



Overcoming Immunological Challenges Limiting Capsid-Mediated Gene Therapy With Machine Learning

Anna Z. Wec¹, Kathy S. Lin², Jamie C. Kwasnieski¹, Sam Sinai², Jeff Gerold² and Eric D. Kelsic^{1,2*}

¹ Applied Biology, Dyno Therapeutics Inc, Cambridge, MA, United States, ² Data Science, Dyno Therapeutics Inc, Cambridge, MA, United States

OPEN ACCESS

Edited by:

Guangping Gao,
University of Massachusetts Medical
School, United States

Reviewed by:

Phillip Tai,
University of Massachusetts Medical
School, United States

Sergei Zolotukhin,

University of Florida, United States

Thomas Weber,

Icahn School of Medicine at Mount

Sinai, United States

Chengwen Li,

University of North Carolina at

Chapel Hill, United States

*Correspondence:

Eric D. Kelsic
eric.kelsic@dynotx.com

Specialty section:

This article was submitted to
Vaccines and Molecular Therapeutics,
a section of the journal
Frontiers in Immunology

Received: 28 February 2021

Accepted: 09 April 2021

Published: 27 April 2021

Citation:

Wec AZ, Lin KS, Kwasnieski JC,
Sinai S, Gerold J and Kelsic ED (2021)
Overcoming Immunological
Challenges Limiting Capsid-Mediated
Gene Therapy With Machine Learning.
Front. Immunol. 12:674021.
doi: 10.3389/fimmu.2021.674021

A key hurdle to making adeno-associated virus (AAV) capsid mediated gene therapy broadly beneficial to all patients is overcoming pre-existing and therapy-induced immune responses to these vectors. Recent advances in high-throughput DNA synthesis, multiplexing and sequencing technologies have accelerated engineering of improved capsid properties such as production yield, packaging efficiency, biodistribution and transduction efficiency. Here we outline how machine learning, advances in viral immunology, and high-throughput measurements can enable engineering of a new generation of de-immunized capsids beyond the antigenic landscape of natural AAVs, towards expanding the therapeutic reach of gene therapy.

Keywords: gene therapy, protein engineering, immune evasion, machine learning, AAV capsid design

INTRODUCTION

Recently approved AAV-based therapeutics and numerous therapeutic candidates in advanced clinical development (1) have demonstrated the transformative and life-saving potential of viral capsids as vectors for gene therapy (GT). The demands on viral capsids to deliver gene replacement and gene editing tools will continue to increase as our understanding of genetic diseases reveals new therapeutic opportunities. Development of next generation capsids that enable more precise, efficient, and durable gene delivery will be key to improving the effectiveness and safety of such therapies. In this perspective, we explore how high throughput (HT) measurement and characterization methods can be combined with machine learning (ML) approaches to identify such capsids by efficiently optimizing capsid sequences for both improved transduction and reduced immunogenicity. Combining these technologies will generate capsid-mediated gene therapies with broader therapeutic uses that are accessible to all individuals in need.

THE NEED TO OPTIMIZE NATURAL AAV CAPSIDS FOR THERAPEUTIC DELIVERY

Most recombinant AAV capsids used clinically today are closely related, or even identical, to naturally occurring AAVs in their amino acid sequences and biological properties. As natural selection did not optimize such capsids for therapeutic use, they display limited specificity of cell

targeting and low overall *in vivo* transduction efficiency in many target tissues, particularly following intravenous administration. Improving *in vivo* transduction of target cells and organs would enable gene therapies to more effectively treat diseases, to perdure, and to address new therapeutic applications. Importantly, pre-existing humoral and cellular immunity against natural AAV capsids limits patient eligibility for therapies as well as their therapeutic efficacy (2). Furthermore, capsids possess inherent immunogenicity — the propensity to activate immune responses — which can impact safety and efficacy, as well as the potential for redose. The challenges of evading both pre-existing immunity and *de novo* adaptive immune responses against AAV vectors are made especially difficult by the heterogeneous nature of patient immune responses and immune histories. Thus, discovering capsids that circumvent the immune system is a significant hurdle facing developers of next generation GT vectors (2).

Established approaches for obtaining novel capsids include mining the naturally-occurring sequence diversity of capsids, rational design and directed evolution (3–5). Each methodology

has contributed valuable capsids to the available catalog of GT vectors, but limitations related to speed and throughput of discovery persist because the total number of possible capsids far exceeds the capacity of current screening approaches. Directed evolution methods often take advantage of ultra-high diversity generated by random mutagenesis in an attempt to overcome the barrier of low discovery yield (i.e. success per individual design). In contrast, rational design approaches rely on expert knowledge and focus on a higher likelihood of success per design, but are relatively low throughput (and overall low yield) as a result. ML approaches offer a promising new option that may mitigate the trade-off between yield and throughput (Figure 1A). ML can be used in combination with these established approaches, or as a stand-alone technique to open new avenues of discovery through high-throughput direct synthesis (6).

The set of desired properties that a capsid should possess in order to be therapeutically transformative can collectively be termed a *capsid profile*, in other words the target of optimization efforts. Capsids that embody every therapeutically desirable property

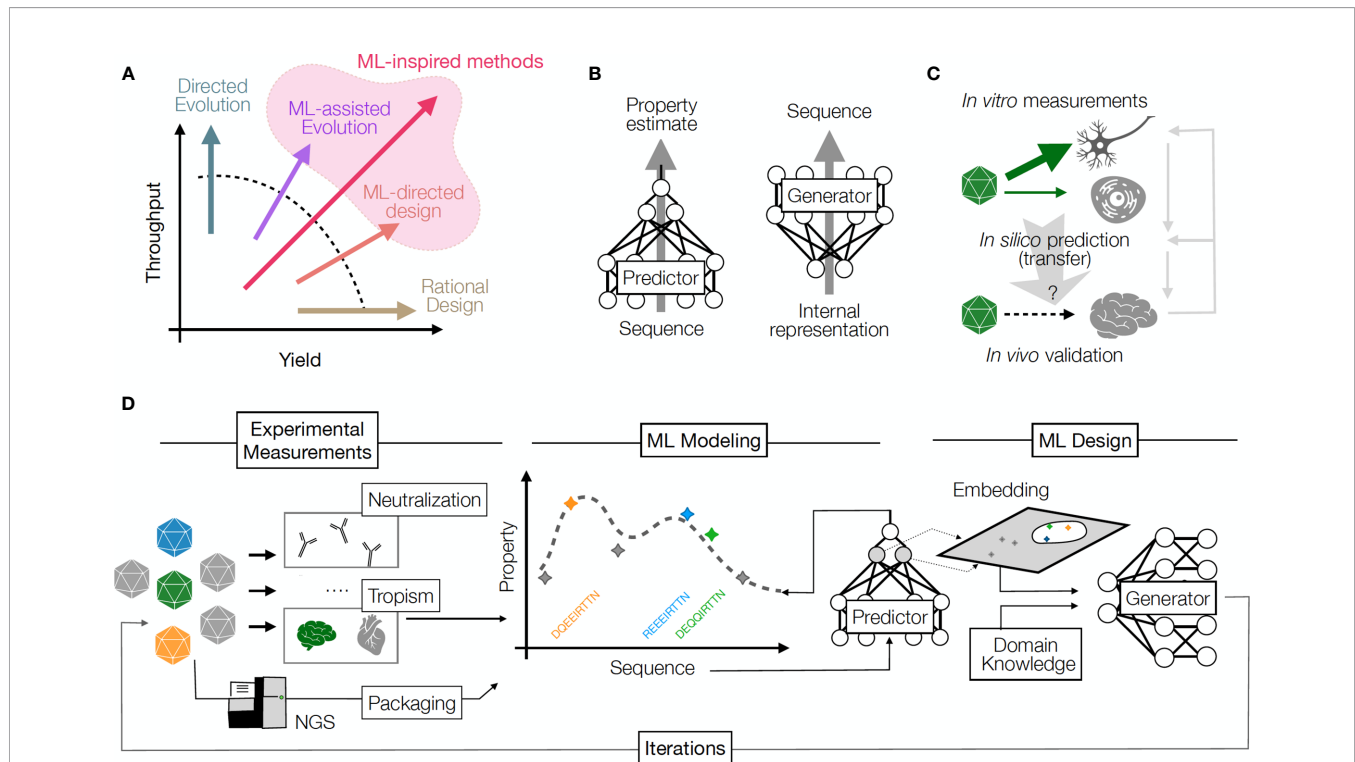


FIGURE 1 | (A) A comparison of throughput (number of samples) and yield (fraction of successful samples generated per attempt) for multiple protein design approaches. Rational design increases yield, directed evolution leverages throughput, and ML methods increase the likelihood of success by balancing yield and throughput. **(B)** Predictive ML models map sequences to their functional properties, while Generative methods can turn an internal data representation back into sequences, producing desirable samples. **(C)** An example of transfer learning whereby a model *transfers* information across cell types and experimental contexts: a model learns based on *in vitro* capsid performance in diverse cell transduction experiments (including neurons), then is applied to predict the result of *in vivo* transduction in the brain neurons, when such experimental data is sparse or missing. Information from *in vivo* validation of the predicted capsid performance is used to refine model performance and understand the relationship between *in vivo* and *in vitro* assays. Right grey arrows illustrate the iterative power of this approach, which refines predictive and generative models over time. **(D)** The design cycle starts with HT screening and measurements of several AAV capsid variant properties. These properties are then used to train predictive models that can impute the property for unseen sequences (predictor model) and can be used to build helpful representations (embeddings), which can then be integrated with auxiliary input (e.g., domain knowledge) to propose a batch of new sequences (generator model). The design process can be repeated in multiple iterations until desired capsids are discovered.

outlined above have eluded discovery despite years of effort. Despite the vast number of possible capsid sequences, it is reasonable to assume capsids which achieve these desired profiles, if they exist, are extremely rare in sequence space (7, 8). Reducing the number of required properties in the context of a particular therapeutic application may increase the chance of finding a candidate capsid, but this may come at the cost of failure in later stages of clinical development. The therapeutic usefulness of a given capsid and our ability to find it are therefore fundamentally in tension. In this perspective, we share how new approaches to immunological data gathering, combined with analysis and design approaches powered by ML, are overcoming this tension towards discovery of capsids that are more therapeutically useful.

KEY CONCEPTS FOR APPLYING MACHINE LEARNING TO ENGINEER NOVEL CAPSIDS

Recent advances in ML enable new solutions to problems inherent to designing immune-evasive capsids. ML is a collection of algorithmic approaches that allow for automatic learning. These approaches are capable of learning rules for predicting the outcome of complex processes directly from input data. Larger and richer datasets pose a challenge for traditional methods of rational design but are the environment in which ML methods thrive (9). ML models can be considered mathematical approximators of physical processes we have measured, and oftentimes have yet to understand mechanistically (10–12). In the context of biological design, ML models can replace labor- or resource-intensive experiments with *in silico* screening. With increasing amounts of data, these approximations can become very accurate, and their rapid and cost-effective application enables the identification of biological designs which would not be accessible by experimentation alone. Importantly, mechanistic knowledge need not be wasted in this approach — biological insights can be incorporated into ML architectures in a way that bolsters model robustness, allowing for more accurate models trained by less data. Additionally, ML can simplify how we represent and understand high-dimensional and high-throughput data, allowing us to substantially improve the experiments themselves. Finally, while many mechanistic details of AAV gene therapy remain poorly understood, ML models trained on empirical data that can predict capsid functions are sufficiently useful for engineering better capsids despite the models being agnostic to mechanism, and in some cases querying such models can guide or improve our mechanistic understanding.

Key ML concepts illustrate the potential for this approach to transform capsid engineering. First, ML algorithms can learn arbitrary *sequence-to-function* relationships. These relationships can be learned automatically from large datasets of capsid sequences and their measured properties. A model can predict one or multiple properties at once. For instance, models can be trained to learn the relationship between the capsid sequence and its ability to produce a viable capsid (6) or its tropism to the

liver (13). These training schemes, termed *supervised*, require collecting data labels (measurements) of the kind we are intending to predict. However, it is also possible to train models solely based on a set of good examples without additional measurements. For instance, training models on the rapidly growing set of publicly available protein sequences to learn relationships among them has shown promise in protein structure and function prediction (12, 14–17). This type of training is known as *unsupervised*. Both supervised and unsupervised training schemes can yield *predictive models* that output property values given an input sequence, or alternatively *generative models* that produce novel sequences given desirable property values as inputs (**Figure 1B**). It is noteworthy that building models with good generalization ability, i.e. ability to predict accurately on samples far from those in the training data, requires care in experimental design and training schemes. Otherwise, models may *overfit* to the training data available, where they perform well on samples similar to their training data, but unexpectedly poorly in novel settings.

Second, effective machine learning methods often make use of internal *latent representations*, also known as *embeddings*, which attempt to represent the information contained in raw inputs in a way that is more amenable to human understanding. One such simple and widely applied method is principal component analysis (PCA), in which a linear transformation of input data allows for the identification of data elements that contribute most to the variance in the data set. PCA and other more complex non-linear dimensionality reduction methods transform high-dimensional raw input data to a lower-dimensional representation (a latent space) that is easier to interpret, visualize, and optimize (14, 18–21). If these and other methods can be applied to the problem of AAV capsid engineering, AAV variant sequences with similar properties to each other would be close together in latent space after being transformed into their latent representations, even if they are far apart in sequence space. A similar strategy was recently used to predict the emergence of escape mutations in multiple viruses (22).

Finally, modern ML can utilize auxiliary data to make inference about domains where information is sparse, a process known as *transfer learning* (**Figure 1C**) (23, 24). An illustrative conceptual example for this technique in machine vision involves “style-transfer” where particular painting styles are learned from an artist’s work, and can then be applied to any new image, converting the style to that of the original artist (25). This type of learning can be used in many contexts in biology (23, 26). For instance, predictive models around AAV serotypes for which little data is available could be improved by training them on data available from other related serotypes or even a larger set of related proteins. Similarly, population level data for immunity profiles of specific patient groups could be used to reduce the amount of data required to make inferences for individual patients. Along with the ability to integrate information from multiple modalities, transfer learning can rapidly accelerate the application of ML models in areas where data is limited, and open new domains for prediction and design.

An example of a ML-driven design pipeline is illustrated in **Figure 1D**. These concepts will be useful for designing immune-evasive capsids, as we explain below.

SAFE AND EFFECTIVE TREATMENT AT LOWER DOSES

Among all capsid properties that could be improved, increased tissue-specific transduction is key to enabling safe and effective gene therapies. Improving this attribute would allow for a higher proportion of injected capsids to deliver their payloads to the intended cells, reducing the dose needed for effective treatment. This in turn would make treatment safer by reducing activation of the innate immune responses and of B and T cell responses, which increase in magnitude relative to the amount of antigenic stimulus (vector dose) delivered (27).

Making viral vectors safer and more effective will require optimization towards multi-property capsid profiles. However, many capsid properties are intrinsically coupled to one another and efforts to optimize or re-direct any single attribute often result in capsids that fail basic tests of functionality, such as capsid assembly and genome packaging. ML models can greatly reduce the burden of multi-property optimization through *in silico* screening of variants (28), ensuring that optimization toward one property does not break other desired functions (29, 30), shifting the engineering burden away from experimental approaches (28). For instance, four supervised models can be trained to learn sequence-to-function maps between capsid sequences and their ability to (i) transduce the liver, (ii) bypass off-target organs, (iii) evade neutralization, and (iv) produce at high yield. The first model can be used in an *in silico* search for variants with better transduction, and the other models can be used to eliminate sequences proposed by the first model that do not meet the specificity, immune evasion and capsid production requirements. A significant body of work in the interface of ML and biology is focused on algorithms that use such supervised models to optimally design protein sequences (31). Notably, while non-human primates are at present the industry-preferred model for measuring transduction, the ability for ML to integrate diverse sources of information may increase the utility of data from other animal models (including transgenic animals with humanized immune systems), as well as human cell culture models, for predicting transduction patterns in human patients and lead to better rates of clinical translation. Capsids optimized towards a profile of improved and specific transduction, reduced immunogenicity, and production efficiencies equivalent to natural AAV capsids would already be transformative relative to currently available vectors.

PERDURING GENE THERAPY

In an ideal therapeutic scenario, a single dose of GT would provide a durable, curative effect throughout a recipient's lifetime. In practice, this goal has been difficult to realize as

therapeutic transgene expression from current vectors decays over time (32). Waning transgene expression can result from silencing of the viral genome through epigenetic mechanisms, from cell division, or from transduced cell death, among other factors. One mechanism underlying the loss of transduced cells observed in a number of clinical studies (33–35) was the induction of cytotoxic CD8⁺ T lymphocyte (CTL) responses against cells presenting capsid antigens, for which immunosuppression is the primary clinically viable remedy.

Engineering capsids that reduce or even eliminate CTL responses will facilitate perduring therapeutic gene expression. Transduced cells process viral capsids through the intracellular proteolytic machinery and present capsid-derived peptides on their surface through the major histocompatibility (MHC) class I molecules (33, 34). CD8⁺ T cells recognize presented peptides *via* their highly specific T cell receptors, which in turn determines cell stimulation, proliferation and cytotoxic activity. CTL activation results in killing of transduced cells as well as generation of immunologic memory that poses a barrier for vector redosing. Unlike B cells, which interact with surface exposed capsid epitopes, T cells can in theory sample the full peptidome of an AAV capsid, including buried capsid sequences that drive assembly or disassembly, and which may be more difficult to alter by conventional engineering approaches. Extensive mapping of CD8⁺ T cell epitopes within AAV capsid proteins and evaluation of their propensity to activate T cell responses would identify the key sequences which must be modified to de-immunize AAV capsids. The large diversity of HLA alleles among people and distinct patterns of peptide presentation and recognition determined by them makes this challenging. While it is currently not possible to exhaustively assess peptide presentation by all variants of MHC class I found in humans, emerging ML methods in peptide presentation and immunogenicity prediction (36, 37) will increase the accuracy of these predictions compared to tools available today. Recently developed strategies of experimental immunopeptidome characterization using mass spectrometry (38, 39) will provide a rich source of data for training such models.

Understanding the determinants of capsid antigen presentation (40) and their effect on CTL activation will provide the foundations for ML models to engineer capsids that evade them. The rules of peptide presentation are shared across the entire proteome based upon an individual patient's HLA alleles (41). This means that ML models can benefit from all existing datasets that catalog CD8⁺ T cell epitopes and learn general properties that influence which peptides tend to be presented in particular genetic backgrounds (17). Through transfer learning, such general models could be tuned toward more accurate models that predict CD8⁺ T cell epitopes for AAV capsid variants specifically. This would require relatively small amounts of additional data that is specific to AAV capsids and would enable engineering of capsids depleted of T cell-activating peptides. While predictions of MHC class I presentation have advanced significantly, meaningful annotation of peptide immunogenicity that enables more accurate models for immunogenicity prediction will require development of HT

functional assays and remains an open challenge for the field of T cell biology.

GENE THERAPY FOR ALL: OVERCOMING PRE-EXISTING ANTI-CAPSID ANTIBODIES

A majority of prospective GT recipients have pre-existing antibodies against one or more natural AAV serotypes, often excluding them from treatment (42–44). Pre-existing antibodies accelerate vector clearance, redirect vector biodistribution, and can directly inhibit capsid-mediated cell entry (33). To overcome these activities of antibodies, it is critical to identify capsids that cannot be efficiently bound and neutralized by them – in other words, capsids with surface-exposed sequence and structural features not previously encountered by the adaptive immune response. Altering antibody recognition of capsids in a therapeutically meaningful way is challenging because serum antibody responses are highly diverse and can target the entire capsid surface (45, 46). Antibodies bind both linear and discontinuous epitopes on the capsid exterior surface, sometimes spanning across neighboring capsid subunits, making rational approaches to altering these sites challenging. Moreover, neutralizing antibodies often target capsid regions involved in critical functions such as cell receptor recognition, meaning that mutations which prevent antibody binding can also adversely affect vector transduction (47).

Much remains to be learned about how human antibodies bind to and neutralize capsids, however several technologies now enable high-throughput mapping of antibody responses at the monoclonal level. The study of both serum antibodies and antibodies encoded by memory B cells in donors with recent AAV exposures can reveal key characteristics of human anti-capsid antibody responses and provide a more complete picture of anti-capsid antibody immunity. While serum antibodies are maintained at steady state by long lived plasma cells, the memory B cell repertoire approximates the antibody repertoire that will be mobilized on AAV re-encounter and their characterization is methodologically useful as a means of identifying anti-capsid antibody sequences for in depth functional studies. For example, efforts in the infectious diseases therapeutic space have yielded multiple approaches to fine mapping of *de novo* and memory B cell responses, where hundreds or even thousands of virus-specific antibodies encoded by B cells can now be routinely sequenced, cloned and produced (48). Epitopes of such antibodies can be characterized using HT competition assays (49, 50) and correlations can be derived between binding site location and neutralization activity. Recently developed approaches utilizing cryo-electron microscopy (51, 52) and high resolution, quantitative, proteomics-based approaches (53–55) enable serum antibody specificities to be characterized in unprecedented detail, to inform their identities and their binding sites. These and other studies revealed for a number of pathogens that just one class of antibodies can contribute the majority of neutralizing activity in the serum despite the overall high diversity of antibody responses (56–58). Identifying any

dominant human neutralizing antibody types against AAVs would inform the sites where capsid engineering can be most effectively applied.

Data with resolution at the individual antibody level would enable ML models to learn how antibody responses target a particular capsid and how to predict their effect on other (designed) capsids. Models can serve as *in silico* evaluators of capsids before they are administered to patients with pre-existing antibodies based on characterization using the methods described above. Through sequencing of capsid-specific B cells and characterization of serum antibodies, a personal ‘immunological fingerprint’ can be created with the aid of ML models, which could also be used to find general patterns in human anti-capsid antibody responses (59). For instance, unsupervised models can directly learn from genetic data to predict immune profile responses. Supervised models could use patient serum data together with other measurements [e.g. sequencing of immune repertoires (59) or genome scanning antibody profiling (60)] to predict likelihood of therapeutic success, or to help select vector administration options. With such models in hand, panels of antibody-evading AAV capsids could be recommended based on a patients’ pre-existing antibody repertoire to maximize the chance of effective antibody evasion.

Many gaps remain in our understanding of how anti-capsid antibodies can be evaded. Serology studies with naturally occurring AAVs have been useful in defining population-level prevalence of anti-AAV immunity but such bulk-level measurements have had limited value for engineering antibody-evading capsids. Some monoclonal antibodies isolated from mice have been characterized in detail (46, 61) providing important insights about the antigenic sites on AAV capsids targeted by neutralizing antibodies. However, it remains a challenge to generalize these results to human antibody responses, which are encoded by distinct germline genes, are more diverse (62), and are shaped in response to a distinct set of natural AAVs endemic in humans. An in-depth large-scale characterization of human antibodies targeting capsids would facilitate our ability to engineer capsids with maximal therapeutic impact.

One such promising approach would be to measure the activity of serum antibodies against highly diverse libraries of capsid variants using immune human serum samples. Such data would enable ML models to learn the quantitative relationship between AAV capsid sequences and their abilities to evade pre-existing antibodies, and to learn commonalities in anti-capsid antibody responses among people. Similarly, intravenous immunoglobulin (IVIg) preparations containing antibodies from thousands of donors may be useful in such screens for identifying the predominant patterns in human antibody responses. Recent work characterizing B cell and antibody responses to a number of important human pathogens (56, 63–65) reveal common features of antibody responses elicited by a given pathogen across donors. If similar shared antibody types arise against AAV capsids, resurfacing the epitopes they target would allow engineering of capsids that more broadly

evade antibody activity, towards the goal of creating *universal capsids* capable of treating all patients.

FUTURE DIRECTIONS

ML-powered capsid design and engineering will transform the landscape of GT delivery modalities, however non-capsid improvements are also relevant from an immunological perspective and can also increase therapeutic effectiveness. Reducing the activation of innate immunity by engineering the vector genome (66, 67), co-administration with targeted immune-modulators to induce tolerance toward the vector (68) or depletion of pre-existing anti-capsid antibodies (69) should work in synergy with engineered capsids to pave a path for repeat vector administration, while further increasing the safety and tolerability of next generation GTs.

As we have outlined, ML approaches to engineer improved AAV capsids have multiple applications: enabling gene therapies that are effective in a lower dose regimen, removing capsid peptides which elicit cytotoxic T cell responses thereby leading to longer lasting gene expression, and resurfacing capsid exteriors allowing potentially universal treatment of all patients. While these goals are ambitious and each individually worthy of study, combining all such properties in a single capsid would be transformative for the field. ML approaches will facilitate this goal by incorporating information from diverse experimental systems and improving the efficiency of multi-trait capsid optimization. We are optimistic that safe, efficient, target-

specific, non-immunogenic and universal capsids will one day enable gene therapy to reach its full potential by delivering therapeutic DNA to cure, treat and prevent disease and even to improve overall health for all patients. Interdisciplinary collaborations focused on combining HT measurements with ML-powered sequence design algorithms will dramatically accelerate progress towards achieving these goals.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

AW, KL, JK, SS, JG and EK conceptualized, wrote and edited the manuscript. AW and SS prepared figures. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

We thank George Church, Jakub Otwinowski, Sam Wolock, Alexander Brown, Sylvain Lapan, Adrian Veres and Tomas Björklund for their helpful discussions and comments on the manuscript.

REFERENCES

- Wang D, Tai PWL, Gao G. Adeno-Associated Virus Vector as a Platform for Gene Therapy Delivery. *Nat Rev Drug Discovery* (2019) 18:358–78. doi: 10.1038/s41573-019-0012-9
- Verdera HC, Kuranda K, Mingozi F. Aav Vector Immunogenicity in Humans: A Long Journey to Successful Gene Transfer. *Mol Ther* (2020) 28:723–46. doi: 10.1016/j.jymth.2019.12.010
- Davidsson M, Wang G, Aldrin-Kirk P, Cardoso T, Nolbrant S, Hartnor M, et al. A Systematic Capsid Evolution Approach Performed In Vivo for the Design of AAV Vectors With Tailored Properties and Tropism. *Proc Natl Acad Sci USA* (2019) 116(52):27053–62. doi: 10.1073/pnas.1910061116
- Byrne LC, Day TP, Visel M, Strazzeri JA, Fortuny C, Dalkara D, et al. In Vivo-Directed Evolution of Adeno-Associated Virus in the Primate Retina. *JCI Insight* (2020) 5(10):e135112. doi: 10.1172/jci.insight.135112
- Qian R, Xiao B, Li J, Xiao X. Directed Evolution of AAV Serotype 5 for Increased Hepatocyte Transduction and Retained Low Humoral Seroreactivity. *Mol Ther Methods Clin Dev* (2021) 20:122–32. doi: 10.1016/j.omtm.2020.10.010
- Bryant DH, Bashir A, Sinai S, Jain NK, Ogden PJ, Riley PF, et al. Deep Diversification of an AAV Capsid Protein by Machine Learning. *Nat Biotechnol* (2021). doi: 10.1038/s41587-020-00793-4
- Povolotskaya IS, Kondrashov FA. Sequence Space and the Ongoing Expansion of the Protein Universe. *Nature* (2010) 465:922–6. doi: 10.1038/nature09105
- Bartel DP, Szostak JW. Isolation of New Ribozymes From a Large Pool of Random Sequences. *Science* (1993) 261:1411–8. doi: 10.1126/science.7690155
- Webb S. Deep Learning for Biology. *Nature* (2018) 554:555–7. doi: 10.1038/d41586-018-02174-z
- Yuan B, Shen C, Luna A, Korkut A, Marks DS, Ingraham J, et al. Cellbox: Interpretable Machine Learning for Perturbation Biology With Application to the Design of Cancer Combination Therapy. *Cell Syst* (2021) 12:128–40.e4. doi: 10.1016/j.cels.2020.11.013
- Madani A, McCann B, Naik N, Keskar NS, Anand N, Eguchi RR, et al. Progen: Language Modeling for Protein Generation. *arXiv [q-bioBM]* (2020). doi: 10.1101/2020.03.07.982272
- Senior AW, Evans R, Jumper J, Kirkpatrick J, Sifre L, Green T, et al. Improved Protein Structure Prediction Using Potentials From Deep Learning. *Nature* (2020) 577:706–10. doi: 10.1038/s41586-019-1923-7
- Ogden PJ, Kelsic ED, Sinai S, Church GM. Comprehensive AAV Capsid Fitness Landscape Reveals a Viral Gene and Enables Machine-Guided Design. *Science* (2019) 366:1139–43. doi: 10.1126/science.aaw2900
- Sinai S, Kelsic E, Church GM, Nowak MA. Variational Auto-Encoding of Protein Sequences. *arXiv [q-bioQM]* (2017).
- Riesselman AJ, Ingraham JB, Marks DS. Deep Generative Models of Genetic Variation Capture the Effects of Mutations. *Nat Methods* (2018) 15:816–22. doi: 10.1038/s41592-018-0138-4
- Marks DS, Colwell LJ, Sheridan R, Hopf TA, Pagnani A, Zecchina R, et al. Protein 3D Structure Computed From Evolutionary Sequence Variation. *PLoS One* (2011) 6:e28766. doi: 10.1371/journal.pone.0028766
- Ogishi M, Yotsuyanagi H. Quantitative Prediction of the Landscape of T Cell Epitope Immunogenicity in Sequence Space. *Front Immunol* (2019) 10:827. doi: 10.3389/fimmu.2019.00827
- Becht E, McInnes L, Healy J, Dutertre C-A, Kwok IWH, Ng LG, et al. Dimensionality Reduction for Visualizing Single-Cell Data Using UMAP. *Nat Biotechnol* (2018) 37:38–44. doi: 10.1038/nbt.4314
- van der Maaten L. *Visualizing Data Using T-SNE* (2008). Available at: <http://jmlr.org/papers/v9/vandermaaten08a.html>.
- Ringnér M. What is Principal Component Analysis? *Nat Biotechnol* (2008) 26:303–4. doi: 10.1038/nbt0308-303

21. Belkin M, Niyogi P. Laplacian Eigenmaps for Dimensionality Reduction and Data Representation. *Neural Comput* (2003) 15:1373–96. doi: 10.1162/08997660321780317
22. Hie B, Zhong ED, Berger B, Bryson B. Learning the Language of Viral Evolution and Escape. *Science* (2021) 371:284–8. doi: 10.1126/science.abd7331
23. Rao R, Bhattacharya N, Thomas N, Duan Y, Chen X, Canny J, et al. Evaluating Protein Transfer Learning With TAPE. *Adv Neural Inf Process Syst* (2019) 32:9689–701. doi: 10.1101/676825
24. Tan C, Sun F, Kong T, Zhang W, Yang C, Liu C. A Survey on Deep Transfer Learning. *arXiv [csLG]* (2018). doi: 10.1007/978-3-030-01424-7_27
25. Gatys LA, Ecker AS, Bethge M. Image Style Transfer Using Convolutional Neural Networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition: Computer vision foundation* (2016). p. 2414–23.
26. Wang J, Agarwal D, Huang M, Hu G, Zhou Z, Ye C, et al. Data Denoising With Transfer Learning in Single-Cell Transcriptomics. *Nat Methods* (2019) 16:875–8. doi: 10.1038/s41592-019-0537-1
27. Vandenberghe LH, Wilson JM. AAV as an Immunogen. *Curr Gene Ther* (2007) 7:325–33. doi: 10.2174/156652307782151416
28. Marques AD, Kummer M, Kondratov O, Banerjee A, Moskalenko O, Zolotukhin S. Applying Machine Learning to Predict Viral Assembly for Adeno-Associated Virus Capsid Libraries. *Mol Ther Methods Clin Dev* (2021) 20:276–86. doi: 10.1016/j.omtm.2020.11.017
29. Biswas M, Marsic D, Li N, Zou C, Gonzalez-Aseguinolaza G, Zolotukhin I, et al. Engineering and In Vitro Selection of a Novel Aav3b Variant With High Hepatocyte Tropism and Reduced Seroreactivity. *Mol Ther Methods Clin Dev* (2020) 19:347–61. doi: 10.1016/j.omtm.2020.09.019
30. Patrick Havlik L, Simon KE, Kennon Smith J, Klinc KA, Tse LV, Oh DK, et al. Coevolution of Adeno-associated Virus Capsid Antigenicity and Tropism Through a Structure-Guided Approach. *J Virol* (2020) 94(19):e00976–20. doi: 10.1128/JVI.00976-20
31. Sinai S, Kelsic ED. A Primer on Model-Guided Exploration of Fitness Landscapes for Biological Sequence Design. *arXiv [q-bioQM]* (2020).
32. Colella P, Ronzitti G, Mingozzi F. Emerging Issues in AAV-Mediated in Vivo Gene Therapy. *Mol Ther Methods Clin Dev* (2018) 8:87–104. doi: 10.1016/j.omtm.2017.11.007
33. Vandamme C, Adjali O, Mingozzi F. Unraveling the Complex Story of Immune Responses to AAV Vectors Trial After Trial. *Hum Gene Ther* (2017) 28:1061–74. doi: 10.1089/hum.2017.150
34. Mingozzi F, Maus MV, Hui DJ, Sabatino DE, Murphy SL, Rasko JEJ, et al. Cd8 (+) T-cell Responses to Adeno-Associated Virus Capsid in Humans. *Nat Med* (2007) 13:419–22. doi: 10.1038/nm1549
35. Manno CS, Pierce GF, Arruda VR, Glader B, Ragni M, Rasko JJ, et al. Successful Transduction of Liver in Hemophilia by AAV-Factor IX and Limitations Imposed by the Host Immune Response. *Nat Med* (2006) 12:342–7. doi: 10.1038/nm1358
36. O'Donnell TJ, Rubinsteyn A, Bonsack M, Riemer AB, Laserson U, Hammerbacher J. Mhcflurry: Open-Source Class I Mhc Binding Affinity Prediction. *Cell Syst* (2018) 7:129–32.e4. doi: 10.1016/j.cels.2018.05.014
37. Paul S, Croft NP, Purcell AW, Tschärke DC, Sette A, Nielsen M, et al. Benchmarking Predictions of MHC Class I Restricted T Cell Epitopes in a Comprehensively Studied Model System. *PLoS Comput Biol* (2020) 16:e1007757. doi: 10.1371/journal.pcbi.1007757
38. Weingarten-Gabbay S, Klaeger S, Sarkizova S, Pearlman LR, Chen D-Y, Bauer MR, et al. Sars-CoV-2 Infected Cells Present HLA-I Peptides From Canonical and Out-of-Frame Orfs. *bioRxiv* (2020). doi: 10.1101/2020.10.02.324145
39. Sarkizova S, Klaeger S, Le PM, Li LW, Oliveira G, Keshishian H, et al. A Large Peptidome Dataset Improves HLA Class I Epitope Prediction Across Most of the Human Population. *Nat Biotechnol* (2020) 38:199–209. doi: 10.1038/s41587-019-0322-9
40. Hui DJ, Edmonson SC, Podsakoff GM, Pien GC, Ivanciu L, Camire RM, et al. AAV Capsid CD8+ T-Cell Epitopes are Highly Conserved Across AAV Serotypes. *Mol Ther Methods Clin Dev* (2015) 2:15029. doi: 10.1038/mtm.2015.29
41. Neeffes J, Jongasma MLM, Paul P, Bakke O. Towards a Systems Understanding of MHC Class I and MHC Class II Antigen Presentation. *Nat Rev Immunol* (2011) 11:823–36. doi: 10.1038/nri3084
42. Kruzik A, Fetahagic D, Hartlieb B, Dorn S, Koppensteiner H, Horling FM, et al. Prevalence of Anti-Adeno-Associated Virus Immune Responses in International Cohorts of Healthy Donors. *Mol Ther Methods Clin Dev* (2019) 14:126–33. doi: 10.1016/j.omtm.2019.05.014
43. Rajavel K, Ayash-Rashkovsky M, Tang Y, Gangadharan B, de la Rosa M, Ewenstein B. Co-Prevalence of Pre-Existing Immunity to Different Serotypes of Adeno-Associated Virus (AAV) in Adults With Hemophilia. *Blood* (2019) 134:3349–9. doi: 10.1182/blood-2019-123666
44. Boutin S, Monteilhet V, Veron P, Leborgne C, Benveniste O, Montus MF, et al. Prevalence of Serum IgG and Neutralizing Factors Against Adeno-Associated Virus (AAV) Types 1, 2, 5, 6, 8, and 9 in the Healthy Population: Implications for Gene Therapy Using AAV Vectors. *Hum Gene Ther* (2010) 21:704–12. doi: 10.1089/hum.2009.182
45. Tse LV, Klinc KA, Madigan VJ, Castellanos Rivera RM, Wells LF, Havlik LP, et al. Structure-Guided Evolution of Antigenically Distinct Adeno-Associated Virus Variants for Immune Evasion. *Proc Natl Acad Sci USA* (2017) 114:E4812–21. doi: 10.1073/pnas.1704766114
46. Tseng Y-S, Agbandje-McKenna M. Mapping the AAV Capsid Host Antibody Response Toward the Development of Second Generation Gene Delivery Vectors. *Front Immunol* (2014) 5:9. doi: 10.3389/fimmu.2014.00009
47. Emmanuel SN, Mietzsch M, Tseng YS, Smith JK, Agbandje-McKenna M. Parvovirus Capsid-Antibody Complex Structures Reveal Conservation of Antigenic Epitopes Across the Family. *Viral Immunol* (2021) 34:3–17. doi: 10.1089/vim.2020.0022
48. Walker LM, Burton DR. Passive Immunotherapy of Viral Infections: “Super-Antibodies” Enter the Fray. *Nat Rev Immunol* (2018) 18:297–308. doi: 10.1038/nri.2017.148
49. Sivasubramanian A, Estep P, Lynaugh H, Yu Y, Miles A, Eckman J, et al. Broad Epitope Coverage of a Human In Vitro Antibody Library. *MAbs* (2017) 9:29–42. doi: 10.1080/19420862.2016.1246096
50. Bornholdt ZA, Turner HL, Murin CD, Li W, Sok D, Souders CA, et al. Isolation of Potent Neutralizing Antibodies From a Survivor of the 2014 Ebola Virus Outbreak. *Science* (2016) 351:1078–83. doi: 10.1126/science.aad5788
51. Bianchi M, Turner HL, Nogal B, Cottrell CA, Oyen D, Pauthner M, et al. Electron-Microscopy-Based Epitope Mapping Defines Specificities of Polyclonal Antibodies Elicited During HIV-1 Bg505 Envelope Trimer Immunization. *Immunity* (2018) 49:288–300.e8. doi: 10.1016/j.immuni.2018.07.009
52. Nogal B, Bianchi M, Cottrell CA, Kirchoerfer RN, Sewall LM, Turner HL, et al. Mapping Polyclonal Antibody Responses in Non-human Primates Vaccinated With HIV Env Trimer Subunit Vaccines. *Cell Rep* (2020) 30:3755–65.e7. doi: 10.1016/j.celrep.2020.02.061
53. Wine Y, Horton AP, Ippolito GC, Georgiou G. Serology in the 21st Century: The Molecular-Level Analysis of the Serum Antibody Repertoire. *Curr Opin Immunol* (2015) 35:89–97. doi: 10.1016/j.coi.2015.06.009
54. Lavinder JJ, Wine Y, Giesecke C, Ippolito GC, Horton AP, Lungu OI, et al. Identification and Characterization of the Constituent Human Serum Antibodies Elicited by Vaccination. *Proc Natl Acad Sci USA* (2014) 111:2259–64. doi: 10.1073/pnas.1317793111
55. Lee J, Boutz DR, Chromikova V, Joyce MG, Vollmers C, Leung K, et al. Molecular-Level Analysis of the Serum Antibody Repertoire in Young Adults Before and After Seasonal Influenza Vaccination. *Nat Med* (2016) 22:1456–64. doi: 10.1038/nm.4224
56. Wec AZ, Haslwanter D, Abdiche YN, Shehata L, Pedreño-Lopez N, Moyer CL, et al. Longitudinal Dynamics of the Human B Cell Response to the Yellow Fever 17D Vaccine. *Proc Natl Acad Sci USA* (2020) 117:6675–85. doi: 10.1073/pnas.1921388117
57. Piccoli L, Park Y-J, Tortorici MA, Czudnochowski N, Walls AC, Beltramello M, et al. Mapping Neutralizing and Immunodominant Sites on the SARS-CoV-2 Spike Receptor-Binding Domain by Structure-Guided High-Resolution Serology. *Cell* (2020) 183:1024–42.e21. doi: 10.1016/j.cell.2020.09.037
58. Goodwin E, Gilman MSA, Wrapp D, Chen M, Ngwuta JO, Moin SM, et al. Infants Infected With Respiratory Syncytial Virus Generate Potent Neutralizing Antibodies That Lack Somatic Hypermutation. *Immunity* (2018) 48:339–49.e5. doi: 10.1016/j.immuni.2018.01.005
59. Miho E, Yermanos A, Weber CR, Berger CT, Reddy ST, Greiff V. Computational Strategies for Dissecting the High-Dimensional Complexity

- of Adaptive Immune Repertoires. *Front Immunol* (2018) 9:224. doi: 10.3389/fimmu.2018.00224
60. Xu GJ, Kula T, Xu Q, Li MZ, Vernon SD, Ndung'u T, et al. Viral Immunology. Comprehensive Serological Profiling of Human Populations Using a Synthetic Human Virome. *Science* (2015) 348:aaa0698. doi: 10.1126/science.aaa0698
 61. Tseng Y-S, Gurda BL, Chipman P, McKenna R, Afione S, Chiorini JA, et al. Adeno-Associated Virus Serotype 1 (AAV1)- and AAV5-antibody Complex Structures Reveal Evolutionary Commonalities in Parvovirus Antigenic Reactivity. *J Virol* (2015) 89:1794–808. doi: 10.1128/JVI.02710-14
 62. Collins AM, Wang Y, Roskin KM, Marquis CP, Jackson KJL. The Mouse Antibody Heavy Chain Repertoire is Germline-Focused and Highly Variable Between Inbred Strains. *Philos Trans R Soc Lond B Biol Sci* (2015) 370:1676. doi: 10.1098/rstb.2014.0236
 63. Robbani DF, Gaebler C, Muecksch F, Lorenzi JCC, Wang Z, Cho A, et al. Convergent Antibody Responses to SARS-CoV-2 in Convalescent Individuals. *Nature* (2020) 584:437–42. doi: 10.1038/s41586-020-2456-9
 64. Parameswaran P, Liu Y, Roskin KM, Jackson KKL, Dixit VP, Lee J-Y, et al. Convergent Antibody Signatures in Human Dengue. *Cell Host Microbe* (2013) 13:691–700. doi: 10.1016/j.chom.2013.05.008
 65. Setliff I, McDonnell WJ, Raju N, Bombardi RG, Murji AA, Scheepers C, et al. Multi-Donor Longitudinal Antibody Repertoire Sequencing Reveals the Existence of Public Antibody Clonotypes in HIV-1 Infection. *Cell Host Microbe* (2018) 23:845–54.e6. doi: 10.1016/j.chom.2018.05.001
 66. Faust SM, Bell P, Cutler BJ, Ashley SN, Zhu Y, Rabinowitz JE, et al. CpG-depleted Adeno-Associated Virus Vectors Evade Immune Detection. *J Clin Invest* (2013) 123:2994–3001. doi: 10.1172/JCI68205
 67. Chan YK, Wang SK, Chu CJ, Copland DA, Letizia AJ, Costa Verdera H, et al. Engineering Adeno-Associated Viral Vectors to Evade Innate Immune and Inflammatory Responses. *Sci Transl Med* (2021) 13:580. doi: 10.1126/scitranslmed.abd3438
 68. Kishimoto TK. Development of ImmTOR Tolerogenic Nanoparticles for the Mitigation of Anti-Drug Antibodies. *Front Immunol* (2020) 11:969. doi: 10.3389/fimmu.2020.00969
 69. Leborgne C, Barbon E, Alexander JM, Hanby H, Delignat S, Cohen DM, et al. IgG-cleaving Endopeptidase Enables In Vivo Gene Therapy in the Presence of anti-AAV Neutralizing Antibodies. *Nat Med* (2020) 26:1096–101. doi: 10.1038/s41591-020-0911-7
- Conflict of Interest:** AW, KL, JK, SS, JG and EK are employees and shareholders in Dyno Therapeutics Inc.
- Copyright © 2021 Wec, Lin, Kwasnieski, Sinai, Gerold and Kelsic. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*