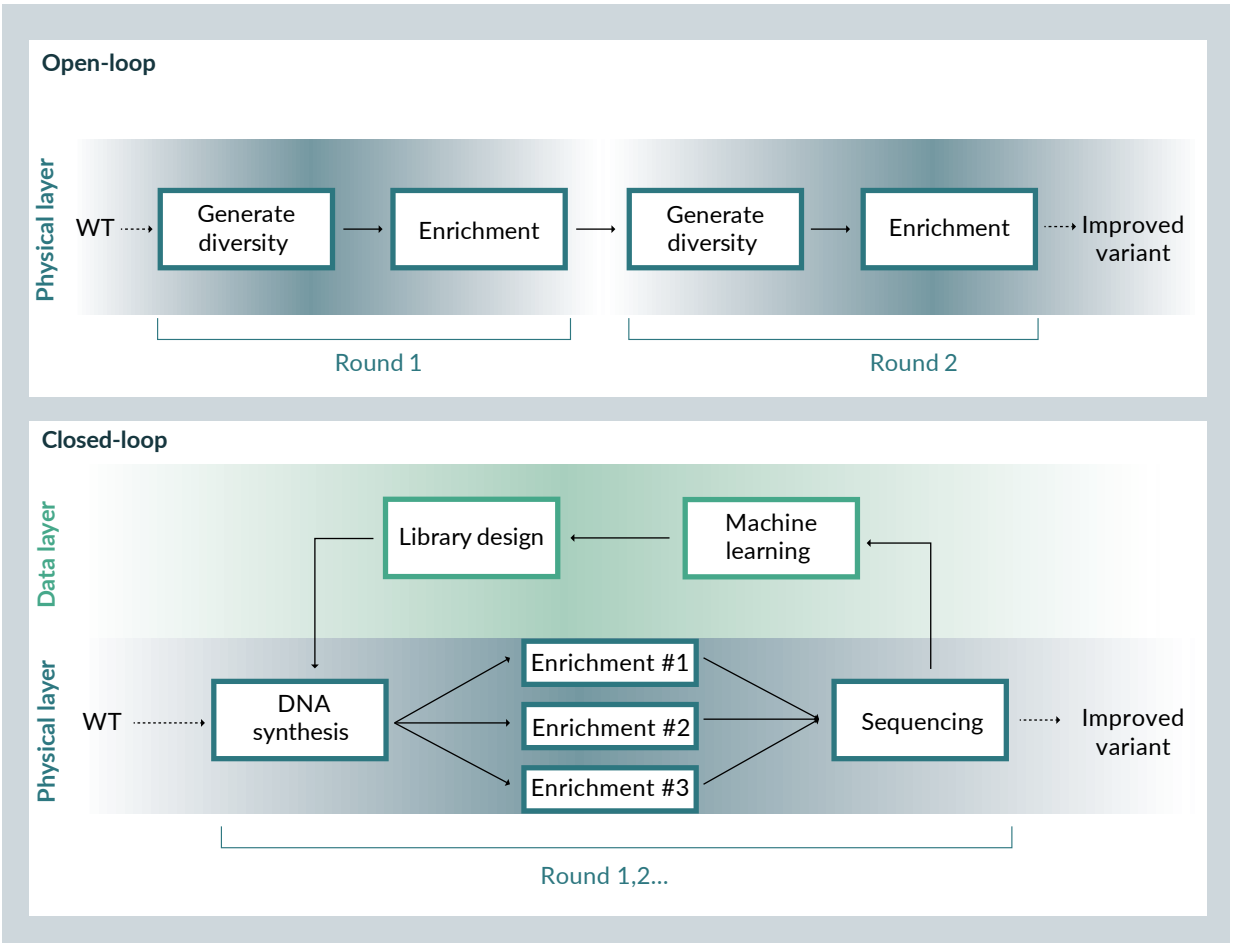
With new high-throughput DNA synthesis capabilities, it becomes possible to shift from an open-loop to a closed-loop data-driven workflow. The library design strategies described above were all open-loop: either being just a single round of screening, or multiple rounds where the source material to each round of engineering might change but, importantly, the *strategy* for generating new diversity remained fixed (usually employing some form of degenerate synthesis or random mutation). With oligo pools offering complete flexibility as to what sequences are tested, a closed-loop workflow becomes possible. By closing the loop, the design of new libraries is primarily informed by the data from previous rounds of screening, maximizing the amount of information incorporated into the design process. This is advantageous mainly due to the complexities and uncertainties of AAV biology. Many aspects of AAV biology are poorly understood: we have only a limited and incomplete view of how AAV capsids interact with cellular receptors and with internal host factors. It is therefore difficult to predict the effect of any given mutation on capsid function. It follows then that for informing an efficient search of sequence space, nothing beats experimental data.

Fortunately, new technologies also provide a way to obtain such data in large quantities: by barcoding libraries of AAV capsids, pooling the libraries and measuring their delivery properties in multiplexed using high-throughput DNA sequencing [7,23]. While these libraries generate a huge amount of data that can be difficult to analyze, new machine learning approaches that are capable of recognizing patterns within large complex datasets are rising to meet this challenge. We recently began experimenting with closed-loop data-driven workflows incorporating high-throughput DNA synthesis to build the libraries, DNA sequencing to measure the results, and machine learning to analyze the data and design future libraries [24] (Figure 2). We expect this machine-guided workflow to become dominant in protein engineering in the near future. For capsid engineering, we expect these technologies to impact the field in three ways:

1. The design of smarter libraries
2. Better selections
3. Optimization across multiple properties

FIGURE 2

Open-loop vs closed-loop engineering workflows.



Top: Open-loop workflows use the same strategy for generating diversity in each round of screening, and all connection between rounds occur within the physical layer.
Bottom: For closed-loop approaches the design occurs in a data layer. Here the strategy for designing new libraries in each round depends strongly on data from prior rounds. Engineering workflows become more closed-loop as more information feeds back on library design. An example closed-loop workflow shows library design using the output of a machine learning model trained on experimental sequencing data across multiple enrichment assays.

SMARTER LIBRARIES

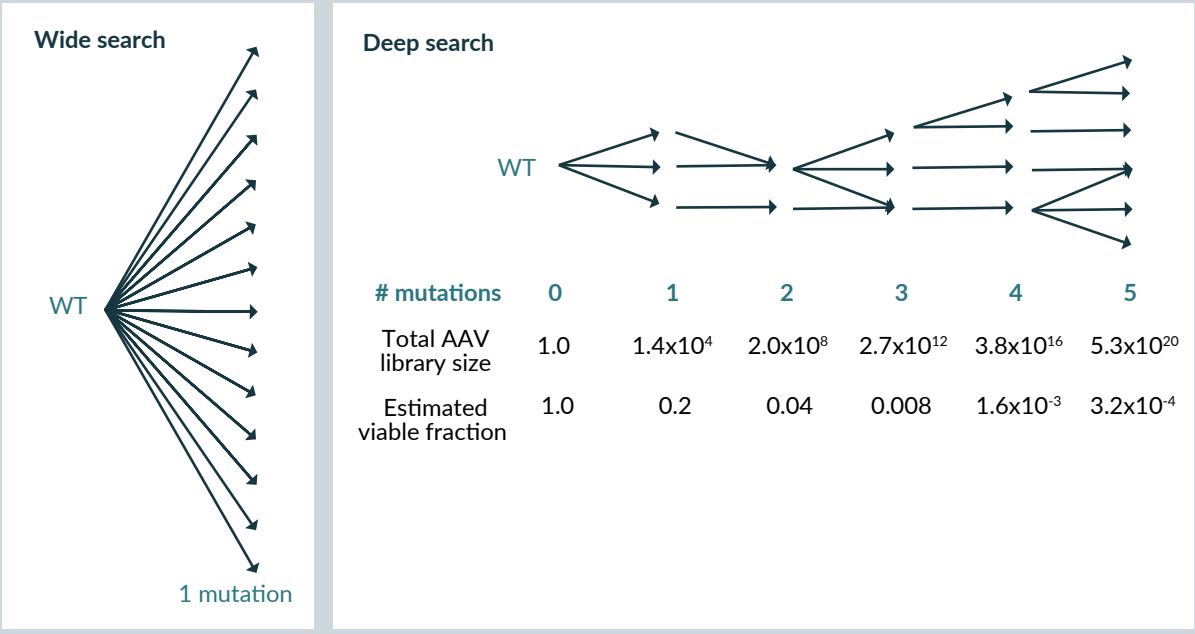
Machine-guided approaches will enable smarter libraries because machine learning approaches are more adept at navigating the variable terrain of sequence space than approaches that rely on randomly generated diversity. Most of sequence space is non-functional, but machine learning models trained on data from prior rounds of screening can learn to identify mutations that are likely to break capsid function and prune these away from future libraries, increasing viability while maintaining high diversity. Such a strategy highlights an important trade-off in the machine-guided workflow between exploration and exploitation. A search strategy must first explore the landscape and learn what is good and what is bad,

DOI: 10.18609/cgti.2019.058

then can pivot to exploiting this knowledge to generate optimized mutants with greater likelihood of improved function. We have constructed libraries toward both aims, and we refer to these search strategies as either "wide" or "deep" (Figure 3). In wide search the emphasis is on exploration, and in our first attempt at a library of this type we generated all possible single amino acid substitutions, insertions and deletions [8]. These libraries are "wide" because the branching factor of the search tree is large and the number of simultaneous mutations per capsid variant typically few, in order to minimize the number of unviable sequences that are created while maximizing learning. Although a comprehensive single-mutation scan is within the scale of current oligo synthesis, comprehension scans for several mutations remain beyond current capabilities. After gathering relevant info on the protein landscape, the next step is a deep search, aiming to combine beneficial mutations together toward generating a highly functional capsid. The selection of deep mutants for synthesis can be informed by a machine learning model, and as more and more combinations of mutations are tested these models begin to learn how pairwise and higher-order interactions between different residues affect function. The follow up to a deep search library is then a better informed and subsequently deeper search. Toward the end of a capsid selection project, all of the resulting learnings should be applied to prioritize and test the most likely sequences to achieve the

FIGURE 3

Wide and deep search strategies.



# mutations	0	1	2	3	4	5
Total AAV library size	1.0	$1.4x10^{4}$	$2.0x10^{8}$	$2.7x10^{12}$	$3.8x10^{16}$	$5.3x10^{20}$
Estimated viable fraction	1.0	0.2	0.04	0.008	$1.6x10^{-3}$	$3.2x10^{-4}$

Wide search emphasizes exploration, using fewer mutations per variant to maximize learning and maintain higher library viability. Deep search emphasizes exploitation, aimed at generating capsids with better function by applying information from prior rounds of screening. The viable fraction of a library is expected to be low when large numbers of mutations are randomly generated. Thus truly random libraries will not be very deep. A simple strategy for generating viable deep libraries, and one that beats a randomly generated library, is to combine just observed beneficial and neutral changes together in various combinations. More complicated machine learning approaches can automate the process of extracting information from deep search data, then further propose new sequences with a higher likelihood of improved function than randomly generated sequences.

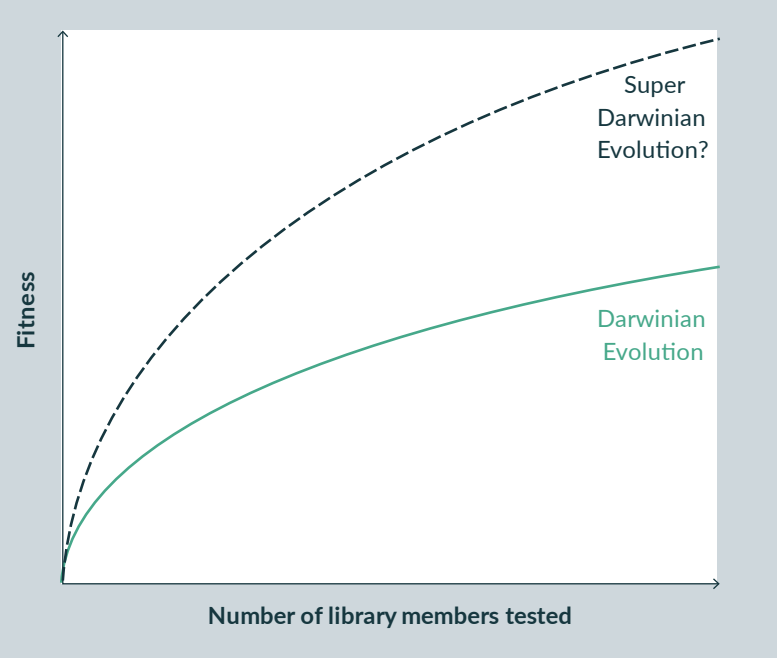
desired function, bottlenecking the population size down to the single variant that will ultimately be validated and used for the therapeutic delivery of a transgene.

An open question is how to optimally balance the tradeoff between exploration and exploitation, and whether new technologies enable strategies that are faster than what has been done before. The most popular search strategy for the last 4 billion years has been that of Darwinian Evolution (DE): the repeated process of

1. Random mutagenesis; followed by
2. Enrichment of beneficial mutations within the population.

FIGURE 4

The possibility of Super Darwinian Evolution.



Darwinian Evolution is the repeated iteration of random diversification and enrichment through selection. This process occurs within the physical layer due to physical constraints on generating mutations and organismal replication. With closed-loop workflows, containing both physical and data layers, these constraints are removed. This enables Super Darwinian algorithms that can increase in fitness faster than Darwinian processes. Plotted are hypothetical trajectories. Many formulations would be possible, for illustrative purposes we imagine plotting the expected best improvement in fitness that would achieved after a given number of variants have been tested.

The popularity of the Darwinian approach was due to two constraints. First, mutations had to be physically generated at the molecular level, typically by error-prone replication, DNA damage, or via recombination. Second, with Darwinian Evolution the frequency of an allele in the population encodes a memory for what is known to be good, so that new mutations are more likely to occur on good backgrounds. In both cases, with *de novo* DNA synthesis capabilities, these constraints are removed. Now the information from a first round of experiments can be used as input to a computer model, and the next generation of mutations generated without any physical link between the molecules from one generation to the next. Likewise, the record of what sequences are good and bad can be stored in computer memory. Removing these constraints, machine-guided workflows suggest the possibility of Super Darwinian Evolution (SDE), which we define as any exploration strategy that beats a Darwinian approach of random mutagenesis and selection on a given fitness landscape (Figure 4). While well-targeted mutagenesis libraries are arguably already Super Darwinian since the mutations are not completely random, these are still open-loop workflows. The greatest gains from SDE will come from closed-loop data-driven workflows as described above. The opportunity for SDE is clear, but it is still unknown what algorithms for SDE will be best suited for AAV and for other proteins as the complex structure of protein fitness landscapes is not well understood. For this reason, the only method of evaluating such algorithms will

DOI: 10.18609/cgti.2019.058

be through empirical head-to-head comparison. These topics are the subject of much active research [25].

BETTER SELECTIONS

The first law of directed evolution is that "you get what you screen for", though perhaps that is best interpreted in tandem with "be careful what you wish for" [26]. Progress in the AAV engineering field has been characterized by regular improvements in the quality of screens and selections, so as to ensure better translation into human therapeutics. Early studies selected for capsid DNA genomes in tissues of interest, but these did not always result in better vectors. That is because an increase in the abundance of a particular genome within a tissue is not necessarily an improvement, since it is not easily possible to distinguish DNA genomes that are stuck in the extracellular matrix or trapped in cell cytoplasms from those that make it all the way to the nucleus and express their payloads [27]. With this realization, recent works have focused on RNA as an indicator of full transduction [23,28]. Cre-based methods represent an alternative approach, wherein DNA genomes that become double stranded in the nucleus of target cells are re-configured within cells expressing Cre recombinase [15,29]. To engineer capsids that infect specific cell types, both RNA and Cre-based selections can be focused using promoters that restrict expression to particular subpopulations of cells. Finally, these experiments are still confronted by the fact that enhanced transduction in cell culture or in mice may unfortunately fail to translate into better use *in vivo* for humans [30,31]. The desire to develop capsids that better translate to humans is now motivating screening of capsid libraries on more closely related non-human primates, mouse-human hybrid models [32], and on human primary cells [33] or explant organs [34,35] In the near future, data from high-throughput sequencing will also make it possible to compare these approaches and identify the best experimental models for a given therapeutic application.

Machine-guided approaches will lead to better selections in two ways, first by improving protocols currently in use, and second by enabling advanced selections for the most challenging engineering goals. Improving current protocols can be as simple as comparing data from technical replicas and optimizing the methods until the results are highly correlated and reproducible. Such quality controls will guard against protocols that inadvertently bottleneck library diversity or that generate extremely noisy selections. For a given *in vivo* experiment, all the required information on delivery properties to every major organ could in principle be gathered from a single animal, or perhaps two animals to verify good correlation. Better selections thus reduce experimental times and costs, encouraging screening of libraries in the best possible model systems. A second area for improvement is the development of hybrid selections. It may be that certain delivery challenges are too difficult to succeed in a single round of library generation. Intravitreal delivery in the

eye and transduction of the deeper cellular layers is a good example of such a complex, multi-step task. While the first FDA approved ocular gene therapy is administered through sub-retinal injection [3], a limited number of patients can be treated due to the complex nature of this surgery, and there is a not insignificant risk of surgical complications. For these reasons, developing capsids that can be administered through a more routine intravitreal injection is a major goal for the field. Achieving the goal requires simultaneously solving several problems unfamiliar to a natural capsid: diffusion through the vitreous fluid, crossing the internal-limiting membrane (ILM), sliding past many cell types and finally preferentially transducing target cells. Why is it so challenging to find sequences that simultaneously overcome these multiple functional challenges? The success of directed evolution depends on being able to achieve a measurable difference in function within a given assay, in other words finding some sequences that are better than others and then moving the library in the direction of more sequences with increased fitness. But when all sequences are unviable because too many different challenges must be solved at once, there is no gradient to the fitness landscape that can be climbed. With a machine-guided approach, individual assays can be developed for each of the steps: by breaking up a difficult selection into separate tasks, the gradient of the selection for individual steps can be measured, enabling improvements to be made first on earlier steps, then over time enabling more of the library to proceed to be challenged by later steps, until ultimately resulting in sequences that are functional for the final, composite goal. Selections for each individual property are also informative even when a significant portion of the library is capable of reaching the final goal, in which case studying the correlations among individual steps can reveal inherent physical tradeoffs among these functions. Learning of this high-dimensional manifold will provide a better understanding of what it means to be truly optimal for specific delivery goals.

OPTIMIZATION ACROSS MULTIPLE PROPERTIES

As highlighted above, an ideal *in vivo* delivery technology must perform well across multiple dimensions: efficiency, specificity, low immunogenicity, and manufacturability. However, the bias in randomly selected mutants toward non-functional capsids leads to an unfortunate conclusion: a variant selected for its improved function relative to WT in one dimension will most likely be worse than WT in any other given dimension. For example, a variant selected for better transduction of a target tissue may also be more difficult to produce. This can be extremely problematic, for example in treatments using high-dose systemic administration, where the high costs and long times required to produce enough virus for a single patient become rate-limiting factors for treatment. Clearly the ideal selection method would enable improvements to be achieved across multiple distinct properties, or at a minimum not make delivery worse in any critical dimensions.

DOI: 10.18609/cgti.2019.058

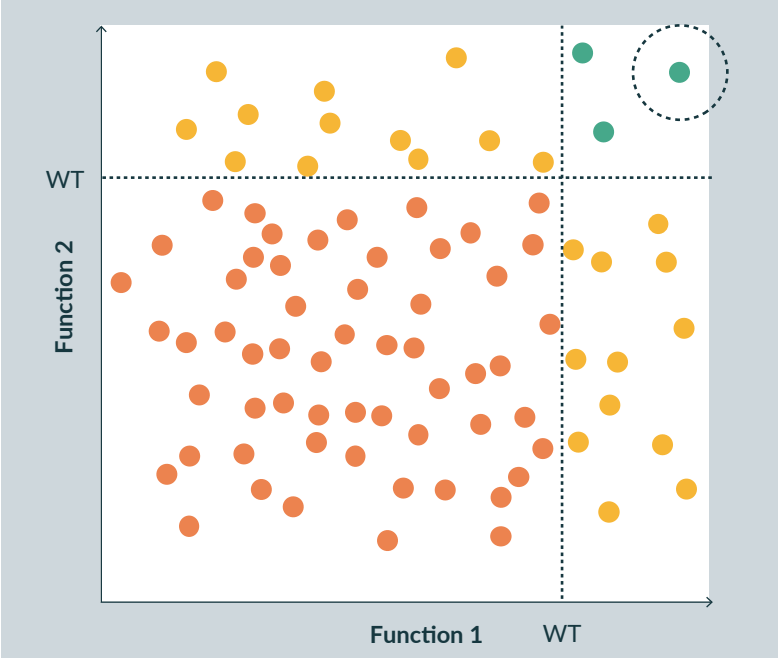
Machine-guided approaches enable AAVs to be optimized for multiple functional properties. Since it is possible to measure each of the relevant functions in high-throughput, selecting a variant that performs well across all functions is as simple as choosing an *in silico* multi-dimensional objective function and taking the best of the ranked variants (Figure 5). This process can be easily customized. First, each assay can be independently refined, and repeated, in order to improve the quality of these measurements. Second, a capsid engineer can perform selections that would be impossible to do through physical transfer of molecules: for example, while physical methods work well for enrichment strategies, it is difficult to use such assays for negative selection. To illustrate, consider the goal of generating an anti-immunogenic AAV. It would be possible to isolate AAVs that are enriched for entering primary immune cells, then from this data design a library in which these variants are removed. Such a negative selection would be difficult to perform otherwise. More complex compound selections are also easy to program: for example, an objective function might prefer a CNS-enhanced, liver-detargeted capsid that can be manufactured with same or better efficiency as the WT serotype. The relative preference for each property can now be set based upon their relevance for the therapeutic application, rather than from experimental limitations. Once these methods are fully implemented, researchers working at the cutting edge of gene delivery will never have to settle for AAV capsids with sub-par functional characteristics.

UPCOMING CHALLENGES AND CONCLUSIONS

While machine-guided engineering approaches have great promise, many challenges remain. The high-throughput technologies enabling this workflow have only recently become available, and future advances in the quality and quantity of DNA synthesis and DNA sequencing will have a major impact on what can be accomplished. Of note, it would be particularly helpful to increase the length of error-free DNA oligo synthesis; similar benefits would arise from extending deep sequencing read lengths. On the synthesis side, at present it is not cost-effective to generate large quantities of oligos (>1,000,000) at lengths

FIGURE 5

Optimization across multiple functional properties.



Since most random mutations are less fit than WT for a given property, selection for one function will most likely result in variants that are worse in an independent function (yellow). Measuring each function in high-throughput via DNA sequencing enables identification of the rare variants that are improved across multiple functions (green) and aids in selecting the best performing variant for validation (dotted circle). By focusing libraries on viable and diverse variants, machine-guided design increases the chance that the rare variants improved across the many important functional properties of AAV will be tested.

much greater than 300nt. On the sequencing side, researchers must choose between a large quantity of short-read sequencing or a lower quantity of longer reads. For now, these limitations can be overcome by clever assembly strategies and DNA barcoding. Advances in DNA synthesis and sequencing will enable a more fluid connection from experimental measurement to design and synthesis. For data-analysis, challenges include knowing how to split data into training and test sets to ensure models are not overfit on experimental noise. It will also be important to learn what sources of input data can lead to generalizable predictions of new sequences that far surpass the capabilities of sequences used for model training, in other words, models that move beyond mere interpolation to extrapolation. Advances in data analysis will come from multiple areas:

1. The greater availability and lower cost of computational cloud platforms for building data-driven models
2. Advances in applying high-capacity neural net architectures to the problem of predicting protein function
3. Availability of more data for training these models, including general lessons learned from the machine-guided engineering of proteins for diverse functions beyond AAV

With such great promise, still the use of any new AAV capsid in humans comes with a risk of failure. However, with patients in urgent need of new treatments, and with delivery being so critical to therapeutic success, the challenge of improving AAV capsids is impossible to ignore. In our view, a machine-guided approach is now best suited to solving these problems. Today's technologies enable the efficient search of sequence space, the refinement of experimental selections used to identify improved candidates, and optimization of AAVs across all therapeutically relevant delivery properties. When used appropriately, these tools can reduce the risk of complications associated with the use of natural capsids and better ensure translation to humans. Fully leveraging these advanced technologies will enable our field to engineer synthetic capsids that maximize the chance of therapeutic success.

ACKNOWLEDGEMENT

The authors would like to thank Sam Sinai, Shimyn Slomovic, Adrian Veres, Milo Lin, Ben Deverman and Tomas Bjorklund for helpful feedback on early versions of this manuscript, and additionally Pierce Ogden, Nina Jain, Noah Davidsohn, Gleb Kuznetsov and Surojit Biswas plus members of the Church lab for inspirational discussions.

FINANCIAL & COMPETING INTERESTS DISCLOSURE

Eric D Kelsic receives salary and equity from Dyno Therapeutics. GMC receives equity and consulting fees from Dyno Therapeutics. EDK and GMC are inventors on pending patents filed at Harvard University on improved AAV capsid sequences. No writing assistance

DOI: 10.18609/cgti.2019.058

was utilized in the production of this manuscript.

REFERENCES

1. Colella P, Ronzitti G, Mingozzi F. Emerging Issues in AAV-Mediated *In vivo* Gene Therapy. *Mol. Ther.: Methods Clin. Develop*. (2018); 8.
2. Dunbar CE, High KA, Joung JK, Kohn DB, Ozawa K, Sadelain M. Gene therapy comes of age. *Science* (New York NY.) (2018); 359(6372).
3. Russell S, Bennett J, Wellman JA *et al.* Efficacy and safety of voretigene neparvovec (AAV2-hRPE65v2) in patients with RPE65-mediated inherited retinal dystrophy: a randomised, controlled, open-label, phase 3 trial. *Lancet* (London, England) (2017); 390(10097), 849–860.
4. Mendell JR, Al-Zaidy S, Shell R *et al.* Single-Dose Gene-Replacement Therapy for Spinal Muscular Atrophy. *N. Engl. J. Med.* (2017); 377(18), 1713–1722.
5. Eyre-Walker A, Keightley P D. The distribution of fitness effects of new mutations. *Nat. Rev. Genetics* (2007); 8(8), 610–618.
6. Fowler DM, Fields S. Deep mutational scanning: a new style of protein science. *Nat. Methods* (2014); 11(8), 801–807.
7. Adachi K, Enoki T, Kawano Y, Veraz M, Nakai H. Drawing a high-resolution functional map of adeno-associated virus capsid by massively parallel sequencing. *Nat. Comm.* (2014); 5.
8. Kelsic E, Ogden P, Church G. Systematic Functional Characterization of the AAV Capsid Fitness Landscape. *Mol. Ther.* (2018); 26(5S1), 29.
9. Grimm D, Zolotukhin S. E Pluribus Unum: 50 Years of Research, Millions of Viruses, and One Goal—Tailored Acceleration of AAV Evolution. *Mol. Ther.* (2015); 23(12), 1819–1831.
10. Wang D, Tai PWL, Gao G. Adeno-associated virus vector as a platform for gene therapy delivery. *Nat. Rev. Drug Discovery* (2019).
11. Maheshri N, Koerber JT, Kaspar BK, Schaffer D V. Directed evolution of adeno-associated virus yields enhanced gene delivery vectors. *Nat. Biotechnol.* (2006); 24(2), 198–204.
12. Müller OJ, Kaul F, Weitzman MD *et al.* Random peptide libraries displayed on adeno-associated virus to select for targeted gene therapy vectors. *Nat. Biotechnol.*(2003); 21(9), 1040–1046.
13. Perabo L, Büning H, Kofler DM *et al. In Vitro* Selection of Viral Vectors with Modified Tropism: The Adeno-associated Virus Display. *Mol. Ther.* (2003); 8(1).
14. Grimm D, Lee JS, Wang L *et al. In Vitro* and *In vivo* Gene Therapy Vector Evolution via Multispecies Interbreeding and Retargeting of Adeno-Associated Viruses. *J. Virol.* (2008); 82(12), 5887–5911.
15. Ojala DS, Sun S, Santiago-Ortiz JL, Shapiro MG, Romero PA, Schaffer DV. *In vivo* Selection of a Computationally Designed SCHEMA AAV Library Yields a Novel Variant for Infection of Adult Neural Stem Cells in the SVZ. *Mol. Ther.* (2018); 26(1).
16. Marsic D, Govindasamy L, Currlin S *et al.* Vector design Tour de Force: integrating combinatorial and rational approaches to derive novel adeno-associated virus variants. *Mol. Ther.: J. Am. Soc. Gene Ther.* (2014); 22(11), 1900–1909.
17. Zinn E, Pacouret S, Khaychuk V et al. *In Silico* Reconstruction of the Viral Evolutionary Lineage Yields a Potent Gene Therapy Vector. *Cell Reports* (2015); 12.
18. Shen S, Horowitz ED, Troupes AN *et al.* Engraftment of a galactose receptor footprint onto adeno-associated viral capsids improves transduction efficiency. *J. Bio. Chem.* (2013); 288(40), 28814–28823.
19. Tse LV, Klinc KA, Madigan VJ *et al.* Structure-guided evolution of antigenically distinct adeno-associated virus variants for immune evasion. *Proceedings of the National Academy of Sciences* (2017); 201704766.
20. Kosuri S, Eroshenko N, LeProust EM *et al.* Scalable gene synthesis by selective amplification of DNA pools from high-fidelity microchips. *Nat. Biotechnol.*(2010); 28(12), 1295–1299.
21. Raveh-Sadka T, Levo M, Shabi U *et al.* Manipulating nucleosome disfavoring sequences allows fine-tune regulation of gene expression in yeast. *Nature Genetics* (2012); 44(7), 743–750.
22. Sharon E, Kalma Y, Sharp A *et al.* Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed

promoters. *Nat. Biotechnol.*(2012); 30(6), 521–530.

23. Davidsson M, Wang G, Aldrin-Kirk P *et al.* Barcoded Rational AAV Vector Evolution enables systematic in vivo mapping of peptide binding motifs. *BioRxiv* (2018); 335372.

24. Kelsic E, Sinai S, Ogden P, Nowak M, Church G. Deep Search: Next-Gen Strategies for Accelerating AAV Capsid Engineering. *Mol. Ther.* (2018); 26(5S1), 168.

25. Yang KK, Wu Z, Arnold FH. Machine learning-guided directed evolution for protein engineering. https://arxiv.org/abs/1811.10775

26. You L, Arnold FH, Directed evolution of subtilisin E in Bacillus subtilis to enhance total activity in aqueous dimethylformamide, *Protein Engineering, Design and Selection*, Volume 9, Issue 1, January 1996, Pages 77-83.

27. Zincarelli C, Soltys S, Rengo G, Rabinowitz J E. Analysis of AAV Serotypes 1–9 Mediated Gene Expression and Tropism in Mice After Systemic Injection. *Mol. Ther.* (2008); 16(6), 1073–1080.

28. Weinmann J, Weis S, Sippel J, Lenter M, Lamla T, Grimm D. Massively Parallel *In vivo* Characterization of >150 Adeno-Associated Viral (AAV) Capsids Using DNA/RNA Barcoding and Next-Generation Sequencing. *Mol. Ther.* (2018); 26(5S1), 319–320.

29. Deverman BE, Pravdo PL, Simpson BP *et al.* Cre-dependent selection yields AAV variants for widespread gene transfer to the adult brain. *Nat. Biotechnol.*(2016); 34(2), 204–209.

30. Hordeaux J, Wang Q, Katz N, Buza EL, Bell P, Wilson JM. The Neurotropic Properties of AAV-PHP.B Are Limited to C57BL/6J Mice. *Mol. Ther.: J. Am. Soc. Gene Ther.* (2018); 26(3), 664–668.

31. Huang Q, Chan KY, Tobey IG *et al.* Delivering genes across the blood-brain barrier: LY6A, a novel cellular receptor for AAV-PHP.B capsids. *BioRxiv* (2019); 538421.

32. Lisowski L, Dane AP, Chu K *et al.* Selection and evaluation of clinically relevant AAV variants in a xenograft liver model. *Nature* (2014); 506(7488), 382–386.

33. Paulk NK, Pekrun K, Charville GW *et al.* Bioengineered Viral Platform for Intramuscular Passive Vaccine Delivery to Human Skeletal Muscle. *Mol. Ther.: Methods Clin. Develo*p. (2018); 10, 144–155.

34. Buck TM, Pellissier LP, Vos RM, van Dijk EHC, Boon CJF, Wijnholds J. AAV Serotype Testing on Cultured Human Donor Retinal Explants. In *Methods in molecular biology* (Clifton NJ.) (2018); (Vol. 1715, pp. 275–288).

35. Orlans HO, Edwards TL, De Silva SR, Patrício MI, MacLaren R E. Human Retinal Explant Culture for Ex Vivo Validation of AAV Gene Therapy. In: *Methods in molecular biology* (Clifton NJ.) (2018); 1715, pp. 289–303.

AFFILIATIONS

Eric D Kelsic

Dyno Therapeutics, Cambridge, MA, USA

Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA, USA

George M Church

Department of Genetics, Harvard Medical School, Boston, MA, USA

Wyss Institute for Biologically Inspired Engineering, Harvard University, Cambridge, MA, USA

DOI: 10.18609/cgti.2019.058