

Exploring protein orthogonality in immune space: a case study with AAV and Cas9 orthologs

Ana M. Moreno^{a,*}, Nathan Palmer^{b,*}, Fernando Alemán^c, Genghao Chen^a, Andrew Pla^a, Wei Leong Chew^d, Mansun Law^c, Prashant Mali^{a,†}

^aDepartment of Bioengineering, University of California San Diego, CA, USA.

^bDivision of Biological Sciences, University of California San Diego, CA, USA.

^cDepartment of Immunology and Microbiology, The Scripps Research Institute, CA, USA.

^dSynthetic Biology, Genome Institute of Singapore, Singapore.

*These authors contributed equally.

†Correspondence: pmali@ucsd.edu.

ABSTRACT

A major hurdle in protein-based therapeutics is the interaction with the adaptive immune system, which can lead to neutralization by circulating antibodies and clearance of treated cells by cytotoxic T-lymphocytes. One method of circumventing these issues is to use human or humanized proteins which avoid the immune response by self-recognition. However, this approach limits potential protein therapeutics to those of human origin, excluding many exciting effectors and delivery vehicles such as CRISPR-Cas9 and adeno-associated viruses (AAVs). To address this issue, we propose here the sequential use of orthologous proteins whose function is constrained by natural selection, but whose structure is subject to diversification by genetic drift. This would, in principle, allow for repeated treatments by ‘immune orthogonal’ orthologs without reduced efficacy due to lack of immune cross-reactivity among the proteins. To explore and validate this concept we chose 91 Type II CRISPR-Cas9 orthologs and 167 AAV capsid protein orthologs, and developed a pipeline to compare total sequence similarity as well as predicted binding to class I and class II Major Histocompatibility Complex (MHC) proteins. Interestingly, MHC binding predictions revealed wide diversity among the set of Cas9 orthologs, with 83% of pairs predicted to have non cross-reacting immune responses, while no global immune orthogonality among AAV serotypes was observed. To confirm these findings we selected two Cas9 orthologs, from *S. pyogenes* and *S. aureus*, predicted to be orthogonal in immune space, and delivered them into mice via multiple AAV serotypes. We observed cross-reacting antibodies against AAV but not Cas9 orthologs in sera from immunized mice, validating the computationally predicted immune orthogonality among these proteins. Moving forward, we anticipate this framework can be applied to prescribe sequential regimens of immune orthogonal

protein therapeutics to circumvent pre-existing or induced immunity, and eventually, to rationally engineer immune orthogonality among protein orthologs.

INTRODUCTION

Protein therapeutics, including protein-based gene therapy, have several advantages over small-molecule drugs. They generally serve complex, specific functions, and have minimal off-target interference with normal biological processes. However, one of the fundamental challenges to any protein-based therapeutic is the interaction with the adaptive immune system. Neutralization by circulating antibodies through B-cell activation and clearance of treated cells by CD8⁺ cytotoxic T-lymphocytes (CTLs) create a substantial barrier to effective protein therapies¹⁻⁴. Although the delay in the adaptive immune response to novel proteins may allow sufficient time for the initial dose to work, subsequent doses face faster and stronger secondary immune responses due to the presence of memory T- and B- cells. In addition, gene transfer studies have shown that host immune responses against the delivery vector and/or therapeutic transgene can eliminate treated cells, thus limiting the efficacy of the treatment⁵⁻¹⁰.

A common approach to circumventing these issues has been to utilize human proteins, or to humanize proteins by substitution of non-human components^{11,12}. However, this approach is limited to a small set of therapeutic proteins naturally occurring in humans or closely related species. In addition, although the humanization of proteins can result in a significantly less immunogenic product, they still carry immunological risk¹². Another way to circumvent an immune response to protein therapeutics is the removal of immunogenic T cell epitopes^{13,14}. Once immunogenic T cell epitopes are identified, substitution of key amino acids may reduce the protein's immunogenicity since modification of amino acids at critical anchor residues can abrogate binding to MHC molecules and prevent antigen presentation. However, this can prove difficult due to the massive diversity at HLA loci. As epitope engineering must account for the substrate specificity of each different HLA allele, therapeutics would likely have to be uniquely modified for each patient. All the same, epitope deletion has been successfully applied to several proteins¹⁵, but can only preserve protein function when limited to small numbers of HLA alleles unrepresentative of the full diversity. Structural modifications such as PEGylation have also been known to reduce immunogenicity by interfering with antigen-processing mechanisms. However, there is evidence that PEG-specific antibodies are elicited in patients treated with PEGylated therapeutic enzymes¹⁶⁻¹⁹.

Furthermore, protein therapies have required repeated treatments due to degradation of the protein or turnover of treated cells, or, in the case of gene therapy, reduced expression of

the transgene^{20,21}. This provides an even greater challenge as repeated exposure to the same antigen can elicit a more robust secondary immune response²², which may completely inhibit subsequent dosage or even sensitize the immune system to antigens remaining from the initial exposure. In order to facilitate efficacious repeat protein therapies, we propose the use of orthologous proteins whose function is constrained by natural selection, but whose structure is subject to diversification by genetic drift. An ortholog, given sufficient sequence divergence, will not cross-react with the immune response generated by exposure to the others, allowing repeat doses to avoid neutralization by existing antibodies and treated cells to avoid clearance by activated CTLs.

As a case study for exploring this approach we focused on the CRISPR-Cas9 system, perhaps the most anticipated therapeutic for gene editing^{23–33}. Comparative genomics has demonstrated that Cas9 proteins are widely distributed across bacterial species and have diversified over an extensive evolutionary history^{34–37}. Although there may be pre-existing immunity to Cas9s from some commensal species³⁸, we hypothesized this diversity could provide a mechanism to circumvent inducing immunological memory by utilizing orthologous Cas9 proteins for each treatment. Additionally, the immunogenicity due to the delivery vehicle or administration route for the Cas9 and the associated guide RNA (gRNA) must also be considered. In this regard, adeno-associated viruses (AAVs) have emerged as a highly preferred vehicle for gene delivery, as these are associated with low immunogenicity and toxicity^{8,9}, which promotes long-term transgene expression^{39,40} and treatment efficacy. Despite the relatively low immunogenicity of AAV vectors, antibodies against both the capsid and transgene may still be elicited^{41–46}. Additionally, the prevalence of neutralizing antibodies (NAB) against AAVs in the human population⁴⁷ and cross-reactivity between serotypes⁴⁸ remains a hurdle for efficacious AAV therapy. Although AAVs were initially considered non-immunogenic due to their poor transduction of antigen-presenting cells (APCs)⁴⁹, it is now known that they can transduce dendritic cells (DCs)⁵⁰ and trigger innate immune responses through Toll-like receptor (TLR) signaling pathways⁵¹. The ability to transduce DCs is dependent on AAV serotype and genome, and may be predictive of overall immunogenicity⁵².

To evaluate the immune orthogonality of AAV-delivered CRISPR-Cas systems, we analyzed 91 Cas9 orthologs, and 167 AAV VP1 orthologs. By comparing total sequence similarity as well as predicted binding strengths to class I and class II MHC molecules, we constructed graphs of immune cross-reactivity and computed cliques of proteins that are orthogonal in immunogenicity profiles. Although MHC epitopes do not predict antibody epitopes, the induction of the more powerful memory response is primarily dependent on reactivation of

memory B-cells with help from memory T-cells through the presentation of antigens on class II MHC molecules^{53,54}. Finally, we experimentally confirmed our immunological predictions by assaying treated mice for induction of protein-targeting antibodies.

RESULTS

Humoral immune response to AAV and Cas9

One of the major obstacles for sequential gene therapy treatments is the presence of neutralizing antibodies against the delivery vehicle and transgene cargo induced by the first administration of the therapy. To determine the humoral immune response kinetics to the AAV-8 capsid and the Cas9 transgene, we first injected C57BL/6J mice retro-orbitally with 10^{12} vg of AAV-8-SaCas9 targeting proprotein convertase subtilisin/kexin type 9 (PCSK9), a promising gene target that when disrupted can reduce Low Density Lipoprotein (LDL) levels and protect against cardiovascular disease. Consistent with a previous study⁵⁵, mice had reduced PCSK9 serum levels as early as one week post-injection due to successful SaCas9 mediated gene-editing, which was sustained for the entire duration of the experiment (4 weeks) (**Figure 1C**). We noted that a subset of the mice developed IgG1 antibodies against the SaCas9 protein (**Figure 1D**). Additionally, mice developed humoral immunity to the AAV8 capsid withing one week post-injection (**Figure 1E**). To evaluate the feasibility of multiple dosing with AAV-Cas9, we next investigated whether immune orthogonal sets of AAV and Cas9 orthologs exist.

Identifying immune-orthogonal proteins

Natural selection produces diverse structural variants with conserved function in the form of orthologous genes. We assayed the relevance of this diversity for immunological cross-reactivity of 91 Type II Cas9 orthologs and 167 AAV orthologs by first comparing their overall amino acid sequence similarities, and second, using a more specific constraint of how their respective amino acid sequences are predicted to bind MHC Type I and II molecules (**Figure 1F**). From these analyses we obtained first an estimate of the comprehensive immune overlap among Cas9 and AAV orthologs based purely at the sequence level, and second a more stringent estimate of predicted immune overlap based on predicted MHC binding. By sequence-level clustering and clique finding methods, we defined many sets of Cas9 orthologs containing up to 9 members with no 6-mer overlap (**Figure S1**). Notably, based on MHC-binding predictions, we find among the set of Cas9 orthologs that 83% of pairs are predicted to have non cross-reacting immune responses, i.e. they are predicted to be orthogonal in immune space (**Figure 1G**). On the contrary, among AAV capsid (VP1 protein) orthologs we did not find full

orthogonality up to the 14-mer level, even when restricting predictions with MHC-binding strengths (**Figure 1H**), likely reflecting the strong sequence conservation and shorter evolutionary history of AAVs⁵⁶. This analysis suggests, consistent with previous observations^{57,58}, that exposure to one AAV serotype can induce broad immunity to all AAVs, which presents a significant challenge to AAV delivery platforms, as some serotypes are prevalent in human populations. Despite the most divergent AAV serotype (AAV-5) showing the fewest shared immunogenic peptides, there remain tracts of sequences fully conserved within the VP1 orthologs. As expected, predicted immune cross-reaction negatively correlates with phylogenetic distance (**Figure S2**), though there is significant variation not captured by that regression, suggesting that MHC-binding predictions can refine the choice of sequential orthologs beyond phylogenetic distance alone.

Confirming humoral immune-orthogonality among Cas9 proteins

To test our immunological predictions and to establish the utility of this approach, we narrowed in on a 5-member clique containing the ubiquitously used *S. pyogenes* Cas9 in addition to the well-characterized *S. aureus* Cas9 (**Figure S1**). To determine whether either of these proteins have cross-reacting antibody responses, we injected mice with 10^{12} vg of either AAV8 or AAVDJ capsids containing SaCas9 or SpCas9 transgenes via retro-orbital injections and harvested serum at days 0 (pre-injection), and periodically over 4-6 weeks (**Figure 2A**). SpCas9-specific antibodies were detected in the plasma of all mice injected with SpCas9 (n=6), and notably none of the mice injected with SaCas9 (n=12) (**Figure 2B**). Although SaCas9 appeared to induce a weaker response, as only half of the mice injected with SaCas9 AAVs (n=12) developed detectable antibodies against SaCas9, none of the mice injected with SpCas9 AAVs (n=6) developed an antibody response against SaCas9. These results were confirmed in an independent study in which SpCas9-specific antibodies, but not SaCas9-specific antibodies, were detected in the plasma of mice injected with AAV-SpCas9 (n=12). These mice were injected retro-orbitally with 10^{12} vg of AAV8-SpCas9 or AAVDJ-SpCas9, and also received an additional intramuscular injection with 10^{11} vg at week 4. (**Figure 2C**). Taken together, our data confirms that SpCas9 and SaCas9 have humoral immune-orthogonality.

Broad cross-reactivity among AAV serotypes

AAVs are becoming a preferred delivery vehicle due to their ability to avoid induction of a strong CD8⁺ T-cell response, however, the presence of neutralizing antibodies remains a significant barrier to successful application of AAV therapies. Consistent with previous results⁵⁷,

we found shared immunogenic peptides among all the various human AAV serotypes, (**Figure S3**). We confirmed the lack of orthogonality for two serotypes, AAV8 and AAVDJ, in which we found that antibodies produced in mice injected with AAV8 and AAVDJ react to both AAV8 and AAVDJ antigens (**Figure 2D**). Our analysis suggests that there are no two known AAVs for which exposure to one would guarantee immune naïveté to another across all HLA genotypes. However, immune cross-reaction could be minimized through the use of AAV5^{58,59}, the most phylogenetically divergent serotype. Our predictions identify only a single shared highly immunogenic peptide between AAV5 and the commonly used AAV2 and AAV8 in the mouse model (though several other shared peptides of mild MHC affinity exist). We confirmed this via ELISAs, where mice injected with AAV2 did not elicit antibodies against AAV5 and AAV8, and mice injected with AAV5 did not elicit antibodies against AAVDJ and AAV8 (**Figure 2E**).

DISCUSSION

The use of protein therapeutics requires ways to evade the host's immune response. Cas9, as an example, has prokaryotic origins and can evoke a T-cell response, which may lead to clearance of transduced cells. In addition, circulating antibodies can neutralize the AAV vector and prevent efficient transduction upon repeated doses. Immunosuppressive drugs could mitigate some of these aspects, but not without significant side-effects, as well as not being applicable to patients in poor health⁶⁰⁻⁶³. Similar to what has been done in cancer antibody therapeutics⁶⁴, the SpCas9 protein could also be de-immunized by swapping high-immunogenicity domains. This is a promising approach, however, it will be complex and laborious as we anticipate tens of mutations to achieve stealth, and could result in a reduction in activity and an overall less effective therapy.

To circumvent this issue, we developed here a framework to compare protein orthologs and their predicted binding to MHC I and MHC II by checking a sliding window of all k-mers in a protein for their presence in another, focusing on peptides predicted to bind to at least one MHC allele. Through this analysis, we identified cliques of Cas9 proteins that are immune orthogonal. Based on these predictions, specific T-cell responses from one ortholog would not cross-react with another ortholog of the same clique, preventing the re-activation of CD8+ cytotoxic T-cells, as well as the CD4+ T-cell help necessary to re-activate memory B-cells. We confirmed these results through ELISAs, and verified two well-characterized Cas9 proteins to be immune orthogonal, SpCas9 and SaCas9. Therefore, we expect that proteins belonging to the same clique can be used sequentially without eliciting memory T- and B- cell responses.

Due to the importance of AAVs as a delivery agent in gene therapy, we also analyzed AAV serotypes through our MHC I and II comparison framework, and have demonstrated that no two AAVs are mutually immune orthogonal. However, with a known HLA genotype, it may be possible to define a personalized regimen of immune orthogonal AAVs using currently defined serotypes. For instance, use of AAV5 minimizes immune cross-reactivity in mice and non-human primates, as demonstrated by a recent study in which chimeric-AAV5 immunized mice and non-human primates successfully received a second dose of treatment with AAV1⁵⁹. However, in the human setting we predict that there may be substantially more immune overlap between AAV5 and other AAVs. Additionally, it has been shown that memory B-cells heavily contribute to the antibody response to similar but not identical antigens⁶⁵, indicating that partial orthogonality may not be sufficient. Our analysis suggests that creating a pair of globally orthogonal AAV capsids for human application would require >10 mutations in one of the two proteins. This hypothetical orthogonal AAV capsid presents a substantial engineering challenge, as it requires mutating many of the most conserved regions to achieve immune orthogonality.

Previous work has identified that MHC affinity is highly dependent on anchor residues at either end of the binding pocket⁶⁶. Residue diversity is more tolerated in the center of the binding pocket, though it may be these residues that most impact antigen specificity, as it is thought that they are central to interaction with the T-cell receptor (TCR). Comparing the number of orthologous pairs in 9-mer space with the number of predicted orthologous pairs based on class II binding predictions suggests that only approximately 65% of 9-mer peptides serve as appropriate MHC class II binding cores, even across the thousands of HLA-2 combinations we explore here. This under-sampling of peptide space by MHC molecules likely reflects the requirement for hydrophobic anchor residues and leaves some space for protein de-immunization by mutation of immunogenic peptides to ones which never serve as MHC binding cores. Achieving this while preserving protein function however, has proven difficult even for few HLA alleles, and remains a major protein engineering challenge. New technologies for directly measuring TCR affinity with MHC-presented antigens⁶⁷ will also further clarify the key antigenic peptides contributing to the immune response, and will be useful to inform approaches here.

We also note some limitations to our work. Mainly, we have used inbred C57BL/6J as our mice model, which have very limited MHC diversity⁶⁸, and might not recapitulate other human immunological features, such as differences in antigen processing and presentation. In this regard, we attempted to measure the T-cell response with the IFN- γ ELISPOT assay for a subset of predicted MHCI and MHC II peptides. Concordant with a recent study of pre-existing immunity to Sp- and SaCas9 in humans³⁸, we observed a CD4+ T-cell response against some

MHCII peptides with mice injected with SaCas9, but not SpCas9. Although we note that the C57BL/6J mice did not show robust responses in general to the AAV-CRISPRs (**Figure S4**), the detection of antibody and T-cell responses to both Cas9³⁸ (only SaCas9 in the case of T-cells) and AAVs^{1,52} in humans demonstrates that the immune consequences of treatment will be critical for success in human therapy. Moving forward, this work can be potentially repeated using other mouse models, such as mice expressing human HLA allotypes, however, these models come with their own technical challenges, such as restricted HLA alleles (representing only main MHC II subgroups) as well as a restricted TCR repertoire⁶⁸. In addition, B-cell epitopes can also be predicted and incorporated into immune orthogonality analysis. However, since B-cell epitopes may be both linear and conformational, these are more difficult to predict. Advances and further validation of these *in silico* models will allow for better predictions in the future^{69–73}. Finally, recent work has indicated that MHC class I peptides may have significant contribution from spliced host and pathogen-derived peptides created by proteasomal processing⁷⁴. It is unclear how this may affect cross-recognition of proteins we predict to be immune orthogonal. On the one hand, it provides a mechanism whereby very short antigenic sequences spliced to the same host protein may result in cross-recognition of substantially different foreign antigens, however, we expect this to be unlikely due to the massive number of possible spliced peptides between the antigen and entire host proteome.

Overall, we believe our framework provides a potential solution for efficacious gene therapy, not solely for Cas9-mediated genome engineering, but also for other protein therapeutics that might necessitate repetitive treatments. Although using this approach still requires mitigating the primary immune response, particularly CTL clearance, we expect that epitope deletion and low-immunogenicity delivery vectors such as AAVs will mitigate this problem, and the potential for repeated dosage will reduce the need for very high first-dose efficiency.

ACKNOWLEDGEMENTS

We thank members of the Mali lab for advice and help with experiments, and the Salk GT3 viral core for help with AAV production. This work was supported by UCSD Institutional Funds, the Burroughs Wellcome Fund (1013926), the March of Dimes Foundation (5-FY15-450), the Kimmel Foundation (SKF-16-150), and NIH grants (R01HG009285, RO1CA222826). A.M. acknowledges a graduate fellowship from CONACYT and UCMEXUS. A.M., N.P., and P.M. have filed patents based on this work.

METHODS

Computational Methods

For Cas9, we chose 91 orthologs cited in exploratory studies cataloguing the diversity of the Cas9 protein⁷⁵, including several that are experimentally well-characterized. For AAVs, we analyzed 167 sequences, focusing in on all 13 characterized human serotypes, as well as one isolate from rhesus macaque (rh32), one engineered variant (DJ), and one reconstructed ancestral protein (Anc80L65). We then compared total sequence similarity (immunologically uninformed) as well as predicted binding to class I and class II MHC molecules (immunologically informed) between these proteins. Immunologically uninformed sequence comparison was carried out by checking a sliding window of all contiguous k-mers in a protein for their presence in another protein sequence with either zero or one mismatch. Immunologically informed comparison was done in a similar fashion, but using only those k-mers predicted to bind to at least one of 81 HLA-1 alleles using netMHC 4.0⁷⁶ for class I (alleles can be found at http://www.cbs.dtu.dk/services/NetMHC/MHC_allele_names.txt), and at least one of 5,620 possible MHC II molecules based on 936 HLA-2 alleles using netMHCIIpan 3.1⁷⁷ for class II (alleles can be found at http://www.cbs.dtu.dk/services/NetMHCIIpan-3.1/alleles_name.list). We compared the use of netMHC to alternative immune epitope prediction platforms such as the Immune Epitope Database (iedb.org)⁷⁸ and found very strong agreement across software. Ultimately, we chose netMHC because of the larger number of HLA alleles it supports. Sequences were defined as binding if the predicted affinity ranked in the top 2% of a test library of 400,000 random peptides as suggested in the software guidelines. Generation of immune orthogonal cliques was carried out using the Bron-Kerbosch algorithm. Briefly, a graph was constructed with each ortholog as a vertex, where the edges are defined by the number of shared immunogenic peptides between the connecting vertices. Sets of proteins for which every pair in the set is immune orthogonal constitutes a clique. Phylogenetic distance between protein sequences was measured using the BLOSUM 62 matrix excluding indels.

Experimental Methods

AAV Production

AAV2/8, AAV2/2, AAV2/5, AAV2/DJ virus particles were produced using HEK293T cells via the triple transfection method and purified via an iodixanol gradient⁷⁹. Confluency at transfection was between 80% and 90%. Media was replaced with pre-warmed media 2 hours before transfection. Each virus was produced in 5 x 15 cm plates, where each plate was transfected with 7.5 µg of pXR-capsid (pXR-8, pXR-2, pXR-5, pXR-DJ), 7.5 of µg recombinant

transfer vector, and 22.5 µg of pAd5 helper vector using PEI (1µg/µL linear PEI in 1x DPBS pH 4.5, using HCl) at a PEI:DNA mass ratio of 4:1. The mixture was incubated for 10 minutes at RT and then applied dropwise onto the media. The virus was harvested after 72 hours and purified using an iodixanol density gradient ultracentrifugation method. The virus was then dialyzed with 1x PBS (pH 7.2) supplemented with 50 mM NaCl and 0.0001% of Pluronic F68 (Thermo Fisher) using 100kDA filters (Millipore), to a final volume of ~1 mL and quantified by qPCR using primers specific to the ITR region, against a standard (ATCC VR-1616).

AAV-ITR-F: 5'-CGGCCTCAGTGAGCGA-3' and

AAV-ITR-R: 5'-GGAACCCCTAGTGATGGAGTT-3'.

Animal studies

All animal procedures were performed in accordance with protocols approved by the Institutional Animal Care and Use Committee (IACUC) of the University of California, San Diego. All mice were acquired from Jackson labs. AAV injections were done in adult C57BL/6J mice (10 weeks) through retro-orbital injections using 1×10^{12} vg/mouse.

ELISA

PCSK9: Levels of serum PCSK9 were measured using the Mouse Proprotein Convertase 9/PCSK9 Quantikine ELISA kit (R&D Systems) according to manufacturer's guidelines. Briefly, serum samples were diluted 1:200 in Calibrator diluent and allowed to bind for 2 h onto microplate wells that were precoated with the capture antibody. Samples were then sequentially incubated with PCSK9 conjugate followed by the PCSK9 substrate solution with extensive intermittent washes between each step. The amount of PCSK9 in serum was estimated colorimetrically using a standard microplate reader (BioRad iMark).

Cas9 and AAV: Recombinant SpCas9 protein (PNA Bio, cat. no. CP01), or SaCas9 protein (ABM good, cat no. K144), was diluted in 1x coating buffer (Bethyl), and 0.5 µg was used to coat each well of 96-well Nunc MaxiSorp Plates (ab210903) overnight at 4 °C. For AAV experiments, 10^9 vg of AAV-2, -5, -8 or -DJ in 1x coating buffer was used to coat each well of 96-well Nuc MaxiSorp Plates. Plates were washed three times for 5 min with 350 µl of 1x Wash Buffer (Bethyl) and blocked with 300 µl of 1x BSA Blocking Solution (Bethyl) for 2 h at RT. The wash procedure was repeated. Serum samples were added at 1:40 dilution, and plates were incubated for 5 h at 4 °C with shaking. Wells were washed three times for 5 min, and 100 µl of HRP-labeled goat anti-mouse IgG1 (Bethyl; diluted 1:100,000 in 1% BSA Blocking Solution) was added to each well. After incubating for 1hr at RT, wells were washed four times for 5 min, and 100 µl of TMB Substrate (Bethyl) was added to each well. Optical density (OD) at 450 nm was measured using a plate reader (BioRad iMark).

Epitope prediction and peptide synthesis

The MHC-binding peptides for our mouse model were predicted using the netMHC-4.0 and netMHCIpan-3.1 online software with the alleles H-2-Db and H-2-Kb for class I and H-2-IAb for class II. For MHCII, the top 10 peptides for Sp- and SaCas9 and top 5 peptides for AAV-8 and AAV-DJ by percentile binding were selected for synthesis by Synthetic Biomolecules as crude materials. For MHCI, we selected the top 20 peptides for Sp- and SaCas9 and the top 10 for AAV-8 and AAV-DJ. All peptides were dissolved in DMSO with a concentration of 40 mg ml⁻¹ and stored at -20 °C.

IFN-γ ELISPOT assay

CD8+ T cells were isolated from splenocytes using magnetic bead positive selection (Miltenyi Biotec) 6 weeks after virus infection. A total of 2×10^5 CD8+ T cells were stimulated with 1×10^5 LPS-blasts loaded with 10 µg of individual peptide in 96-well flat-bottom plates (Immobilon-P, Millipore) that were coated with anti-IFN-γ mAb (clone AN18, Mabtech) in triplicate. Concanavalin A (ConA) was used as positive control. After 20 h of incubation, biotinylated anti-mouse IFN-γ mAb (R4-6A2; Mabtech), followed by ABC peroxidase (Vector Laboratories) and then 3-amino-9-ethylcarbazole (Sigma-Aldrich) were added into the wells. Responses are expressed as number of IFN-γ SFCs per 1×10^6 CD8+ T cells.

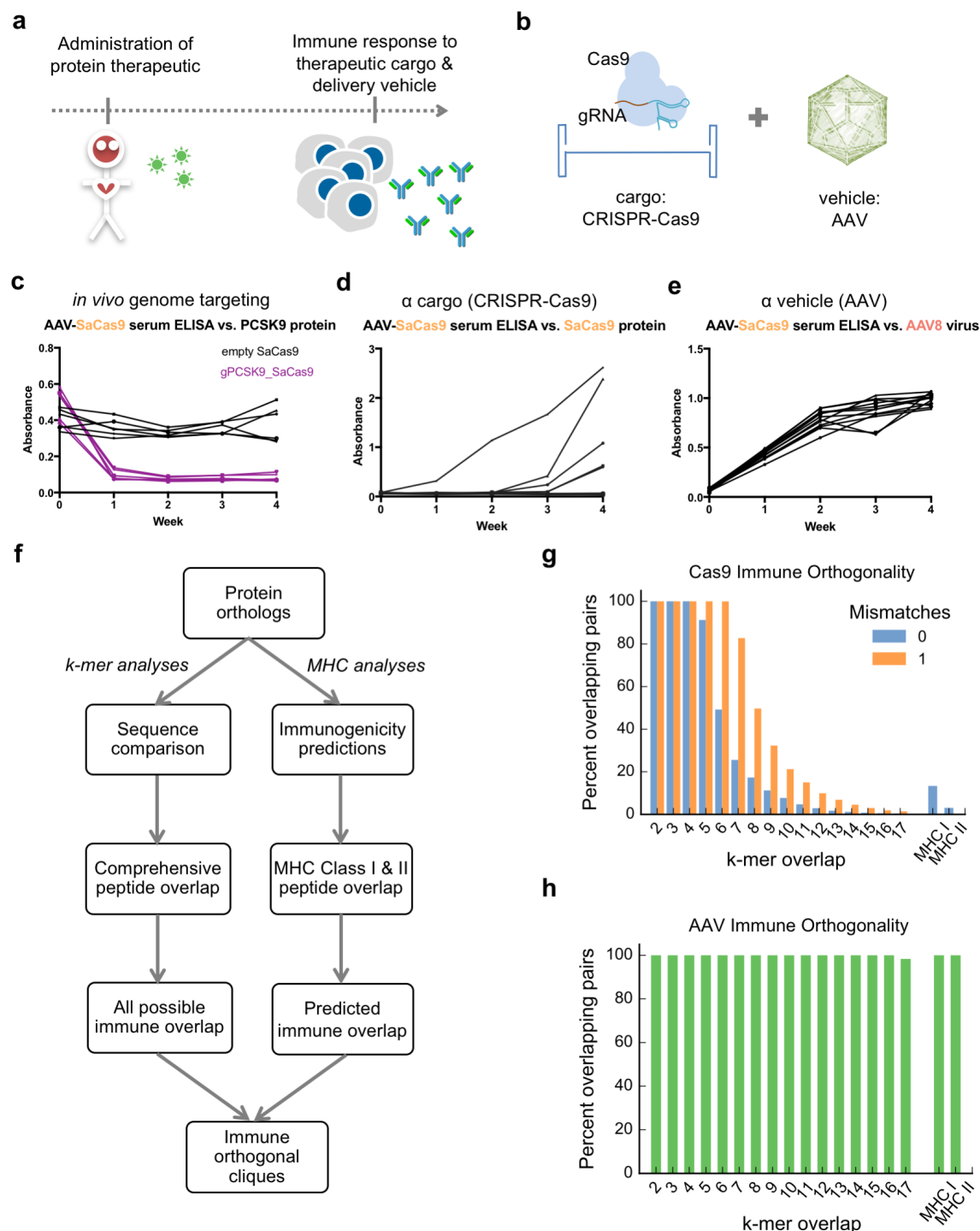


Figure 1: Protein based therapeutics elicit an adaptive immune response: experimental and *in silico* analyses. (a) Proteins have substantial therapeutic potential, but a major drawback is the immune response to both the therapeutic protein and its delivery vehicle. (b) As a case study, we explored the CRISPR-Cas9 systems and corresponding delivery vehicles based on AAVs. (c) Mice were injected retro-orbitally with 10^{12} vg/mouse of AAV8-SaCas9

targeting the PCSK9 gene or a non-targeting control (empty vector). A decrease in PCSK9 serum levels, due to successful gene targeting, can be seen in mice receiving AAV-SaCas9-PCSK9 virus (n=6 mice for each group). **(d)** Immune response to the payload was detected in ELISAs for the SaCas9 protein. (n=12) **(e)** Immune response to the delivery vehicle was detected in ELISAs for the AAV8 virus capsid (n=12 mice). **(f)** *In silico* workflow used to find immune orthogonal protein homolog cliques. **(g)** Immunologically uninformed sequence comparison was carried out by checking all *k*-mers in a protein for their presence in another protein sequence with either zero or one mismatch. The x-axis corresponds to *k*, while MHC I and MHC II show overlap only of peptides predicted to bind to MHC class I and class II molecules. 48% of Cas9 pairs show no 6-mer overlap, and 83% of pairs show no overlapping MHC-binding peptides. **(h)** Same as (g) but for AAV VP1 capsid proteins. All AAV pairs contain overlapping MHC-binding peptides.

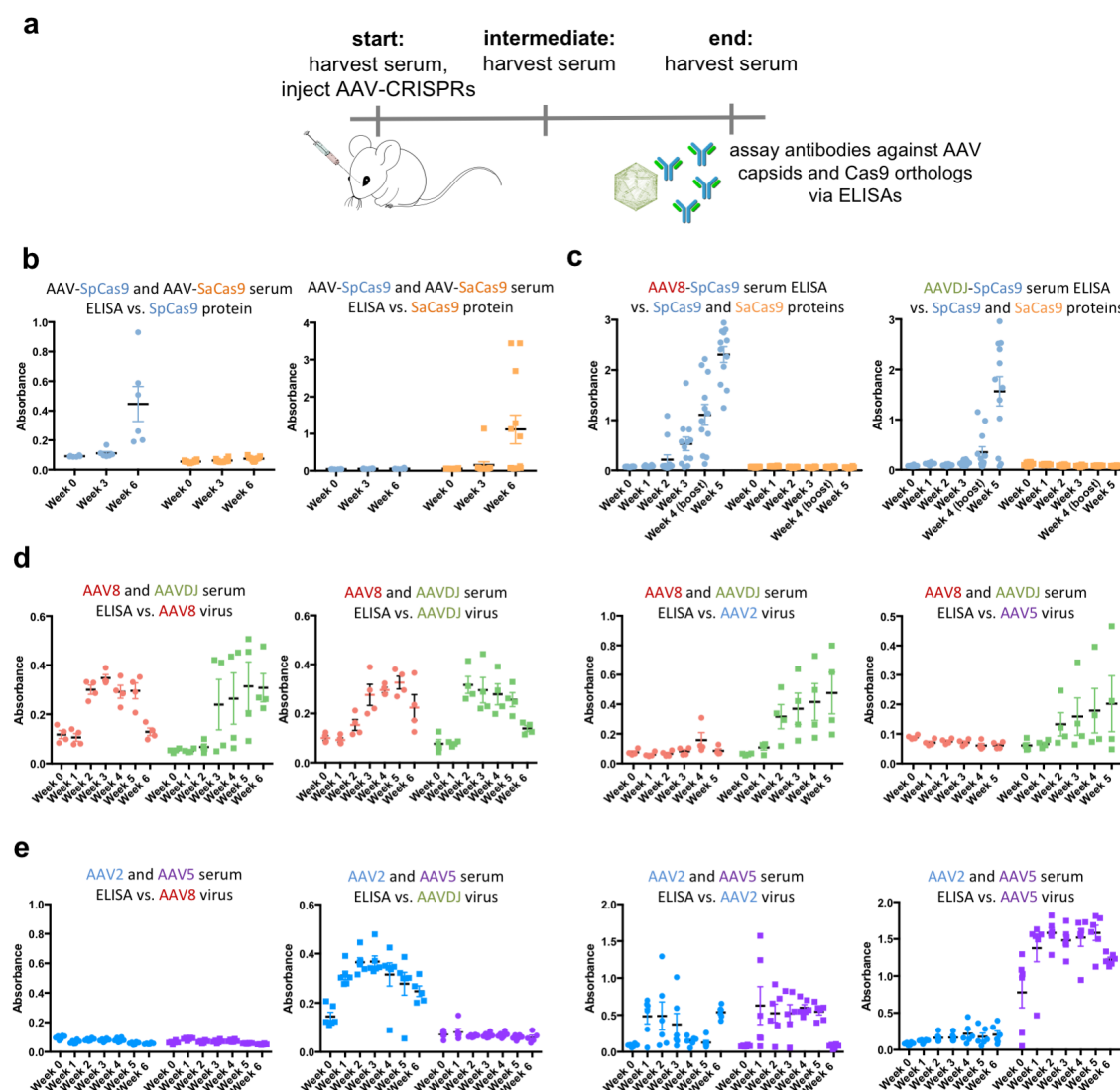
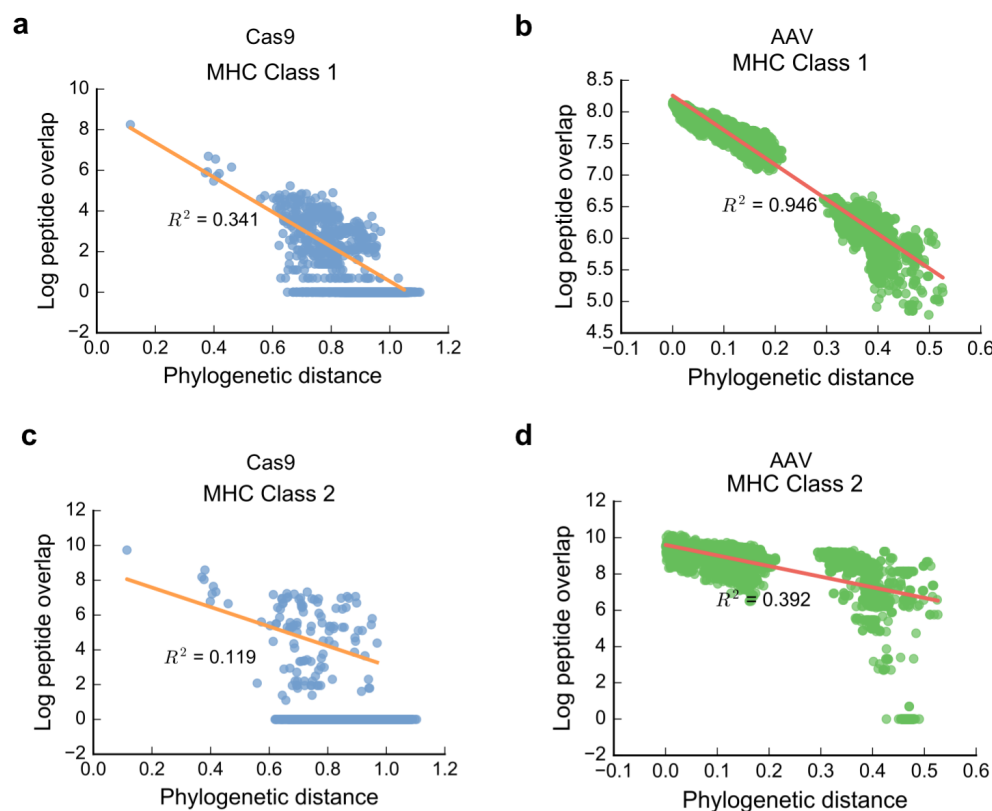
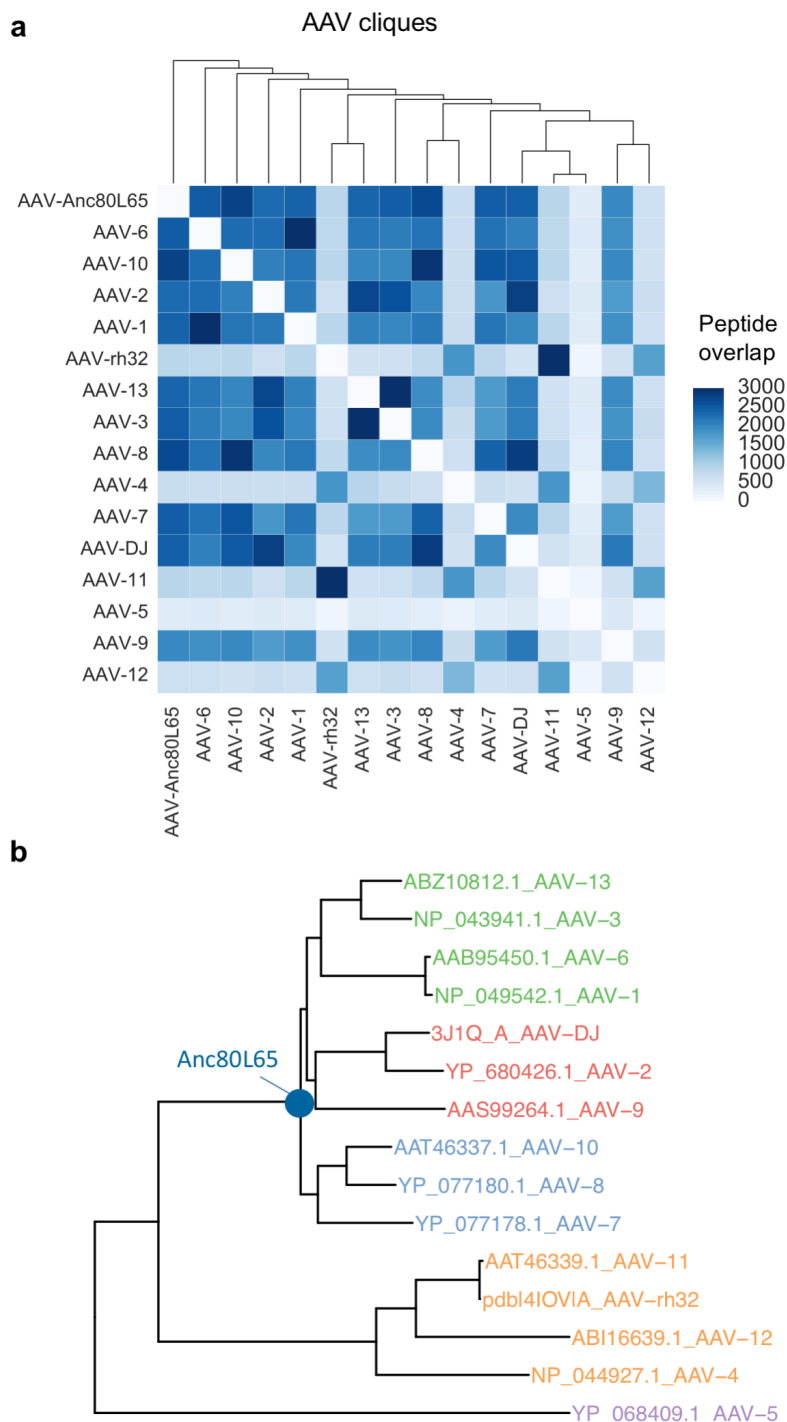


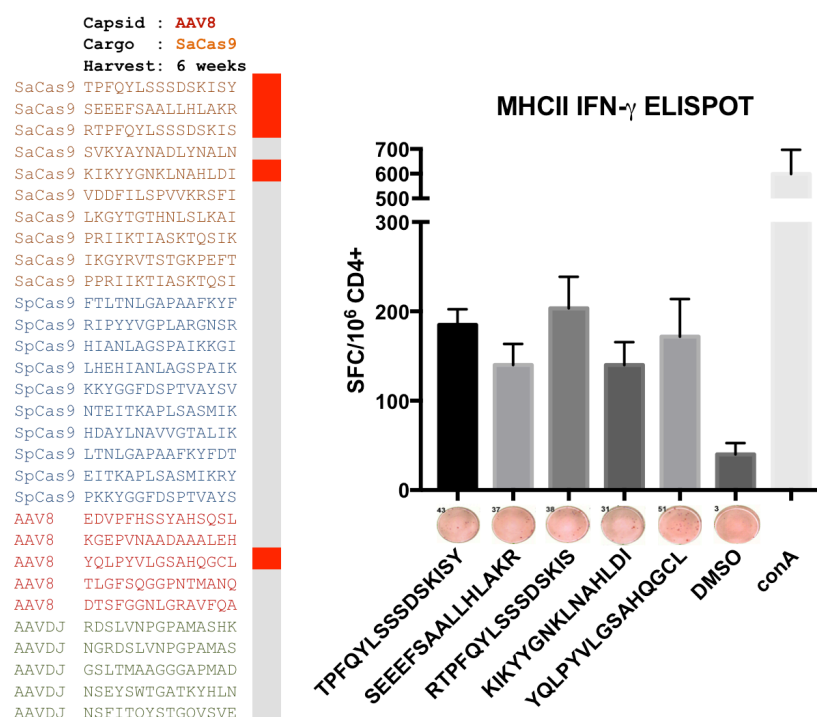
Figure 2: Experimental validation of Cas9 and AAV immunogenicity predictions. (a) Mice were exposed to antigens via retro-orbital injections at 10^{12} vg/mouse. Serum was harvested prior to injection on day 0, and at multiple points over the course of 4-6 weeks. (b) anti-SpCas9 antibodies generated in mice injected with SpCas9 (n=6) and SaCas9 (n=12), and anti-SaCas9 antibodies generated in mice injected with SpCas9 (n=6) and SaCas9 (n=12). (c) anti-SpCas9 and anti-SaCas9 antibodies generated by mice injected with AAV8 SpCas9 (n=12; left panel), or AAVDJ SpCas9 (n=12; right panel). (d) anti-AAV8/DJ/2/5 antibodies generated against mice injected with AAV8 or AAVDJ (n=4 for all panels). (e) anti-AAV8/DJ/2/5 antibodies generated against mice injected with AAV2 or AAV5 (n=5 for all panels).



Supplementary Figure 2: In silico analyses and comparisons of immunogenicity of Cas9 and AAV orthologs. Linear regressions exclude pairs with no overlap. **(a)** Cas9 MHC class I peptide overlap vs. phylogenetic distance. **(b)** AAV MHC class I peptide overlap vs. phylogenetic distance. **(c)** Cas9 MHC class II peptide overlap vs. phylogenetic distance. **(d)** AAV MHC class II peptide overlap vs. phylogenetic distance.



Supplementary Figure 3: Major AAV serotype groups. (a) AAV immune orthogonal cliques over 81 HLA alleles. AAV5 is the most immune-divergent in comparison to the other serotypes. No orthogonal cliques exist. (b) AAV phylogeny showing major serotype groupings as well as the position of the reconstructed sequence Anc80L65.



Supplementary Figure 4: Experimental validation of a MHCII peptide predictions via IFN- γ ELISPOT. Mice were injected retro-orbitally with 10^{12} vg/mouse of AAV8-SaCas9 targeting the PCSK9 gene and were sacrificed after 6 weeks (n=6). Purified CD4⁺ T cells from splenocytes were seeded at 2×10^5 cells per well in triplicate. 1×10^5 lipopolysaccharide-activated antigen presenting cells (APCs) from control mice were added to each well. Cells were incubated with highly immunogenic MHC-II predicted peptides for 20h. Spots were developed with biotinylated anti-IFN- γ . A one-way ANOVA with post hoc Dunnett's test was performed to determine statistical differences with DMSO for all peptides (left panel), and the significant peptides were plotted (right panel).

REFERENCES

1. Mingozzi, F. & High, K. A. Immune responses to AAV vectors: overcoming barriers to successful gene therapy. *Blood* **122**, 23–36 (2013).
2. Zaldumbide, A. & Hoebe, R. C. How not to be seen: Immune-evasion strategies in gene therapy. *Gene Therapy* **15**, 239–246 (2008).
3. Yang, Y., Li, Q., Ertl, H. C. & Wilson, J. M. Cellular and humoral immune responses to viral antigens create barriers to lung-directed gene therapy with recombinant adenoviruses. *J. Virol.* **69**, 2004–15 (1995).
4. Jawa, V. *et al.* T-cell dependent immunogenicity of protein therapeutics: Preclinical assessment and mitigation. *Clinical Immunology* **149**, 534–555 (2013).
5. Mays, L. E. & Wilson, J. M. The Complex and Evolving Story of T cell Activation to AAV Vector-encoded Transgene Products. *Mol. Ther.* **19**, 16–27 (2011).
6. Basner-Tschakarjan, E., Bijjiga, E. & Martino, A. T. Pre-clinical assessment of immune responses to adeno-associated virus (AAV) vectors. *Frontiers in Immunology* **5**, (2014).
7. Ertl, H. C. J. & High, K. A. Impact of AAV Capsid-Specific T-Cell Responses on Design and Outcome of Clinical Gene Transfer Trials with Recombinant Adeno-Associated Viral Vectors: An Evolving Controversy. *Hum. Gene Ther.* **28**, 328–337 (2017).
8. Kotterman, M. A., Chalberg, T. W. & Schaffer, D. V. Viral Vectors for Gene Therapy: Translational and Clinical Outlook. *Annu. Rev. Biomed. Eng.* **17**, 63–89 (2015).
9. Mingozzi, F. & High, K. A. Therapeutic in vivo gene transfer for genetic disease using AAV: progress and challenges. *Nat. Rev. Genet.* **12**, 341–355 (2011).
10. Manno, C. S. *et al.* Successful transduction of liver in hemophilia by AAV-Factor IX and limitations imposed by the host immune response. *Nat. Med.* **12**, 342–347 (2006).
11. Sathish, J. G. *et al.* Challenges and approaches for the development of safer immunomodulatory biologics. *Nat Rev Drug Discov* **12**, 306–324 (2013).
12. Harding, F. A., Stickler, M. M., Razo, J. & DuBridge, R. B. The immunogenicity of humanized and fully human antibodies: Residual immunogenicity resides in the CDR regions. *MAbs* **2**, 256–265 (2010).
13. De Groot, a S., Knopp, P. M. & Martin, W. De-immunization of therapeutic proteins by T-cell epitope modification. *Dev. Biol. (Basel)*. **122**, 171–194 (2005).
14. Tangri, S. *et al.* Rationally Engineered Therapeutic Proteins with Reduced Immunogenicity. *J. Immunol.* **174**, 3187 LP-3196 (2005).
15. Salvat, R. S., Choi, Y., Bishop, A., Bailey-Kellogg, C. & Griswold, K. E. Protein deimmunization via structure-based design enables efficient epitope deletion at high

- mutational loads. *Biotechnol. Bioeng.* **112**, 1306–1318 (2015).
16. Armstrong, J. K. *et al.* Antibody against poly(ethylene glycol) adversely affects PEG-asparaginase therapy in acute lymphoblastic leukemia patients. *Cancer* **110**, 103–111 (2007).
17. Ganson, N. J., Kelly, S. J., Scarlett, E., Sundy, J. S. & Hershfield, M. S. Control of hyperuricemia in subjects with refractory gout, and induction of antibody against poly(ethylene glycol) (PEG), in a phase I trial of subcutaneous PEGylated urate oxidase. *Arthritis Res. Ther.* **8**, R12–R12 (2006).
18. Veronese, F. M. & Mero, A. The impact of PEGylation on biological therapies. *BioDrugs* **22**, 315–329 (2008).
19. Jevševar, S., Kunstelj, M. & Porekar, V. G. PEGylation of therapeutic proteins. *Biotechnol. J.* **5**, 113–128 (2010).
20. Jacobs, F., Gordts, S. C., Muthuramu, I. & De Geest, B. The liver as a target organ for gene therapy: state of the art, challenges, and future perspectives. *Pharmaceuticals (Basel)*. **5**, 1372–92 (2012).
21. Kok, C. Y. *et al.* Adeno-associated Virus-mediated Rescue of Neonatal Lethality in Argininosuccinate Synthetase-deficient Mice. *Mol. Ther.* **21**, 1823–1831 (2013).
22. Courtenay-Luck, N. S., Epenetos, A. A. & Moore, R. Development of primary and secondary immune responses to mouse monoclonal antibodies used in the diagnosis and therapy of malignant neoplasms. *Cancer Res.* **46**, 6489–6493 (1986).
23. Jinek, M. *et al.* A Programmable Dual-RNA – Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337**, 816–822 (2012).
24. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823–6 (2013).
25. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci.* **109**, E2579–E2586 (2012).
26. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–23 (2013).
27. Ran, F. A. *et al.* In vivo genome editing using Staphylococcus aureus Cas9. *Nature* **520**, 186–190 (2015).
28. Jinek, M. *et al.* RNA-programmed genome editing in human cells. *Elife* **2013**, (2013).
29. Mali, P., Esvelt, K. M. & Church, G. M. Cas9 as a versatile tool for engineering biology. *Nat. Methods* **10**, 957–963 (2013).

30. Hsu, P. D., Lander, E. S. & Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157**, 1262–1278 (2014).
31. Kelton, W. J., Pesch, T., Matile, S. & Reddy, S. T. Surveying the Delivery Methods of CRISPR/Cas9 for ex vivo Mammalian Cell Engineering. *Chim. Int. J. Chem.* **70**, 439–442 (2016).
32. Cho, S. W., Kim, S., Kim, J.-S. J. M. & Kim, J.-S. J. M. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat. Biotechnol.* **31**, 230–2 (2013).
33. Moreno, A. M. & Mali, P. Therapeutic genome engineering via CRISPR-Cas systems. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine* **9**, (2017).
34. Koonin, E. V., Makarova, K. S. & Zhang, F. Diversity, classification and evolution of CRISPR-Cas systems. *Current Opinion in Microbiology* **37**, 67–78 (2017).
35. Makarova, K. S. *et al.* An updated evolutionary classification of CRISPR–Cas systems. *Nat. Rev. Microbiol.* **13**, 722–736 (2015).
36. Chylinski, K., Makarova, K. S., Charpentier, E. & Koonin, E. V. Classification and evolution of type II CRISPR-Cas systems. *Nucleic Acids Research* **42**, 6091–6105 (2014).
37. Shmakov, S. *et al.* Diversity and evolution of class 2 CRISPR–Cas systems. *Nat. Rev. Microbiol.* **15**, 169–182 (2017).
38. Charlesworth, C. T. *et al.* Identification of Pre-Existing Adaptive Immunity to Cas9 Proteins in Humans. doi:10.1101/243345
39. Wagner, J. a *et al.* Safety and biological efficacy of an adeno-associated virus vector-cystic fibrosis transmembrane regulator (AAV-CFTR) in the cystic fibrosis maxillary sinus. *Laryngoscope* **109**, 266–74 (1999).
40. Song, S. *et al.* Sustained secretion of human alpha-1-antitrypsin from murine muscle transduced with adeno-associated virus vectors. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 14384–8 (1998).
41. Chirmule, N. *et al.* Humoral Immunity to Adeno-Associated Virus Type 2 Vectors following Administration to Murine and Nonhuman Primate Muscle. *J. Virol.* **74**, 2420–2425 (2000).
42. Fields, P. a *et al.* Risk and prevention of anti-factor IX formation in AAV-mediated gene transfer in the context of a large deletion of F9. *Mol. Ther.* **4**, 201–210 (2001).
43. Herzog, R. W. *et al.* Influence of vector dose on factor IX-specific T and B cell responses in muscle-directed gene therapy. *Hum. Gene Ther.* **13**, 1281–91 (2002).
44. Lozier, J. N., Tayebi, N. & Zhang, P. Mapping of genes that control the antibody response

- to human factor IX in mice. *Blood* **105**, 1029–1035 (2005).
45. Zhang, H. G. *et al.* Genetic analysis of the antibody response to AAV2 and factor IX. *Mol. Ther.* **11**, 866–874 (2005).
46. Tam, H. H. *et al.* Sustained antigen availability during germinal center initiation enhances antibody responses to vaccination. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E6639–E6648 (2016).
47. Benveniste, O. *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.* **21**, 704–712 (2010).
48. Gao, G.-P. *et al.* Novel adeno-associated viruses from rhesus monkeys as vectors for human gene therapy. *Proc. Natl. Acad. Sci.* **99**, 11854–11859 (2002).
49. Jooss, K., Yang, Y., Fisher, K. J. & Wilson, J. M. Transduction of Dendritic Cells by DNA Viral Vectors Directs the Immune Response to Transgene Products in Muscle Fibers. *J. Virol.* **72**, 4212–4223 (1998).
50. Gernoux, G. *et al.* Early Interaction of Adeno-Associated Virus Serotype 8 Vector with the Host Immune System Following Intramuscular Delivery Results in Weak but Detectable Lymphocyte and Dendritic Cell Transduction. *Hum. Gene Ther.* **26**, 1–13 (2015).
51. Zhu, J., Huang, X. & Yang, Y. The TLR9-MyD88 pathway is critical for adaptive immune responses to adeno-associated virus gene therapy vectors in mice. *J. Clin. Invest.* **119**, 2388–2398 (2009).
52. Gernoux, G., Wilson, J. M. & Mueller, C. Regulatory and Exhausted T Cell Responses to AAV Capsid. *Hum. Gene Ther.* **28**, 338–349 (2017).
53. Kurosaki, T., Kometani, K. & Ise, W. Memory B cells. *Nat. Rev. Immunol.* **15**, 149–159 (2015).
54. Zabel, F. *et al.* Distinct T helper cell dependence of memory B-cell proliferation versus plasma cell differentiation. *Immunology* **150**, 329–342 (2017).
55. Ding, Q. *et al.* Permanent Alteration of PCSK9 With In Vivo CRISPR-Cas9 Genome Editing. *Circ. Res.* **115**, 488–492 (2014).
56. Zinn, E. *et al.* In Silico Reconstruction of the Viral Evolutionary Lineage Yields a Potent Gene Therapy Vector. *Cell Rep.* **12**, 1056–1068 (2017).
57. Calcedo, R. & Wilson, J. M. AAV Natural Infection Induces Broad Cross-Neutralizing Antibody Responses to Multiple AAV Serotypes in Chimpanzees. *Hum. Gene Ther. Clin. Dev.* **27**, 79–82 (2016).
58. Harbison, C. E. *et al.* Examining the cross-reactivity and neutralization mechanisms of a

- panel of mabs against adeno-associated virus serotypes 1 and 5. *J. Gen. Virol.* **93**, (2012).
59. Majowicz, A. *et al.* Successful Repeated Hepatic Gene Delivery in Mice and Non-human Primates Achieved by Sequential Administration of AAV5^{ch} and AAV1. *Mol. Ther.* **25**, 1831–1842 (2017).
 60. McIntosh, J. H. *et al.* Successful attenuation of humoral immunity to viral capsid and transgenic protein following AAV-mediated gene transfer with a non-depleting CD4 antibody and cyclosporine. *Gene Ther* **19**, 78–85 (2012).
 61. Mingozi, F. *et al.* Prevalence and pharmacological modulation of humoral immunity to AAV vectors in gene transfer to synovial tissue. *Gene Ther* **20**, 417–424 (2013).
 62. Mingozi, F. *et al.* Pharmacological Modulation of Humoral Immunity in a Nonhuman Primate Model of AAV Gene Transfer for Hemophilia B. *Mol. Ther.* **20**, 1410–1416 (2017).
 63. Unzu, C. *et al.* Transient and intensive pharmacological immunosuppression fails to improve AAV-based liver gene transfer in non-human primates. *J. Transl. Med.* **10**, 122 (2012).
 64. Riechmann, L., Clark, M., Waldmann, H. & Winter, G. Reshaping human antibodies for therapy. *Nature* **332**, 323–7 (1988).
 65. Vollmers, C., Sit, R. V., Weinstein, J. A., Dekker, C. L. & Quake, S. R. Genetic measurement of memory B-cell recall using antibody repertoire sequencing. *Proc. Natl. Acad. Sci.* **110**, 13463–13468 (2013).
 66. Ruppert, J. *et al.* Prominent role of secondary anchor residues in peptide binding to HLA-A2.1 molecules. *Cell* **74**, 929–937 (2017).
 67. Zhang, S.-Q. *et al.* Direct measurement of T cell receptor affinity and sequence from naïve antiviral T cells. *Sci. Transl. Med.* **8**, 341ra77 (2016).
 68. Baker, M. P., Reynolds, H. M., Lumicisi, B. & Bryson, C. J. Immunogenicity of protein therapeutics: The key causes, consequences and challenges. *Self Nonself* **1**, 314–322 (2010).
 69. EL-Manzalawy, Y., Dobbs, D. & Honavar, V. Predicting linear B-cell epitopes using string kernels. *J. Mol. Recognit.* **21**, 243–255 (2008).
 70. Larsen, J. E. P., Lund, O. & Nielsen, M. Improved method for predicting linear B-cell epitopes. *Immunome Res.* **2**, 2 (2006).
 71. Sollner, J. *et al.* Analysis and prediction of protective continuous B-cell epitopes on pathogen proteins. *Immunome Res.* **4**, 1 (2008).

72. Dalkas, G. A. & Roodman, M. SEPIa, a knowledge-driven algorithm for predicting conformational B-cell epitopes from the amino acid sequence. *BMC Bioinformatics* **18**, 95 (2017).
73. Sun, P. *et al.* Bioinformatics resources and tools for conformational B-cell epitope prediction. *Computational and Mathematical Methods in Medicine* **2013**, (2013).
74. Liepe, J. *et al.* A large fraction of HLA class I ligands are proteasome-generated spliced peptides. *Science* (80-.). **354**, (2016).
75. Fonfara, I. *et al.* Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems. *Nucleic Acids Res.* **42**, 2577–2590 (2014).
76. Andreatta, M. & Nielsen, M. Gapped sequence alignment using artificial neural networks: application to the MHC class I system. *Bioinformatics* **32**, 511–517 (2015).
77. Andreatta, M. *et al.* Accurate pan-specific prediction of peptide-MHC class II binding affinity with improved binding core identification. *Immunogenetics* **67**, 641–650 (2015).
78. Vita, R. *et al.* The immune epitope database (IEDB) 3.0. *Nucleic Acids Res.* **43**, D405-12 (2015).
79. Grieger, J. C., Choi, V. W. & Samulski, R. J. Production and characterization of adeno-associated viral vectors. *Nat. Protoc.* **1**, 1412–1428 (2006).