Drawing a high-resolution functional map of adeno-associated virus capsid by massively parallel sequencing

Kei Adachi1,†, Tatsuji Enoki2,†, Yasuhiro Kawano1,2, Michael Veraz1 & Hiroyuki Nakai1,†

Adeno-associated virus (AAV) capsid engineering is an emerging approach to advance gene therapy. However, a systematic analysis on how each capsid amino acid contributes to multiple functions remains challenging. Here we show proof-of-principle and successful application of a novel approach, termed AAV Barcode-Seq, that allows us to characterize phenotypes of hundreds of different AAV strains in a high-throughput manner and therefore overcomes technical difficulties in the systematic analysis. In this approach, we generate DNA barcode-tagged AAV libraries and determine a spectrum of phenotypes of each AAV strain by Illumina barcode sequencing. By applying this method to AAV capsid mutant libraries tagged with DNA barcodes, we can draw a high-resolution map of AAV capsid amino acids important for the structural integrity and functions including receptor binding, tropism, neutralization and blood clearance. Thus, Barcode-Seq provides a new tool to generate a valuable resource for virus and gene therapy research.
A deno-associated virus (AAV) is an attractive gene delivery vector for human gene therapy. However, various issues remain to be overcome, including the requirement of high vector dose for clinically beneficial outcomes\textsuperscript{1,2}, efficacy-limiting host immune responses against viral proteins\textsuperscript{3-5}, promiscuous viral tropism and the high prevalence of pre-existing anti-AAV neutralizing antibodies in humans\textsuperscript{6,7}. Despite these issues, enthusiasm and hope for the use of AAV vectors in gene therapy is growing. This is owed in part to the amenability of the capsids to genetic manipulation for the creation of novel targeting vectors and vectors with a stealth phenotype\textsuperscript{8}. A series of site-directed mutagenesis studies\textsuperscript{9-18} and the elucidation of the atomic structures of the prototype AAV serotype 2 (AAV2)\textsuperscript{19} and other serotypes\textsuperscript{20-23} have provided insights into the structural basis of the AAV capsid functions. Such conventional approaches have helped identify amino-acid residues that play roles in binding to cell surface receptors\textsuperscript{8,24-26} and neutralizing antibodies\textsuperscript{27,28}. They also assisted in designing more potent AAV capsids by surface-exposed tyrosine residue mutations\textsuperscript{29-32} and those with selective tropism by re-engineering a cell surface receptor footprint\textsuperscript{33}. However, structural knowledge-based prediction of viral capsid functions remains a significant challenge. Directed evolution approaches have recently become increasingly common in the development of novel AAV capsids that target specific cell types with an enhanced efficiency\textsuperscript{34-37}. This approach does not require prior knowledge but relies on the power of iterative positive selection from a pool of diverse mutants. Mutants selected by this method, however, often suffer from a lack of structural and functional interpretation of phenotypes, which limits the use of evolved amino-acid sequence information for the development of new logical approaches.

Here, we develop a novel next-generation sequencing (NGS)-based method, termed AAV Barcode-Seq. This new method utilizes a DNA barcode-tagged mutagenesis approach in conjunction with multiplexed Illumina sequencing. We report proof-of-principle and successful application of this method to comprehensively identify the structural and functional roles of a total of 381 amino acids in the entire carboxy (C)-terminal half of the AAV9 capsid and a total of ~70 amino acids within the largest loop in the AAV1, AAV6, AAV7, AAV8 and AAV9 VP capsid proteins. In addition, we present two successful cases that utilized the new knowledge obtained from AAV Barcode-Seq data to design and create new AAV capsids with directed phenotypes. Barcode-Seq is a new approach that significantly advances virus and gene therapy research.

Results

Experimental design. The aim of this study was to establish AAV Barcode-Seq as a novel method that would allow us to characterize the biological phenotypes of many AAV strains in an unprecedented high-throughput manner, and demonstrate its successful application to AAV research. To accomplish this aim, we generated the following seven DNA-barcoded AAV serotype and capsid mutant virus libraries (Table 1 and Supplementary Table 1). Each serotype or mutant viral clone in the libraries carried the wild-type AAV2 rep gene, an AAV cap gene derived from various serotypes or mutants, and a pair (pr) of left (lt) and right (rt) viral clone-specific 12-nucleotide long DNA barcodes (Virus Bar Code or VBC) (Fig. 1a). AAV-Serotype-VBCLib contained nine AAV serotypes plus two AAV chimeric mutants; AAV9-AA-VBCLib’s covered a total of 191 double alanine (AA) scanning mutants that spanned the entire region of the C-terminal half of the AAV9 VP1 capsid protein (Fig. 1b,d), and AAV2R585E-HP-VBCLib’s contained 125 AAV2R585E-derived hexapeptide (HP) scanning mutants with AAV2-derived HPs being replaced with those derived from AAV1, 6, 7, 8 and 9 capsids (Fig. 1c,d). These 125 HP mutants covered the entire region from amino-acid positions 441–484 and from 571–604. Each DNA-barcoded AAV capsid mutant library contained 15–18 clones each of AAV9 and AAV2R585E as internal reference controls in addition to 2 or 3 clones per mutant (Table 1). AAV9 AA scanning primarily focused on loss-of-function phenotypes, while heparin-binding-deficient AAV2R585E HP scanning primarily focused on gain-of-function phenotypes. To construct a large number of mutants, we established a method for high-throughput site-directed capsid mutagenesis (Fig. 1e). With these libraries, we followed the AAV Barcode-Seq experimental procedure depicted in Figure 2 to determine Phenotypic Difference (PD) values of each serotype or mutant. PD indicates ‘fold change’ of a phenotype compared with that of the reference controls.

Validation of AAV Barcode-Seq. We validated the Barcode-Seq approach through analysis of errors. Over a dynamic range of 10\textsuperscript{5}, the relative abundance of DNA templates and Illumina sequencing read numbers in a sample were positively correlated with Pearson’s correlation coefficients of 0.93 (lt-VBCs) and 0.97 (rt-VBCs) on average (Supplementary Fig. 1a). An undersampling simulation study\textsuperscript{38}, using the actual data sets we obtained, showed that when the average Illumina sequence read number per clone is ≥64, the average coefficients of variation of the data obtained by technically replicated experiments are 0.22–0.24 and 0.31–0.34 for lt- and rt-VBCs, respectively (Supplementary Fig. 1b), and Pearson’s correlation coefficients between undersampled and full-size data sets are ≥0.96 (Supplementary Fig. 1c). We determined statistical power of the analysis by a simulated experiment that used a library containing two clones per mutant and three or more reference control clones. This simulation study showed that the statistical power to detect a twofold change with P<0.05 (two-tailed Mann–Whitney U-test) is 0.78–0.87 in a duplicated experiment and 0.87–0.96 in a triplicated experiment when the average reference control read number per clone is ≥64 (Supplementary Fig. 2). The power to detect ≥4-fold changes with P<0.05 (two-tailed Mann–Whitney U-test) was found to be ~1.0 in both duplicated and triplicated experiments (Supplementary Fig. 2).

Proof-of-principle of AAV Barcode-Seq. Here we present four examples demonstrating that AAV Barcode-Seq can reproduce data obtained by conventional approaches. First, using the tissues collected from AAV-Serotype-VBCLib-injected mice, we determined transduction efficiencies of the major AAV serotypes in various tissues by AAV Barcode-Seq (Supplementary Fig. 3). The results reproduced the phenotypes of each AAV serotype we have already known from prior studies. For example, AAV8 and 9 transduced many tissues with high efficiency\textsuperscript{39-43}, while AAV3 was the least efficient serotype in mice\textsuperscript{44}. AAV2R585E significantly lost the ability to transduce the liver\textsuperscript{46}. Second, we injected 1 × 10\textsuperscript{13} vector genomes (vg) per kg of AAV-Serotype-VBCLib (n = 3) via the tail vein as a bolus and determined the blood clearance rate of each serotype. The results again reproduced the pharmacokinetic profiles obtained in our previous study, showing that AAV9 exhibited distinctively delayed blood clearance, while AAV2 was rapidly cleared in the early phase after injection\textsuperscript{44} (Supplementary Fig. 4). Third, we applied AAV-Serotype-VBCLib to Chinese hamster ovary (CHO) Pro5 and Lec2 cells to determine the ability of each serotype to bind different sugar residues in glycan chains. Pro5 cells express terminal sialic acids but do not express terminal galactose, while Lec2 cells express terminal galactose but do not express terminal sialic acids in glycan chains\textsuperscript{45}. Using these cells, we assessed
efficiencies in cell surface binding and transduction by AAV Barcode-Seq. The obtained results were consistent with the established fact that terminal N-linked sialic acids are the primary receptor for AAV1, AAV5 and AAV6 (refs 46,47) and terminal galactose is the receptor for AAV9 (refs 24,25) (Supplementary Fig. 5). Fourth, by injecting AAV-Serotype-VBCLib into mice pre-immunized with AAV9, we could demonstrate that anti-AAV9 neutralizing antibody crossreacts with AAV8 and

| Library name            | AAV clones contained in the libraries | Amino acids to be investigated | Total no. of clones |
|-------------------------|---------------------------------------|-------------------------------|---------------------|
|                         | AAV strains | No. of mutants | No. of clones |
| AAV-Serotype-VBCLib     | AAV9       | Control       | 100            |
|                         | AAV2R585E  | Control       | 10             |
|                         | AAV1       | Wild type     | 2              |
|                         | AAV2       | Wild type     | 2              |
|                         | AAV3       | Wild type     | 2              |
|                         | AAV4*      | Wild type     | 2              |
|                         | AAV5       | Wild type     | 2              |
|                         | AAV6       | Wild type     | 2              |
|                         | AAV7       | Wild type     | 2              |
|                         | AAV8       | Wild type     | 2              |
|                         | AAVrh10    | Wild type     | 2              |
|                         | AAV1.9-1†  | Hybrid        | 2              |
|                         | AAV1.9-3-5†| Hybrid        | 2              |
| AAV9-AA-VBCLib-1        | AAV9       | Control       | 15             |
|                         | AAV2R585E  | Control       | 15             |
|                         | AAV9       | Wild type     | 2              |
|                         | AAV2R585E  | Wild type     | 2              |
|                         | AAV9       | Wild type     | 2              |
|                         | AAV8       | Wild type     | 2              |
|                         | AAVrh10    | Wild type     | 2              |
|                         | AAV1.9-1†  | Hybrid        | 2              |
|                         | AAV1.9-3-5†| Hybrid        | 2              |
| AAV9-AA-VBCLib-2        | AAV9       | Control       | 15             |
|                         | AAV2R585E  | Control       | 15             |
|                         | AAV9       | Wild type     | 2              |
|                         | AAV2R585E  | Wild type     | 2              |
|                         | AAV9       | Wild type     | 2              |
|                         | AAV8       | Wild type     | 2              |
|                         | AAVrh10    | Wild type     | 2              |
|                         | AAV1.9-1†  | Hybrid        | 2              |
|                         | AAV1.9-3-5†| Hybrid        | 2              |
| AAV9-AA-VBCLib-3        | AAV9       | Control       | 15             |
|                         | AAV2R585E  | Control       | 15             |
|                         | AAV9       | Wild type     | 2              |
|                         | AAV2R585E  | Wild type     | 2              |
|                         | AAV9       | Wild type     | 2              |
|                         | AAV8       | Wild type     | 2              |
|                         | AAVrh10    | Wild type     | 2              |
|                         | AAV1.9-1†  | Hybrid        | 2              |
|                         | AAV1.9-3-5†| Hybrid        | 2              |
| AAV9-AA-VBCLib-4        | AAV9       | Control       | 15             |
|                         | AAV2R585E  | Control       | 15             |
|                         | AAV9       | Wild type     | 2              |
|                         | AAV2R585E  | Wild type     | 2              |
|                         | AAV9       | Wild type     | 2              |
|                         | AAV8       | Wild type     | 2              |
|                         | AAVrh10    | Wild type     | 2              |
|                         | AAV1.9-1†  | Hybrid        | 2              |
|                         | AAV1.9-3-5†| Hybrid        | 2              |
| AAV2R585E-HP-VBCLib-1   | AAV9       | Control       | 15             |
|                         | AAV2R585E  | Control       | 15             |
|                         | AAV2R585E  | Hexapeptide   | 66             |
|                         | AAV2R585E  | Hexapeptide   | 132            |
|                         | AAV2R585E  | Hexapeptide   | 441–484†       |
| AAV2R585E-HP-VBCLib-2   | AAV9       | Control       | 15             |
|                         | AAV2R585E  | Control       | 15             |
|                         | AAV2R585E  | Hexapeptide   | 177            |
|                         | AAV2R585E  | Hexapeptide   | 571–604†       |
| AAV2R585E-HP-VBCLib-3   | AAV9       | Control       | 15             |
|                         | AAV2R585E  | Control       | 15             |
|                         | AAV2R585E  | Hexapeptide   | 177            |
|                         | AAV2R585E  | Hexapeptide   | 571–604†       |

*We lost AAV4 viral particles significantly from the library during the viral stock purification procedure. Accordingly, Illumina sequence read numbers for AAV4 in the viral stock were several logs lower than others. Therefore, AAV4 data were not evaluated or included in the study.

†AAV1.9-1 and AAV1.9-3-5 are AAV1 and AAV9 hybrid capsids44.

‡The following 11 AAV2R585E hexapeptide mutants were included: 463-16000, 465-00700, 465-16000, 467-16000, 479-00009, 573-00780, 581-10000, 583-06000, 585-00080, 593-10000 and 595-10000.

§The amino-acid positions are provided based on AAV2 VP1.
AAVrh10 but does not neutralize AAV1, 2, 3, 5, 6 and 7 at an appreciable level, which is in keeping with the observation reported in the literature. These observations establish proof-of-principle of AAV Barcode-Seq.

AAV9 capsid amino acids important for virion formation. We then applied AAV Barcode-Seq to AAV mutants. First, we produced 382 viral clones representing 191 AAV9 AA mutants and 15 each of AAV9 and AAV2R585E reference control clones in separate culture dishes by DNA transfection, pooled an equal amount of crude cell lysate obtained from each culture dish and sequenced them on the Illumina platform. We then converted raw sequence read number data to PD values by a computational algorithm.

Figure 1 | DNA-barcoded AAV libraries. (a) Viral genome maps of AAV-Serotype-VBCLib, AAV9-AA-VBCLib, and AAV2R585E-HP-VBCLib. Each viral genome contains a pair of 12 nucleotide-long DNA barcodes (lt-VBC and rt-VBC) downstream of the AAV2 polyadenylation signal (pA). PBS1–4, PCR primer-binding sites. Nhe I and Bsr GI indicate the restriction enzyme recognition sites used for DNA barcode cloning. (b) Double alanine (AA) mutagenesis of the C-terminal half of the AAV9 VP1 capsid protein from amino-acid positions 356 to 736. AAxxx’s are the names of AA mutants. (c) HP scanning mutagenesis of the AAV2R585E capsid. We moved HPs derived from AAV1, 6, 7, 8 and 9 on the AAV2R585E capsid at a two amino-acid interval within the regions of interest. (d) Structure of the AAV capsid VP protein indicating the region we investigated in this study. The blue region was the wild type (Fig. 3a). As expected, a strong correlation was made three AAV9-AA-VBCLib libraries as described in Table 1. One of each of AAV9 and AAV2R585E reference control clones in separate culture dishes by DNA transfection, pooled an equal amount of crude cell lysate obtained from each culture dish and sequenced them on the Illumina platform. We then converted raw sequence read number data to PD values by a computational algorithm.
found between the tolerability to amino-acid substitution, the degree of evolutionary conservation and topological locations. Forty-four of the 72 mutants that could not tolerate an AA mutation had changes of amino acids buried inside the virion shell, while 28 of the 119 capsid-forming mutants had such changes ($\chi^2$ test; $P = 0.0002$). Importantly, this analysis could identify amino acids that potentially have functional roles. D384, G385, I560, T561, N562, E563, E564, E565, N704 and Y705 are amino acids that are conserved and exposed on the depressed surface at the twofold symmetry axis of the capsid but tolerated mutations, indicating their functional role. Subsequent studies revealed that the mutations involving these amino acids resulted in a phenotype exhibiting impaired transduction.

**Capsid amino acids responsible for liver transduction.** Second, we injected C57BL/6 mice intravenously with AAV9-9A-VBCLib or AAV2R585E-HP-VBCLib at a dose of $1 \times 10^{12}$ vg per mouse ($n = 3$ per library). We harvested 12 major tissues 6 weeks post injection and subjected them to the AAV Barcode-Seq analysis, which revealed that 31 AAV9 AA mutants exhibited >10-fold reduction in liver transduction (Fig. 3b). A hierarchical clustering analysis (complete linkage method) on the *in vivo* transduction profiles in all the 12 tissues revealed that they can be grouped into two phenotypically distinct groups, that is, 9 mutants that mainly detergate the liver (Liver-Detargeting (LD) mutants) and 22 mutants showing impaired transduction not only in the liver but also in many nonhepatic tissues (Globally Detargeting (GD) mutants) (Supplementary Fig. 7). To validate these observations, we produced AAV-CMV-lacZ vector encapsidated with the mutants (Supplementary Table 2 and 3). In particular, P504A/G505A (LD mutant), N562A/E563A (GD mutant) and Q590A (LD mutant). We injected them into C57BL/6 wild type or Rag1 $–/–$ mice intravenously at $3 \times 10^{11}$ or $1 \times 10^{12}$ vg per mouse. The results obtained by X-Gal staining, qPCR and Southern blot analyses on the tissues harvested 11 days (wild-type mice) and 6 weeks (Rag1 $–/–$ mice) post injection corroborated the AAV Barcode-Seq results (Fig. 4, Supplementary Tables 2 and 3). In particular, P504A/G505A mutant transduced the heart, kidney and brain efficiently at 92%, 318% and 60% of the wild type AAV9’s levels, respectively, with a >200-fold decrease in the liver transduction determined by vector genome copy numbers (Supplementary Table 2). In the HP scanning experiment, replacement of AAPV2R585E capsid amino-acid residues 461–474 with any of the HPs derived from AAV1, 6, 7, 8 and 9 resulted in enhanced liver transduction by 3–68-fold (Fig. 3f) in all but one of the 17 virion-forming AAPV2R585E-derived mutants. We validated this enhancement by injecting mice with the AAV-CMV-lacZ vector packaged with the AAPV2R585E mutant 463–16000 (Fig. 4, Supplementary Tables 2 and 3). We also observed a more than sixfold increase in liver transduction when the AAPV2R585E 581-586 HP was replaced with any of those derived from the other five serotypes (Fig. 3f). These observations clearly delineate the AAPV1, 6, 7, 8 and 9 capsid-amino-acid residues important for hepatic transduction.

**AAV9 capsid amino acids responsible for galactose binding.** Third, we applied AAPV9-9A-VBCLib to CHO Pro5 and Lec2 cells and investigated the ability to bind and transduce these two cell lines by AAP barcode-Seq. We found that 14 AA mutants covering 26 residues exhibited $>80\%$ decrease in the binding to Lec2 cells (Fig. 3c), while they retained >50% ability to bind to Pro5 cells (Fig. 3d). This strongly indicates that these 26 amino acids are directly or indirectly responsible for galactose binding. Supporting this notion, 3 of the 26 residues have recently been found to constitute the galactose binding domain$^9$. Topologically, they cluster in the pocket between the side walls of the threefold protrusions (Supplementary Fig. 8). All the HP scanning AAPV2R585E mutants exhibited Lec2 cell binding >466-fold less than the wild-type AAV9. We also identified seven mutants showing a significantly reduced transduction-to-binding ratio compared with the wild type (Fig. 3e). The amino acids substituted in these mutants are therefore important for postattachment viral processing.

**Capsid amino acids responsible for persistence in the blood.** Fourth, we applied AAPV Barcode-Seq to a pharmacokinetic study of AAV mutants injected intravenously into mice. We infused $1 \times 10^{12}$ vg per kg of each of AAV mutant libraries into mice and determined relative blood concentrations by AAP Barcode-Seq over a period of 72 h ($n = 2$ per library) (Supplementary Fig. 9). At 72 h post injection, 20 AAV9 AA mutants lost persistence in the blood compared with the wild-type AAV9, showing <10% levels of the blood concentration of AAV9 (Loss of Persistence (LP) phenotype) and 13 AAPV2R585E HP mutants showed 2–13-fold higher blood concentrations than AAPV2R585E at 72 h post injection (Delayed Clearance (DC) phenotype). These amino-acid residues responsible for persistent circulation in the bloodstream are surface-exposed and primarily overlap with those responsible for the altered *in vivo* transduction phenotypes. Eighteen of the 20 (90%) AAV9 mutants showing an LP pharmacokinetic phenotype were LD or GD mutants (Fisher’s exact test; $P = 3 \times 10^{-4}$; Supplementary Table 4), and 12 of the 13 (92%) AAPV2R585E mutants with a DC pharmacokinetic phenotype exhibited 7–62-fold enhancement of liver transduction (Fisher’s exact test; $P = 5 \times 10^{-10}$; Supplementary Table 5). These observations indicate that a majority of the surface-exposed amino acids that play a role in viral clearance have a dual functional role in both pharmacokinetics and *in vivo* transduction, although the amino acids important for *in vivo* transduction are not necessarily involved in the clearance (Supplementary Tables 4 and 5). One mutant, AAPV9’484A/R485A, was found to bind to various types of cells with a dramatically increased efficiency (a prominent bar in Fig. 3d) and was cleared at a strikingly accelerated rate (Supplementary Fig. 9a). The rapid blood clearance of this particular mutant could be interpreted by sequestration of viral particles from the bloodstream due to the increased cell attachment; however, the underlying mechanism for this gain-of-function phenotype remains uncharacterized.

**Anti-AAV1 and AAV9-neutralizing antibody epitope mapping.** Fifth, we applied AAP Barcode-Seq to map anti-AAV1 and AAV9 capsid-neutralizing antibody epitopes. To this end, we immunized C57BL/6 mice by intravenous injection of AAPV1- or AAPV9-CMV-lacZ. Three weeks later, we infused AAPV2R585E-HP-VBCLib into the immunized mice at a dose of $1 \times 10^{13}$ vg per kg ($n = 3$ per library) and determined viral concentrations in the blood over 1 hour by AAP Barcode-Seq. We expected that only AAPV2R585E mutants with an HP containing an antibody epitope would be neutralized, and therefore would be cleared faster than other mutants in the same immunized animal or faster than the same mutant in naïve animals. By taking this approach, we successfully identified 452-QSGSAQ-457 and 453-GSGQ-457 as one of the epitopes for anti-AAV1 and AAV9-neutralizing antibodies developed by viral immunization of mice, respectively (Supplementary Fig. 10). Both of the epitopes reside on the highest peak of the threefold capsid protrusions. Many neutralizing antibody epitopes are conformational, as opposed to linear. In this regard, the AAP HP scanning approach displays peptides in the context of appropriately juxtaposed regions from different parts of the sequence in a more native-like quaternary.
**Figure 3 | Results of the AAV Barcode-Seq analysis.** (a) Mapping of the AAV9 capsid amino acids important for viral particle formation. Grey bars indicate viral particle production yield in HEK293 cells determined by AAV Barcode-Seq. Evolutionary conservation scores determined by the ConSurf analysis are shown with a red line. Highly conserved amino acids show low scores. Topological locations of each amino acid are shown as a horizontal colour bar. The red, blue and green regions indicate outer surface-exposed, inner surface-exposed and buried amino acids, respectively. The far left green and red bars represent the wild-type AAV9 and AAV2R585E, respectively. (b) Liver transduction efficiency of each AAV9 AA mutant. The mutants shown with red bars are liver-detargeting mutants (> 90% reduction of liver transduction efficiency). (c, d) Results of the cell surface binding assay of the AAV9 mutants using Lec2 (c) and Pro5 (d) cells. The mutants shown with red bars in c are the 14 mutants exhibiting a >80% loss of the binding ability compared with the wild type. These 14 mutants were L380A/T381A, L382A/N383A, I440A/D441A, Y446A/L447A, T450A/I451A, V465A, P468A/S469A, N470A/M471A, Q474A/G475A, Y484A/R485A, E500A/F501A, W503A, R514A/N515A and S516/L517A. (e) Transduction efficiency of the AAV9 mutants in Lec2 cells. The mutants shown with red or cyan bars are those exhibiting a >80% loss of transduction efficiency compared with the wild type. The mutants with cyan bars are the seven mutants that showed impaired postattachment viral processing deduced by a >80% reduction of the transduction-to-binding ratios. These seven mutants were D384A/G385A, S386A/Q387A, N496A/N497A, P504A/G505A, N562A/E563A, Q590A and Y706A/K707A. In b–e, the grey regions indicate the AAV9 mutants that exhibited a >95% decrease in the viral particle formation compared with the wild type and therefore are devoid of functional phenotype data; the leftmost green bars represent the wild-type AAV9. (f) Liver transduction efficiency of each AAV2R585E HP mutant. Note that yellow bars are all AAV2R585E because HP replacement in them does not change their amino-acid sequence. In all the panels, error bars represent s.e.m. The number of replicates is detailed in Supplementary Data.
structure of viral capsid and therefore may provide a more ideal platform for the identification of epitopes than linear peptide arrays. As for anti-AAV9-neutralizing antibody epitopes, our study has indicated that there is another epitope(s) in the amino (N)-terminal half of the AAV9 capsid because AAV1.9-1 (an AAV1 and AAV9 hybrid capsid consisted of the N-terminal half of the AAV9 capsid and the C-terminal half of the AAV1 capsid) was neutralized with anti-AAV9 antibody (Supplementary Fig. 6i).

Design of a galactose-binding motif in the AAV2 capsid. Now that we had identified the 26 amino acids responsible for galactose binding, we exploited this knowledge to create AAV2 capsids that bind to galactose with minimal amino-acid changes. AAV2 does not bind to galactose; therefore, successful creation of galactose-binding AAV2 mutants based on the knowledge obtained from the AAV Barcode-Seq data would provide a proof of the practical utility of the AAV Barcode-Seq analysis. To create such mutants, we selected 10 residues based on the AAV Barcode-Seq data and the information about the evolutionary conservation and topological locations of the identified amino acids. In brief, we assumed that the amino acids critical for galactose binding are those that are evolutionarily variable, surface exposed and form a cluster on the surface of the capsid. To assess the evolutionary conservation of the AAV capsid amino acids, we compared the capsid amino-acid sequences between AAV1, 2, 3, 6, 7, 8 and 9, which represent each of all the six AAV clades. A sequence alignment analysis revealed that, among the 26 amino acids, the amino acids except for the following 12 amino acids, I451, V465, P468, S469, N470, M471, G475, Y484, E500, F501, N515 and L517, are completely conserved in the AAV1, 2, 3, 6, 7, 8 and 9 capsids. Among these 12 amino acids, I451, V465, P468, S469, N470, E500, F501 and N515 are surface-exposed. These amino acids cluster in the pocket that has recently been reported as the site where AAV9 binds galactose. AAV9 S469 is shared with AAV2; L517 is well conserved between different serotypes and Y484 is not located in the pocket and well conserved. Therefore, we did not select these three amino acids. As for AAV9 I451, the partner of the double alanine mutation, T450, is completely conserved. For this reason, we opted not to introduce a mutation at AAV2 N449 corresponding to AAV9 I451.

Consequently, we introduced the following 10 amino-acid substitutions to the AAV2 capsid to create a new galactose-binding motif in the AAV2R585E capsid: Q464V/A467P/D469N/I470M/R471A/D472V/S474G/Y500F/S501A/D514N (Fig. 5a). R471A/D472V was included because the AAV9AA472 with a V473A mutation within the vicinity of this amino-acid cluster also showed impairment of Lec2 cell binding (PD = 0.36).

We then replaced some or all of these 10 residues spanning from 464 to 514 in the AAV2R585E capsid with those of AAV9 and produced AAV2R585E.9-1, 9-2, 9-3, 9-4 and 9-5 vectors.
expressing GFP or lacZ driven by the CMV promoter (Fig. 5a). 9-2 contained all the 10 residue substitutions while 9-1 and 9-3 contained only the right arm left of the motif, respectively. 9-4 and 9-5 were among the AAV2R585E HP mutants. We infected Pro5 and Lec2 cells with all the five mutants together with controls and injected mice with 9-2 in the same manner as described above. This experiment revealed that 9-2 has biological properties that are almost the same as those of the wild-type AAV9 both in vitro (Fig. 5b–d) and in vivo (Fig. 4, Supplementary Tables 2 and 3). 9-2 bound and transduced Lec2 cells and exhibited a high liver and heart transduction efficiency. The other four mutants showed some reduction compared with the wild-type AAV9. This experiment also revealed that AAV9 F501, A502 and/or N515 are important not only for galactose binding but also for postattachment viral processing because 9-3 bound to Lec2 cells at a level comparable to that of 9-2 and AAV9 but showed significantly attenuated Lec2 transduction. Shen et al.49 have recently created an AAV2 capsid that binds to terminal galactose based on our AAV Barcode-Seq data. This experiment was performed once. Scale bar, 400 µm.

Design of liver-detargeting mutants. P504A/G505A and Q590A mutations in AAV9 capsid led to a liver-detargeting phenotype that preserved the ability to mediate efficient cardiac transduction in mice. T503/G504 and Q589 in AAV2 are residues corresponding to P504/G505 and Q590 in AAV9. To investigate whether introduction of alanine mutations to these residues in the AAV2R585E.9-2 capsid also attenuates the ability to mediate liver transduction, we created three AAV2R585E.9-2 capsid mutants carrying T503A/G504A (mtTG), Q589A (mtQ) or T503A/G504A/Q589A (mtTGQ) (Fig. 5a) and packaged the lacZ- or GFP-expressing AAV viral genome into them. We tested these AAV mutant vectors in vitro and in vivo in the same manner as described above. We did not observe any phenotypic changes in vitro (Fig. 5c). However, in mice, we found that mtTG substantially attenuated liver transduction and mtTGQ nearly completely abolished liver transduction, while both of them retained the capability of transducing the heart (Fig. 4, Supplementary Tables 2 and 3). Although Q589A had only an ancillary role in the context of AAV2R585E.9-2, this study also demonstrated that it is feasible to design new AAV mutants with directed phenotypes based on the knowledge obtained from the AAV Barcode-Seq data.

High-resolution functional maps of the AAV9 capsid. Finally, with all the data being combined, we were able to draw high-resolution functional maps of the amino-acid residues in the C-terminal half of the AAV9 capsid. By two-dimensional (2D) heat mapping, one can readily appreciate that functionally important amino acids form four clusters (Clusters I–IV, Fig. 6a). Cluster II contains the most amino-acid residues causing the LD phenotype when mutated. On the other hand, functionally important amino acids in Clusters I, III and IV are dominated by those that play a more fundamental role in any transduction events. In keeping with this functional distinction between the clusters, a three-dimensional (3D) mapping of Clusters I–IV reveals that Cluster II is on the side walls of the threefold capsid protrusions and topologically distinct from the Clusters I, II and IV, which mainly reside in the twofold axis valley on the capsid (Fig. 6b). By displaying the data in a Venn diagram, it is revealed that many of the functionally important...
amino-acid residues, particularly those in Cluster II, play a multifunctional role (Fig. 6c). Although detailed investigation and interpretation of each of many amino acid–phenotype correlations we have identified is out of the scope of this article, all the data presented here demonstrate that the Barcode-Seq approach combined with DNA barcode-tagged mutagenesis offers a novel powerful means to study multifunctional viral capsid proteins and seek clues in designing novel viral vectors for gene therapy.

Discussion

We present here a novel NGS-based methodology to comprehensively characterize various aspects of biological phenotypes of many AAV strains in a high-throughput manner. Unlike the commonly used AAV capsid directed evolution methods to screen random peptide display libraries or shuffled capsid libraries, AAV Barcode-Seq per se is not aimed at identifying improved AAV capsids with desired phenotypes. Rather, AAV Barcode-Seq is a method that has enormous power...
to collect a large data set on capsid amino-acid sequence-phenotype correlation in a short period of time, using only a small number of replicates. We show that, when this new method is applied to capsid mutant libraries, comprehensive functional dissection of multifunctional viral capsid proteins becomes possible. Such functional mapping of viral capsids, which was not possible by conventional methodologies, significantly furthers our understanding of viral capsid biology and unambiguously identifies functional amino-acid residues that would provide important targets for protein engineering directed at altered functions. Similar NGS-based approaches have recently been reported in other experimental settings. However, to the best of our knowledge, this is the first demonstration that an NGS-based approach enables us to draw a comprehensive high-resolution functional map of a multifunctional protein such as the AAV capsid protein we studied in this report.

AAV Barcode-Seq should not be mistaken for a complicated and labour-intensive approach. Although the validation studies reported here required substantial time and efforts, Barcode-Seq per se is straightforward and relatively effortless procedure that involves only the following routine molecular biology techniques and basic bioinformatics tasks: DNA extraction from samples, DNA-PCR, Illumina sequencing of PCR products and data analysis. Illumina sequencing can be done at NGS resources at a steadily decreasing cost, and NGS data analysis has now become a standard bioinformatic task in many molecular biology laboratories as well as bioinformatics laboratories. In addition, the labour required for one DNA-barcoded AAV library production and purification is not substantially more than that for one conventional AAV vector preparation once all plasmid DNAs for AAV production have been prepared. For example, we will be able to compare relative transduction efficiencies of 10 different laboratory-engineered AAV strains of interest in major organs in AAV-injected animals using only three animals and one DNA-barcoded AAV library if we take the AAV Barcode-Seq approach. In contrast, for a side-by-side comparison, conventional approaches would require at least 30 animals (n = 3 or more per AAV strain) and 10 AAV preparations that are individually prepared and purified on a scale comparable to that of the DNA-barcoded AAV library.

We used the AAV rep–cap genome for this proof-of-concept study; however, AAV viral genomes tagged with DNA barcodes can be any type of DNA devoid of the rep and cap genes. The core DNA sequence essential for our Barcode-Seq approach is the 96 nucleotide-long DNA barcode cassette composed of a pair of random 12 nucleotides, three PCR primer-binding sites and two restriction enzyme recognition sites for Nhe I and Bsr GI (Fig. 1a). This DNA barcode cassette has been thoroughly validated for the Barcode-Seq analysis and in theory should function in any context of DNA sequence. The novel approach reported here will therefore readily be evolved into a more universal system by incorporating the DNA barcode cassette into a standard recombinant AAV viral genome devoid of the rep and cap genes and supplying these viral genes in trans as an AAV helper plasmid when DNA-barcoded AAV libraries are produced. Thus, tissue tropism, transduction efficiencies and any aspects of biological properties of a panel of many AAV capsid mutants of interest expressing a marker gene or a therapeutic gene can readily be compared side-by-side by AAV Barcode-Seq using DNA-barcoded AAV libraries produced by the standard three plasmid transfection method without a need of constructing the AAV rep–cap genome with DNA barcodes. A potential caveat of AAV Barcode-Seq is that the quantity of AAV vector genome DNA in cells may not necessarily correlate with the level of transgene expression. This caveat, however, will be addressed by co-transcribing DNA barcodes into RNA. Such an RNA barcode-based system will allow relative quantification of AAV vector-mediated transgene expression levels from many AAV mutants in a limited number of replicates. Thus, the AAV Barcode-Seq approach and more broadly high-throughput phenotypic characterization of barcode-tagged viruses by NGS in general have a potential to expand their utility, immediately and in the future, beyond the scope of the study reported here.

An important question in viral vector research is how a spectrum of biological properties of vectors is determined by viral components. In this regard, the large data set we generated and the AAV capsid functional map we presented here not only reinforce previous observations but also provide new insights into the AAV capsid biology. For example, one can now appreciate that liver transduction of AAV9 is governed by two mechanisms: one requiring galactose binding and another that utilizes an independent pathway(s) (Fig. 6c). This in part supports the previous notion that a decrease in galactose binding avidity results in a liver-detrargeting phenotype with systemic transduction. However, this notion is not entirely correct because our study shows the presence of galactose-binding-independent liver transduction pathways involving several amino acids such as P504, G505 and Q590. The observations from galactose-binding-deficient AAV9 mutants E500A/F501A and W503A and a galactose-binding-proficient and transduction-impaired mutant AAV2R858E.9-3 (or AAV2R858E.9-2 F500Y/A501/S/N514D) indicate the potential presence of an amino-acid residue(s) in the vicinity of E500-W503 in the AAV9 capsid that is not only responsible for galactose binding but also involved in postattachment viral processing. Such a dual role of amino-acid residues provides an alternative interpretation for the liver-detrargeting nature of galactose-binding-deficient mutants, such as AAV9W503R reported by Shen et al. and AAV9W503A described in this study. That is, impaired postattachment viral processing and decreased galactose binding can coincidentally result from a mutation, while the mere ablation of galactose binding does not necessarily result in a liver-detrargeted phenotype.

Another new insight is that AAV9-mediated liver transduction likely requires an additional viral processing step that is not required for transduction in nonhepatic tissues. This insight comes from the following observation: AAV9 mutants that showed a >90% decrease (dark blue in Fig. 6a) in transduction in any of the nonhepatic tissues also exclusively exhibited a >90% decrease in liver transduction (odds >11), while impaired liver transduction did not necessarily accompany significantly attenuated transduction in multiple nonhepatic tissues, as observed in the LD mutants (odds = 0.24–0.81). The straightforward interpretation of this observation is that the liver poses an additional hurdle to AAV9 to achieve transduction and the AAV9 capsid possesses amino-acid residues that are functionally required for overcoming this hurdle but are not required for overcoming transduction barriers in nonhepatic tissues. Although the nature of this liver-specific transduction barrier has yet to be elucidated, our study has indicated that it could be a postattachment barrier because more than a half of the amino acids associated with the LD phenotype also play a role in postattachment viral processing (Fig. 6c).

In summary, we have successfully established a new method to collect a large data set on amino-acid sequence-viral capsid phenotype correlation in a high-throughput manner and used it to draw a high-resolution functional map of the AAV capsid protein. A complete AAV Barcode-Seq data set on phenotypes of all the AAV serotypes and mutants we analysed is provided as Supplementary Data 1–3. As more data accumulate and are explored by further data mining, we will be able to understand AAV virus and vector biology in much greater detail.
Barcode-Seq, when combined with other complementing methodologies, will ultimately allow us to design and create novel AAV vectors endowed with the most desired biological properties for each clinical application. Importantly, this approach can be readily adapted to studies involving animals with more translational relevance, such as nonhuman primates.

Methods

Cell culture conditions and experiments. Human embryonic kidney (HEK) 293 cells, AAV293, were purchased from Stratgene. CHO Pro5 and Lec2 cells were gifted by A. Asokan (University of North Carolina, Chapel Hill). HEK293 cells were grown in Dulbecco’s Modified Eagle’s Medium (DMEM, Lonza) and CHO Pro5 and Lec2 cells were maintained in Alpha-Minimum Essential Medium (Alpha-MEM, Sigma-Aldrich). The media were supplemented with 10% fetal bovine serum (FBS), L-glutamine and penicillin-streptomycin. Frozen cell stocks were created without further authentication from the original vials we received before use for experiments. These cell lines have been tested for mycoplasma contamination using a mycoplasma 16S rDNA PCR assay and found to be negative. The virus cell surface binding and transfection assays were performed using CHO Pro5 and Lec2 cells seeded on 24- or 96-well plates 1 day before each experiment as detailed in Supplementary Methods.

Plasmid construction. pAAV-Serotype-x-VBC-y (Serotype-x = serotype 1, serotype 2, serotype 3, ...), pAAV9- SBBANN-AA-x-VBC-y (AA-x = AA356, AA358, AA360, ...), and pAAV2R585E-SBBXEB-HP-x-VBC-y (HP-x = 441-00700, 441- 16000, 443-00009, 443-16000) are the AAV plasmids with which we produced each AAV serotype and mutant viral clone with a clone-specific pr-VBC-y (y = 1, 2, 3, 4, 5, 6, 7, 8, 9). These plasmids are all derivatives of pAAV9-SBBANN-VBCLIB (accession code, KF032296) or pAAV2R585E-SBBXEB-VBCLIB (accession code, KF032297). The detailed methodological information about plasmid construction and high-throughput double alanine and hexapeptide scanning mutagenesis can be found in Supplementary Methods.

Production of DNA-barcoded AAV libraries and AAV vectors. We produced DNA-barcoded AAV libraries using an adenovirus-free plasmid transfection method and purified them by two cycles of calcium chloride ultracentrifugation.53,54 HEK293 cells were transfected with each AAV library plasmid (that is, either pAAV-Serotype-x-VBC-y, pAAV9-SBBANN-AA-x-VBC-y or pAAV2R585E-SBBXEB-HP-x-VBC-y) together with 15 μg of pHelper (Stratagene) in separate 15-cm dishes by a calcium phosphate transfection method.

Animal experiments. All the animal experiments were performed according to the guidelines for animal care at University of Pittsburgh and Oregon Health & Science University. We used 8-week-old C57BL/6 male mice and C57BL/6 Rag1−/− mice purchased from the Jackson Laboratory. We randomly assigned each animal to each experimental group without considering body weights of animals. In the AAV Barcode-Seq studies, we injected C57BL/6 mice with AAV2 or AAV9 virions intravenously at a dose of 1 × 1012 vg per mouse for the assessment tissue transduction efficiencies, and at a dose of 1 × 1013 vg per kg for pharmacokinetic studies.44 To immunize C57BL/6 mice with AAV1 or AAV9, we injected mice intravenously with 1 × 1012 vg per mouse of AAV-CMV-lacZ vector packaged with the same AAV9 capsid, respectively. In the AAV Barcode-Seq studies as detailed in Supplementary Data 1, 2 and 3. This sample size has been justified by the statistical power of the AAV Barcode-Seq analysis as shown in Supplementary Fig. 2. To validate the AAV Barcode-Seq results and intermediate similarity tropism and transduction efficiency, we injected 8-week-old C57BL/6 male mice and C57BL/6 Rag1−/− male mice with AAV-CMV-lacZ vector packaged with AAV capsids of interest intravenously at two doses, 3 × 1011 and 1 × 1012 vg per mouse. For this validation study, we used n = 3 per group. We sacrificed the vector-injected C57BL/6 mice and C57BL/6 Rag1−/− mice 11 days and 6 weeks post-injection, respectively, to determine transduction efficiencies in the following 12 major organs: brain, heart, lung, liver, kidney, spleen, intestine, pancreas, testis, hind limb skeletal muscle, visceral adipo tise and dorsal skin. All the animal experiments were performed in a non-blinded fashion.

AAV Barcode-Seq analysis. We extracted total DNA from cultured cells and tissues using Nucleospin Tissue Kit (MACHEREY-NAGEL, Duren, Germany) or by phenol–chloroform. We used 100 ng of total DNA to PCR-amplify 11- and 17-bp VBCs (that is, VBC-PCR). We extracted viral DNA from AAV library stocks using Wako DNA Extraction Kit (Wako Chemicals, Richmond, USA) and used 25 μl of extracted nucleic equivalents in each PCR cycle. For blood samples, we amplified VBCs directly from the blood without extracting DNA using the lysis and neutralization buffers that come with Extract-N-Amp Blood PCR Kit (Sigma, Saint Louis, USA). We used 0.1 μl of the whole blood for PCR. This quantity gave us PCR signals sufficient for the downstream analysis and was empirically determined. The PCR primers we used for the amplification of 11-bp and 17-bp VBCs are as follows: It-VBC-For (FSN-SBC-ACCTACGTACTTCCGCTCAT), It-VBC-Rev (FSN-SBC-TGGGCAATGCATTTCTCTTG), rt-VBC-For (FSN-SBC-AACGAAATAACATGTCGGGGA) and rt-VBC-Rev (FSN-SBC-CTTCCTGCGGGCTTCTTTGGC). Each primer had a 3–4 nucleotide-long Sample-specific Bar Code (SBc) and a 9–4 Frame-Shifting Nucleotides (FSN) at the 3′ end. We incorporated SBCs for multiplexed Illumina sequencing and FSNs for overcoming the issue of low sequence diversity of PCR products in reference image construction.36 The PCR cycles used for the VBC-PCR were 2 min at 95 °C, 35 cycles of 15 s at 95 °C and 30 s at 68 °C, and subsequently 5 min at 68 °C. We quantified each PCR product by agarose gel electrophoresis followed by densitometry of ethidium bromide-stained DNA. We then pooled up to 96 multiplexed PCR products at an equimolar ratio and performed Illumina sequencing as detailed in Supplementary Methods. We assessed the quality of Illumina raw sequence reads by FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The following quality control parameters were used: sequence quality, per base sequence quality, per base Q30 content and sequence length distribution, were all met in all the data sets we used in this study. We extracted the data on sequence read numbers of each 11- and 17-bp VBCs from each sample from Illumina fastq files. To do this, we developed an algorithm for binning sequence reads by SBC, subsequently by 11-bp and 17-bp VBCs, and implemented it in Perl at the Pittsburgh Supercomputing Center.

In the AAV Barcode-Seq analysis of a sample, we determined PD (xi, yj), which indicates a Phenotypic Difference between an AAV strain xi(AAVs) and the reference control strain yj(AAVs) contained in an AAV library, based on Illumina sequence read numbers. To determine PD (xi, yj), we defined the following functions: RNLS_lt_VBC(x), RNLS_rt_VBC(x), RNLS_MAP(x), RNLS_flt_VBC(x), RNLS_frt_VBC(x, y), RNLS_MAP(x, y), RNLS_MAP(x, y, n), GNR_lt_VBC(x, y, n), GNR_flt_VBC(x, y, n), k∈{i, j, 1, 2, ..., Ns}, n∈{1, 2, ..., Ns}. In x, y and n refer to AAV strains and AAV clones of the same AAV strain, respectively. j refers reference control AAV strain or AAV clone from two values: AAV9 or AAV9 and AAV9, respectively. For these functions, RNLS, GNR, RR, and PD stand for Read Number in Library Stock, Read Number in Experimental Sample, Raw Ratio, Globally Normalized Ratio and Phenotypic Difference, respectively; Ns is the number of AAV strains contained in an AAV library; Ns is the number of AAV clones derived from the same AAVs(x, y) in the library; Ns is the maximum number of AAV clones representing the same AAV strain, which is the same as the number of clones of the reference controls; Ns is the number of replicates. AAVs9 and AAVs2 are reference control AAV9 and AAV2R585E clones, respectively. For example, AAV9- AA9-VBCLIB-1 (Table 1) contains 15 AAV9 reference control clones, 15 AAV9- AAV9-VBCLIB-1 contains 148 AAV9 mutant clones representing 92 different AAV9 mutants (that is, 2 clones per mutant). Therefore, Ns, Ns, Ns and the number of combinations of this library are 94, 2, 15 and 214, respectively; AAVs9, j∈{i, j, 1, 2, ..., 15}, are AAV9 clones: AAVs9, j∈{i, j, 1, 2, ..., 15} are AAV2R585E clones, and AAVs9, j∈{i, 3, ..., 94}, j∈{i, 1, 2} are AAV9 mutant clones.

Acknowledgments. We thank the Barcode-Seq study team for their expert assistance in performing Illumina sequencing. We also thank R. Damoiseaux for his technical assistance in the AAV Barcode-Seq studies.
RNES.lt_VBC (xij, yk)n and RNES.rt_VBC (xij, yk)n are Illumina sequence read numbers of lt-VBC of AAVx, and those of rt-VBC of AAVx, respectively, representing a sample obtained from the replicated experimental set ‘n’ that uses the same AAV library stock. If an experiment is done in triplicate (experimental sets 1, 2 and 3), n = 1, 2 or 3 is given to each experimental set. With these values, we calculated RR.lt_VBC (xij, yk)n and RR.rt_VBC (xij, yk)n using the following formulas (1, 2):

\[
RR.lt_VBC(xij,yk)n = \frac{RNES.lt_VBC(xij,yk)n}{RNES.rt_VBC(xij,yk)n}
\]

\[
RR.rt_VBC(xij,yk)n = \frac{RNES.rt_VBC(xij,yk)n}{RNES.rt_VBC(xij,yk)n}
\]

GNR.lt_VBC (xij, yk)n and GNR.rt_VBC (xij, yk)n are globally normalized RR.lt_VBC (xij, yk)n and RR.rt_VBC (xij, yk)n, respectively, when the reference control AAVY is selected for comparison. We calculated GNR.lt_VBC (xij, yk)n and GNR.rt_VBC (xij, yk)n using the following formulas (3, 4):

\[
GNR.lt_VBC(xij,yk)n = \frac{\sum n_{xij} RNES.lt_VBC(xij,yk)n/(No.lt_VBC(xij))}{\sum n_{xij} RNES.rt_VBC(xij,yk)n/(No.rt_VBC(xij))}
\]

\[
GNR.rt_VBC(xij,yk)n = \frac{\sum n_{xij} RNES.rt_VBC(xij,yk)n/(No.rt_VBC(xij))}{\sum n_{xij} RNES.rt_VBC(xij,yk)n/(No.rt_VBC(xij))}
\]

Please note that No.lt (xij, yk)n and No.rt (xij, yk)n are the number of outilers in the RR.lt_VBC and RR.rt_VBC values of reference controls (that is, AAVx1 or AAVx2) identified based on the three times the interquartile range. Therefore, when outliers were removed, the number of summation elements in the above formula was fewer by the number of outliers.

We then calculated the PD (xij, yk)n (i ≥ 3), which represents the average of GNR.lt_VBC and GNR.rt_VBC values, using the following formula (5): PD(xij, yk)n = \[\sum n_{xij} GNR.lt_VBC(xij,yk)n + \sum n_{xij} GNR.rt_VBC(xij,yk)n \] / [Nj × Nk × 2 - No.lt(xij)]

No (xij, yk)n is the number of outliers in the GNR.lt_VBC and GNR.rt_VBC values of AAV strains besides the reference controls (AAVx1, AAVx2). We used a non-parametric test because Shapiro–Wilks test for normality revealed that GNR data sets with which PD (xij, yk)n was determined (that is, GNR.lt_VBC (xij, yk)n and GNR.rt_VBC (xij, yk)n values) do not necessarily follow normal distribution. To assess correlation between two phenotypes, we used a r² test or Fisher’s exact test. We applied a Monte Carlo approach to determine statistical power of the Barcode-Seq analysis.

Biocomputers and computer modelling of the AAV capsids. We collected nucleotide sequence information of 128 AAV strains from GenBank and calculated evolutionary scores by ConSurf with the default parameters. We obtained 3D structure coordinates of the AAV2 and AAV9 capsids from the Protein Data Bank (1lp3 and 3u1l for AAV2 and AAV9, respectively). With these coordinates, we generated AAV2 and AAV9 capsid oligomers comprising nine subunits or full capsids using VIPERdb Oligomer Generator49 and visualized them using PyMOL. We determined topological locations of each capsid amino acid in a tertiary and quaternary structure by visual inspection of the surface-rendered structural model of the AAV capsids using PyMOL.

Statistical analysis. We assessed phenotypic differences between AAVx1 and AAVY by two-tailed Mann–Whitney U-test. We used a non-parametric test because Shapiro–Wilks test for normality revealed that GNR data sets with which PD (xij, yk)n was determined (that is, GNR.lt_VBC (xij, yk)n and GNR.rt_VBC (xij, yk)n values) do not necessarily follow normal distribution. To assess correlation between two phenotypes, we used a r² test or Fisher’s exact test. We applied a Monte Carlo approach to determine statistical power of the Barcode-Seq analysis.

References
1. Mannio, 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).
2. Nathwani, A. C. et al. Adenovirus-associated virus vector-mediated gene transfer in hemophilia B. N. Engl. J. Med. 365, 2357–2365 (2011).
3. Boutin, S. 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).
4. Calcedo, R., Vandenberge, L. H., Gao, G., Lin, J. & Wilson, J. M. Worldwide epidemiology of neutralizing antibodies to adeno-associated viruses. J. Infect. Dis. 199, 381–390 (2009).
5. Asokan, A., Schaffer, D. V. & Samulski, R. J. The AAV vector toolkit: poised at the clinical crossroads. Mol. Ther. 20, 699–708 (2012).
6. Shen, S. et al. Glycan binding avidity determines the systemic fate of adeno-associated virus type 9. J. Virol. 86, 10408–10417 (2012).
7. Opie, S. R., Warrington, J. R., K. H., Agbandje-McKenna, M., Zlotukhin, S. & Muzykyna, N. Identification of amino acid residues in the capsid proteins of adeno-associated virus type 9 that contribute to heparin sulfate proteoglycan binding. J. Virol. 77, 6995–7006 (2003).
8. Kern, A. et al. Identification of a heparin-binding motif on adeno-associated virus type 2 capsids. J. Virol. 77, 11072–11081 (2003).
9. Bell, C. L., Gurdz, L. B., Van Vliet, K., Agbandje-McKenna, M. & Wilson, J. M. Identification of the galacose binding domain of the adeno-associated virus type 2 capsid. J. Virol. 76, 7232–7233 (2002).
10. Wu, Z. et al. Single amino acid changes can influence titer, heparin binding, and tissue tropism in different adeno-associated virus serotypes. J. Virol. 80, 11393–11397 (2006).
23. Lerch, T. F., Xie, Q. & Chapman, M. S. The structure of adeno-associated virus capsid. J. Virol. 86, 11877–11885 (2012).

11. Wu, P.

17. DiPrimio, N., Asokan, A., Govindasamy, L., Agbandje-McKenna, M. & Samulski, R. J. Surface loop dynamics in adeno-associated virus capsid assembly. J. Virol. 82, 5178–5189 (2008).

34. Maheshri, N., Koerber, J. T., Kaspar, B. K. & Schaffer, D. V. Directed evolution for human gene therapy. Nat. Biotechnol. 20, 1073–1080 (2002).

41. Zincarelli, C., Soltsy, S., Rengo, G. & Rabinowitz, J. E. Analysis of AAV serotypes 1–9 mediated gene expression and tropism in mice after systemic injection. Mol. Ther. 15, 1073–1080 (2008).

40. Inagaki, K. et al. A potential role of distinctively delayed blood clearance of recombinant adeno-associated virus serotype 9 in robust cardiac transduction. Mol. Ther. 19, 1079–1089 (2011).

19. Wu, Z., Miller, E., Agbandje-McKenna, M. & Samulski, R. J. N-linked sialic acids facilitate efficient binding and transduction by adeno-associated virus types 1 and 6. J. Virol. 80, 9093–9103 (2006).

45. Deutscher, S. L., Nuwayhid, N., Stanley, P., Briles, E. F. & Hirschberg, C. B. Translocation across Golgi vesicle membranes: a CHO glycosylation mutant deficient in CMP-sialic acid transport. Cell 39, 295–299 (1984).

44. DiMattia, M. A.

13. Raupp, C. et al. The threefold protrusions of adeno-associated virus type 8 are involved in cell surface targeting as well as postattachment processing. J. Virol. 86, 9396–9408 (2012).

43. Zhang, H. et al. Several rAAV vectors efficiently cross the blood-brain barrier and transduce neurons and astrocytes in the neonatal mouse central nervous system. Mol. Ther. 19, 1440–1441 (2011).

33. Asokan, A.

25. Wu, J. T., Kao, T. K., Hsu, C. S., Huang, L. & Wang, L. Structural insight into the unique properties of adeno-associated virus serotype 8. J. Virol. 86, 6947–6958 (2012).

47. Walters, R. W. et al. Binding of adeno-associated virus type 5 to 2.3-linked sialic acid is required for gene transfer. J. Biol. Chem. 276, 20610–20616 (2001).

46. Deutscher, S. L. et al. Clades of adeno-associated viruses are widely disseminated in human tissues. J. Virol. 78, 6381–6388 (2004).

49. Shen, S. et al. Engraftment of a galactose receptor footprint onto adeno-associated viral capsids improves transduction efficiency. J. Biol. Chem. 288, 28814–28823 (2013).

50. Lu, R. et al. Tracking single hematopoietic stem cells in vivo using high-throughput sequencing in conjunction with viral genetic barcoding. Nat. Biotechnol. 29, 928–933 (2011).

52. Patwardhan, R. P. et al. Massively parallel functional dissection of mammalian enhancers in vivo. Nat. Biotechnol. 30, 265–270 (2012).

53. Melnikov, A. et al. Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. Nat. Biotechnol. 30, 271–277 (2012).

54. Matsuhashi, T. et al. Adeno-associated virus vectors can be efficiently produced without helper virus. Gene Ther. 5, 938–945 (1998).

55. Xiao, X., Li, J. & Samulski, R. J. Production of high-titer recombinant adeno-associated virus vectors in the absence of helper adenovirus. J. Virol. 72, 2224–2232 (1998).

26. Burton, M. et al. Coexpression of factor VIII heavy and light chain adenoviral vectors produces biologically active protein. Proc. Natl Acad. Sci. USA 96, 12260–12271 (2007).

28. Inagaki, K. et al. Engineering liver-detargeted AAV9 vectors for cardiac and adult astrocytes. Nat. Biotechnol. 27, 95–96 (2009).

29. Inagaki, K. et al. A myocardium tropic adeno-associated virus (AAV) evolved by directed evolution of adeno-associated virus to an a myocardium tropic adeno-associated virus (AAV) evolved by directed evolution of adeno-associated virus to an
**Additional information**

**Accession codes**: Sequences of plasmids pAAV9-SBBANN-VBCLib and pAAV2R585E-SBBXEB-VBCLib have been deposited in the GenBank nucleotide core database under the accession codes KF032296 and KF032297, respectively.

**Supplementary Information** accompanies this paper at [http://www.nature.com/naturecommunications](http://www.nature.com/naturecommunications).

**Competing financial interests**: T.E. and Y.K. are employees who receive salary from Takara Bio Inc. K.A. and H.N. are inventors of the technology arising from this work and licensed by Takara Bio Inc.

**Reprints and permission** information is available online at [http://npg.nature.com/reprintsandpermissions/](http://npg.nature.com/reprintsandpermissions/)

**How to cite this article**: Adachi, K. et al. Drawing a high-resolution functional map of adeno-associated virus capsid by massively parallel sequencing. *Nat. Commun.* 5:3075 doi: 10.1038/ncomms4075 (2014).

This article is licensed under a Creative Commons Attribution 3.0 Unported Licence. To view a copy of this licence visit [http://creativecommons.org/licenses/by/3.0/](http://creativecommons.org/licenses/by/3.0/).