Supplemental results:

Characterization of highly diverse barcoded capsid shuffled AAV libraries

We first generated a library with random barcodes (BC) downstream of the cap polyadenylation site. Insertion of the BC sequence did not negatively impact virus function as shown by replication studies for wild-type AAV with and without the BC (data not shown). HTS of the barcode plasmid library showed a high degree of diversity with over 93% of all BC reads containing a unique sequence (Supplemental Table 2). As HTS of the barcodes involved an amplification step, it is most likely that the actual number of repetitive barcodes in the BC library was lower than that. The size of the barcode library was estimated to be $1x10^7$ based on the number of colonies obtained from an aliquot of the transformation reaction (data not shown). Two shuffled capsid libraries were then generated using sequences from 10 related AAVs (10 parent library) or 18 more diverse AAVs (18 parent library) and cloned into the BC library vector. A phylogenetic analysis of the parental sequences used for shuffling is depicted in Supplemental Figure 1A. Importantly, two of those capsid sequences – DJ and LK03 – had been derived from previous screens performed in our laboratory using earlier generation libraries and were added as parentals. The crossover analysis of those shuffled VP1 capsid sequences is shown in Supplemental Figure 1B, the crossover of the AAP sequences is shown in Supplemental Figure 1C. The capsid shuffled barcoded 10 parent library and the 18 parent library each each had a size of approximately $5x10^6$ clones. Given the low number of replicates after BC sequencing the complexity was estimated to be at least $1x10^6$ for each of the libraries.

The degree of library diversity was assessed using three different methods: HTS of the barcodes (MiSeq platform), PacBio sequencing of the capsids including the barcodes, as well as

Sanger sequencing of the capsids including the barcodes. Amplification and HTS of the barcodes from the plasmid libraries as well as the AAV libraries revealed a very high degree of complexity with low numbers of replicates. At least 1 million reads were obtained for each sample (Supplemental Table 2). When the 10 parent library was analyzed on the plasmid level, i.e. prior to production in 293T cells, it was found that over 88% of all reads contained unique sequences, with a maximum number of replicates found to be 7 out of 1.25 million reads. For the 18 parent plasmid library, more than 88% of the reads were unique -the maximum number of reads obtained for one single BC sequence was 8 out of 1.8 million reads. Analysis of the NGS data for the AAV libraries revealed approximately 76% and 80% unique BC sequences for the 10 parent and 18 parent libraries, respectively (Supplemental Table 2).

Several capsids from the 10 parent and the 18 parent plasmid libraries as well as from the 10 parent AAV library were also analyzed by Sanger sequencing and subjected to recombination analysis (Supplemental Figures 2 & 3). Since parental capsids DJ and LK03 are chimeras derived mostly from serotypes also used in this study (Supplemental Figure 1B), they are not included as separate parental sequences here. For analysis of diversity on the AAV level, the region spanning capsid and barcodes was amplified from extracted viral genomes, cloned and sequenced (Supplemental Figure 2, right panel). In addition to depicting crossover figures of the shuffled capsids, we used the parental contribution data generated using the Xover software to perform conservation analysis throughout the capsid sequences. The conservation values for each amino acid residue were calculated (Supplemental Figures 4A & 4B) allowing us to evaluate if sequence diversity within the libraries was drastically diminished when compared to the parental sequences. The conservation patterns of the 10 parent library on the plasmid as well as the AAV level closely matched those of the parental sequences (Supplemental Figure 4A). On the other hand, the 18

parent plasmid library exhibited higher conservation levels than the parents throughout most of the capsid (Supplemental Figure 4B). Crossover analysis of the capsid amino acid sequences showed that the 10 parent library was well shuffled with an average rate of 11 crossover events per capsid. With an average of 8 crossovers, the 18 parent library was less thoroughly shuffled (Supplemental Figure 3). This observation can be explained with the large proportion of parental capsids that have low homology to each other throughout large stretches. Crossovers based on an identity of less than 15 bases are difficult to obtain (1). Thus, it is not surprising that fragments with low homology to each other do not easily recombine when equimolar amounts of each parent are mixed together. The observation that fewer crossovers were found within the 3' half of the library capsids can also be explained by this limitation as this part of the capsid contains several stretches of low sequence homology between the different serotypes. Complexity of the 10 parent library on the AAV level was found to be similar to that on the plasmid level indicating that capsid assembly during AAV production in 293T cells does not reduce diversity (Supplemental Figures 2 & 4A). Moreover, the crossover analysis patterns show that the parent serotypes appear to be represented roughly equally and randomly within the libraries, although AAV3B sequences seem to be slightly favoured in the 3' part of the shuffled clones. In addition to Sanger sequencing of individual capsids, we performed PacBio sequencing of the 10 parent library on the plasmid level prior to transfection into 293T cells as well as on the AAV level. We obtained 21,000 full-length reads for the 10 parent plasmid library and about half of that read number for the 10 parent AAV library. The graphs in Supplemental Figure 4C depict the contribution of each parental amino acid residue throughout the length of the capsid. As already observed by Sanger sequencing, PacBio sequencing revealed that the 5' half of the capsid gene contained a more even distribution of all parental sequences than the 3' half. Particularly residue contributions from AAV-porcine2 and

AAV9hu14 were starkly diminished in the sequence stretch between amino acid positions 450 and 650 while AAV3B sequences were overrepresented. An overrepresentation of AAV3B sequences is possibly at least in part due to the fact that LK03, which contains large sequence stretches from AAV3B (Supplemental Figure 1B), was used as a parental for shuffling.

We also determined if each capsid was linked to a unique barcode sequence and found very high levels of capsid-BC linkage. According to PacBio sequencing, only 0.5% of all sequences had different capsids sharing identical barcodes.

Supplemental methods:

Generation of an AAV vector containing a library of unique barcode sequences. For generation of the barcodes an oligonucleotide with an AgeI restriction site sequence on the 5'- end followed by 12 random nucleotides and the 20nt long spacer sequence (CTA AAC CGG TNN NNN NNN NNN NAC GGA AAT ACG ATG TCG GGA) was annealed to an oligonucleotide containing an Eag I site on the 5' end followed by 12 random nucleotides as well as the antisense spacer sequence (TTC TCG GCC GNN NNN NNN NNN NTC CCG ACA TCG TAT TTC CGT) and extended using Klenow Polymerase devoid of exonuclease activity (NEB). Fragments were purified using a Qiaquick PCR purification kit (Qiagen) and subsequently digested with AgeI and EagI and purified using a Qiaquick PCR purification column. Vector was digested with the same restriction enzymes, dephosphorylated, phenol-chloroform purified and Ethanol precipitated. The optimum vector to insert ratio was evaluated by setting up several ligation reactions initially and testing for possible multiple barcode inserts by performing colony PCR with primers rightF (CGC GCC ACT AGT AAT AAA C) and QSeqRev (TAG AGC AAC TAG AGT TCG). For the scale-up reaction barcodes were ligated with the vector at a molar vector to insert ratio of 1 to 2.5 in a total volume of 30 ul, de-salted using Strataclean resin according to instruction (Agilent Technologies) and electroporated in 2 ul aliquots into DH10B-MegaX cells (Thermo Fisher). Electroporated cells were pooled and used to inoculate 500 ml LB-Amp medium. An aliquot was plated to assess library size and diversity. After 16 hrs in a 37°C shaker bacteria were harvested and plasmid DNA was isolated using a Megaprep kit (Qiagen). ITR integrity was confirmed by digestion with XmaI as well as with AhdI. Several individual clones from the test plate were sequenced to assess barcode diversity.

Generation of capsid shuffled barcoded AAV libraries. All parental capsid sequences used for shuffling contained a PacI site immediately 5' and an AscI site 3' of cap and had been cloned into pBluescript. Initially, parental sequences were amplified individually using primers located in the flanking pBluescript sequences (outer F: AAT TAA CCC TCA CTA AAG G, outer R: GTA ATA CGA CTC ACT ATA GGG C). Phusion Hot Start Flex polymerase (NEB) was used for amplification and 25 PCR cycles were employed (30 sec 98°C, 25 cycles of 10 sec 98°C, 15 sec 56°C, 1 min 15-sec 72°C, followed by 10 min 72°C). PCR products were purified using the Qiaquick PCR purification kit (Qiagen), all 18 capsids (for the 18 parent library) or 10 capsids (for the 10 parent library, AAV1, AAV2, AAV3B, AAV6, AAV8, AAV9hu14, AAV12, AAV rhesus10, DJ, LK03) were pooled in equimolar ratio and fragmented at room temperature (RT) using DNAse I (Sigma). At different incubation time points aliquots were analyzed on an 1.5% agarose gel while the reaction was temporarily stopped by incubation in a dry ice / ethanol bath. Incubation time and DNAse concentration was adjusted until the majority of the fragments ranged from 100 bp to 500 bp. The entire reaction was then loaded on an 1.5% agarose gel and fragments in the desired size range were electroeluted from the gel, purified using two rounds of phenol-chloroform purification followed by one round of chloroform purification and ethanol precipitated. DNA fragments were then re-assembled in a primer-less PCR using Phusion Hot Start Flex polymerase and the following cycling conditions: 30 sec 98°C, 40 cycles of 10 sec 98°C, 30 sec 42°C, 45 sec 72°C, followed by 10 min 72°C.

Full-length capsid sequences were amplified from the assembly reactions using primers rescueF (GTC TGA GTG ACT AGC ATT CG) and rescueR (GTC TAC TGA AGC TCA CTG AG) and the following cycling conditions: 30 sec 98°C, 25 cycles of 10 sec 98°C, 15 sec 57°C, 1 min 15

sec 72°C, followed by 10 min 72°C. Amplicons were diluted 4-fold with fresh PCR mix and subjected to one additional cycle with a 10 min extension to fill up the ends. After concentrating PCR products using a PCR purification kit (Qiagen) they were digested with PacI and AscI and ligated into the BC library vector that had been treated with PacI, AscI, dephosphorylated and phenol-chloroform purified. Ligation reactions were de-salted using the Strataclean resin, electroporated into MegaX DH10B cells and expanded in liquid culture as described above for generation of the BC library vector. Small aliquots of transformed cells were plated to assess library size. Plasmid DNA was extracted from random colonies and *cap* and BC sequences were determined by Sanger sequencing using primers capF (TGG ATG ACT GCA TCT TTG AA), capF2 (ATT GGC ATT GCG ATT CC), and QSeqRev (TAG AGC AAC TAG AGT TCG).

AAV libraries were generated in HEK 293T cells using the calcium phosphate transfection method. Compared to the regular protocol the amount of transfected library plasmid DNA was reduced almost 20-fold to approximately 5000 copies per cell to minimize the likelihood of cross-packaging events taking place during AAV production. Briefly, 25 T225 flasks were seeded with 8E06 cells per flask in 40 ml media two days prior to transfection. On the day of transfection cells were between 80% and 90% confluent. 20 ml of media per flask was replaced with fresh media 1.5 hrs prior to transfection and a mixture of 40 ug pAd5 helper plasmid and 2 ug library plasmid in 4 ml 300mM $CaCl_2$ per T225 was prepared. Equal amounts of $CaCl_2$ / DNA mix and 2xHBS (280 mM NaCl, 50 mM HEPES pH 7.28, 1.5 mM Na_2HPO_4, pH 7.12) were mixed and 8 ml of the mixture was added to each flask. After 3 days cells were detached with 0.5 ml 500 mM EDTA each flask, pelleted, and resuspended in Benzonase digestion buffer (2 mM $MgCl_2$, 50 mM Tris-HCl, pH 8.5). AAVs were released from the cells by submitting them to three freeze-thaw cycles, non-encapsidated DNA was removed by digestion with Benzonase (200 U/ml, 1 hr 37°C), cell

debris was pelleted by centrifugation, followed by another $CaCl_2$ precipitation step (25 mM final concentration,1 hr on ice) of the supernatant and an AAV precipitation step using a final concentration of 8% PEG-8000 and 625 mM NaCl. Virus was resuspended in HEPES-EDTA buffer (50 mM HEPES pH 7.28, 150 mM NaCl, 25 mM EDTA) and mixed with CsCl to a final refractory index (RI) of 1.371 followed by centrifugation for 23 hrs at 45000 Rpm in an ultracentrifuge. Fractions were collected after piercing the bottom of the centrifuge tube with an 18-gauge needle and fractions ranging in RI from 1.3766 to 1.3711 were pooled and adjusted to an RI of 1.3710 with HEPES-EDTA resuspension buffer. A second CsCl gradient centrifugation step was carried out for at least 8 hrs at 65000 Rpm. Fractions were collected and fractions with an RI of 1.3766 to 1.3711 were dialyzed overnight against PBS, followed by another dialysis step against fresh PBS (4 hr), and dialysis (2 hr) against 5% sorbitol in PBS. All dialysis steps were carried out at 4°C. Virus was recovered from the dialysis cassette and pluronic F-68 was added to a final concentration of 0.001%. Virus was sterile-filtered, aliquoted, and stored in aliquots at -80°C. Genomic DNA was extracted from 10 ul of the purified virus using the MinElute Virus Spin Kit (Qiagen Cat#57704), and the viral genome titer was determined by qPCR using an AAV2 rep gene specific primer probe set (repF: TTC GAT CAA CTA CGC AGA CAG, repR: GTC CGT GAG TGA AGC AGA TAT T, rep probe: TCT GAT GCT GTT TCC CTG CAG ACA).

Cell culture conditions

Human islet cultures. Human pancreatic islets from deceased non-diabetic organ donors were provided by the Integrated Islet Distribution Program (IIDP) or the University of Alberta through the Stanford Islet Research Core and cultured in CMRL-1066 with 10% FBS, Pen-Strep, 1%

Insulin Transferrin Selenium (Thermo Fisher), 1mM sodium pyruvate, 2mM Glutamax, 2.5mM HEPES. Ultra-low attachment dishes were used for all islet cell culture experiments.

hESC derived β-cells: Mel1 $INS^{GFP/W}$ human embryonic stem cell (hESC)s were obtained from S. J. Micallef and E. G. Stanley (Monash Immunology and Stem Cell Laboratories, Australia). Cells were maintained and propagated on mouse embryonic fibroblasts (MEFs) in hESC media [DMEM/F12 (Gibco) with 10% KSR (Gibco), 10 ng/ml FGF-2 (R&D Systems)]. A stepwise differentiation of hESC toward β cells was carried out following the protocol described previously (2). Briefly, confluent hESCs were dissociated into single-cell suspensions using TrypLE (Gibco), counted and seeded at 5.5×10^6 cells per well in 6-well suspension plates in 5.5 ml hESC media supplemented with 10 ng /ml activin A (R&D Systems) and 10 ng/ ml heregulinB (Peprotech). The plates were incubated at 37 °C and 5% CO_2 on an orbital shaker at 100 rpm to induce 3D sphere formation. After 24 hours, the spheres were washed with PBS and resuspended in day 1 media. From day 1 to day 20, media was changed every day. Media compositions are as follows: Day 1: RPMI (Gibco) containing 0.2% FBS, 1:5,000 ITS (Gibco), 100 ng /ml activin A and 50 ng /ml WNT3a (R&D Systems). Day 2: RPMI containing 0.2% FBS, 1:2,000 ITS and 100 ng /ml activin A. Day 3: RPMI containing 0.2% FBS, 1:1,000 ITS, 2.5 μM TGFbi IV (CalBioChem) and 25 ng /ml KGF (R&D Systems). Day 4–5: RPMI containing 0.4% FBS, 1:1,000 ITS and 25 ng/ ml KGF. Day 6–7: DMEM (Gibco) with 25 mM glucose containing 1:100 B27 (Gibco) and 3 nM TTNPB (Sigma). Day 8: DMEM with 25 mM glucose containing 1:100 B27, 3 nM TTNPB and 50 ng /ml EGF (R&D Systems). Day 9–11: DMEM with 25 mM glucose containing 1:100 B27, 50 ng /ml EGF and 50 ng/ ml KGF. Day 12–20: DMEM with 25 mM glucose containing 1:100 B27, 1:100 Glutamax (Gibco), 1:100 NEAA (Gibco), 10 μM ALKi II (Axxora), 500 nM LDN-193189 (Stemgent), 1 μm Xxi (Millipore), 1 μM T3 (Sigma-Aldrich), 0.5 mM vitamin C, 1 mM N-

acetyl cysteine (Sigma-Aldrich),10 μM zinc sulfate (Sigma-Aldrich) and 10 μg /ml of heparin sulfate. At day 20, the spheres were collected, incubated with Accumax (innovative cell technologies) for 10 min at 37 °C and dissociated into single cells. Live GFP-high cells were sorted on Aria II at low flow rates and reaggregated in Aggrewell-400 (StemCell Technologies) at 1,000 cells per cluster in CMRL containing 10% FBS, 1:100 Glutamax (Gibco), 1:100 NEAA (Gibco), 10 μM ALKi II (Axxora), 0.5 mM vitamin C, 1 μM T3 (Sigma-Aldrich), 1 mM N-acetyl Cysteine (Sigma-Aldrich), 10 μM zinc sulfate (Sigma-Aldrich) and 10 μg /ml of heparin sulfate. At day 23, the reaggregated enriched β-clusters (eBCs) were transferred from Aggrewells and placed on orbital shakers at 100 rpm, and further cultured for 6 days. Media was changed every third day following reaggregation.

Human skeletal muscle stem cell and myotube cultures. A pool of primary muscle stem cells isolated from 6 individual donors (kind gift from G. Charville, Stanford) was frozen at an early passage and aliquots were used for experiments. Plates were coated with extracellular matrix protein (Sigma) at 1:500 in DMEM with 1% penicillin/streptomycin. The hMuSC medium was a 1:1 mixture of DMEM:MCDB media supplemented with 20% FBS, 1% insulin-transferrin-selenium, 1% antibiotic/antimycotic, and 10 μM p38i (Cell Signaling Technology Cat#SB203580) to maintain the stem state as described (3). Media for differentiating primary hMuSCs into myotubes lacked p38i and included a 2% horse serum starve instead of 20% FBS for 7 days. All media was changed every two days.

Mouse skeletal muscle myoblast cultures. Wild-type C2C12 mouse myoblasts (ATCC Cat# CRL-1772) were maintained in DMEM supplemented with 10% FBS and 1% antibiotic/antimycotic.

293 and 293T cell line cultures. HEK 293 cells (ATCC Cat# CRL-1573) and HEK 293T cells (ATCC Cat#CRL-3216) were cultured in DMEM with 10% FBS, 2 mM glutamine, 1% antimycotic-antibiotic, 11 mM HEPES pH 7.28, and 1 mM sodium pyruvate.

HeLa cell cultures. HeLa cells (ATCC Cat# CCL-2) were cultured in DMEM with 10% FBS, 2mM glutamine, 1% antimycotic-antibiotic.

Mouse pancreatic β-cell cultures. R7T1 cells (kind gift from H. Moeller, Stanford) were cultured in DMEM with 10% FBS, 2mM glutamine, 1% antimycotic-antibiotic, 1 ug/ml Doxycyclin.

Mouse pancreatic α-cell cultures. Alpha TC1 clone 6 cells (ATCC Cat# CRL-2934) were cultured in DMEM with 10% FBS, 2mM glutamine, 1% antimycotic-antibiotic, 15 mM HEPES, 0.1 mM NEAA.

Rat hepatoma cell cultures. H4TG cells (ATCC Cat# CRL-1578) were cultured in DMEM with 10% FBS, 4 mM glutamine, 1% antimycotic-antibiotic.

Human hepatocellular carcinoma cell cultures. SNU-387 cells (ATCC Cat# CRL-2237) and HepG2 (ATCC Cat# HB-8065) were cultured in RPMI with 10% FBS, 2mM glutamine, 1% antimycotic-antibiotic, 1% non-essential amino acids. HuH7 cells were cultured in DMEM with 10% FBS, 2 mM glutamine, 1% antimycotic-antibiotic, 1% non-essential amino acids.

Human keratinocyte cell cultures. HaCaT cells (4) (kind gift from A. Oro, Stanford) were cultured in DMEM with 10% FBS, 2 mM glutamine, 1% penicillin/streptomycin.

Hamster ovary cell cultures. CHO-K1 cells (ATCC Cat# CCL-61) were cultured in Ham's F12 with 10% FBS, 1% penicillin/streptomycin.

Rhesus macaque kidney cell cultures. FRhK-4 cells (ATCC Cat# CRL-1688) were cultured in DMEM with 10% FBS, 2mM glutamine, 1% penicillin/streptomycin.

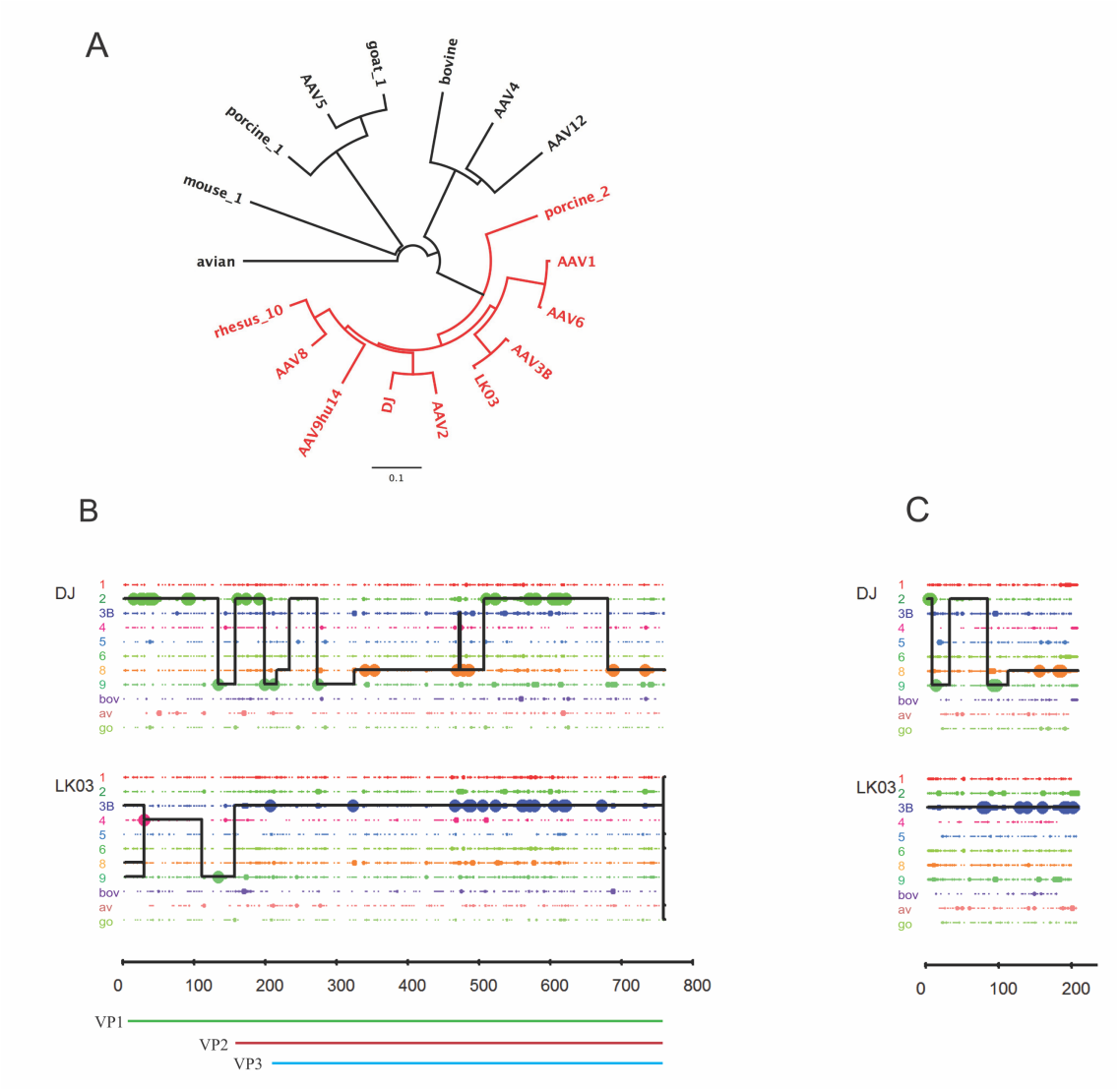
Mouse fibroblast cell cultures. Primary mouse embryonic fibroblasts derived at E14 were cultured in DMEM with 10% FBS, 2 mM glutamine, 1% non-essential amino acids, 1% antimycotic-antibiotic, 55 uM β-Mercaptoethanol.

Liver immunohistochemistry. Slides were fixed in methanol for 1 min at -20°C and air dried at RT. All following steps were done at RT except for notification. Slides were washed in PBT (PBS + 0.1% Tween20) for 3x5 min, permeabilized in 0.3% Triton-X-100 in PBS for 1x10 min, and washed in PBT for 2x5 min. Blocking was performed in PBT + 10% normal donkey serum (Cat. no. ab7475; Abcam) for 1 hr in a humidified chamber. A rabbit anti-human FAH antibody (Cat. no. HPA041370; Sigma) was added in PBT at 1:200 and incubated overnight at 4°C. Post-staining wash was conducted in PBT for 3x5 min. A secondary antibody for donkey anti-rabbit Alexa Fluor 488 IgG antibody (Cat. no. A-21206; ThermoFisher) was added in PBT (1:500, with DAPI at 80 ng/mL) and incubated at RT for 1 hr. Slides were washed in PBT for 3x5 min and PBS for 1x5 min, followed by mounting with ProLong Gold Antifade Reagent (Cat. no. 9071S; Cell Signaling). Antibody validity controls included secondary-only staining and demonstration on positive control human liver tissue sections (Cat. no. HF-314; Zyagen) and negative control untreated mouse liver sections (Cat. no. MF-314-C57; Zyagen). Imaging was performed on an inverted Zeiss laser scanning confocal microscope (LSM 880) by a 20× objective with the Zen Pro software. AAV-RFP signals were scanned and captured directly. Quantification of human hepatocyte transduction was done using the Volocity software (v6.3) and confirmed with counts by eye. Briefly, six to nine different areas of interests across different lobes from different sections of each liver sample were scanned, counting ~1,000 cells on average per section. Areas with roughly 30%-80% of FAH-positive staining were chosen for

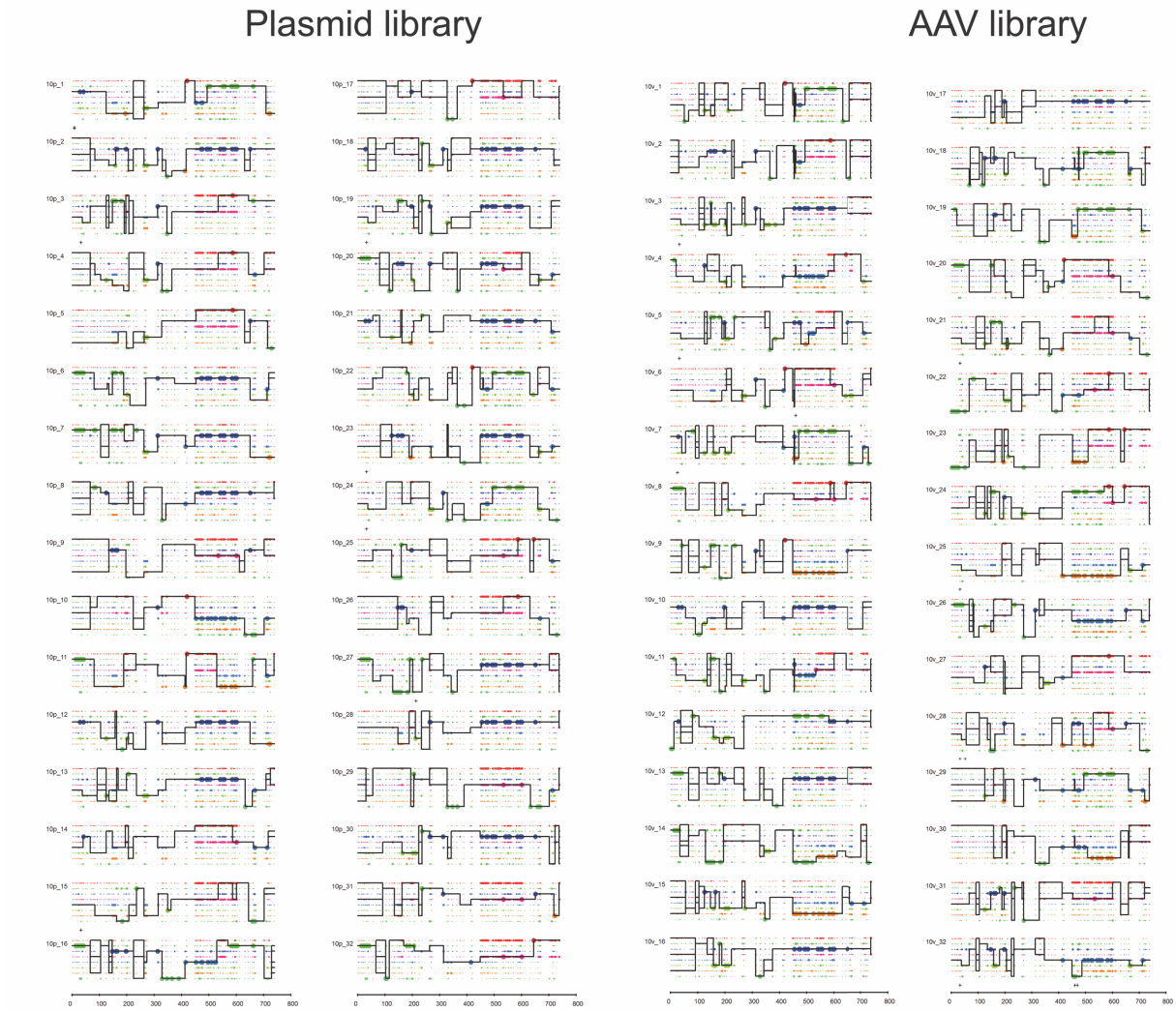
analysis. The percentage of human hepatocytes per scanned area was presented as the number of cells with Alexa Fluor 488 signals (FAH-positive) divided by the number of cells with DAPI signals (total number of cells). Total transduction efficiency was calculated by dividing the number of cells with RFP signals (AAV-positive) by the number of cells with DAPI signals. The overlap of these two numbers represented the transduction efficiency for human hepatocytes for different AAV serotypes.

Supplemental references:

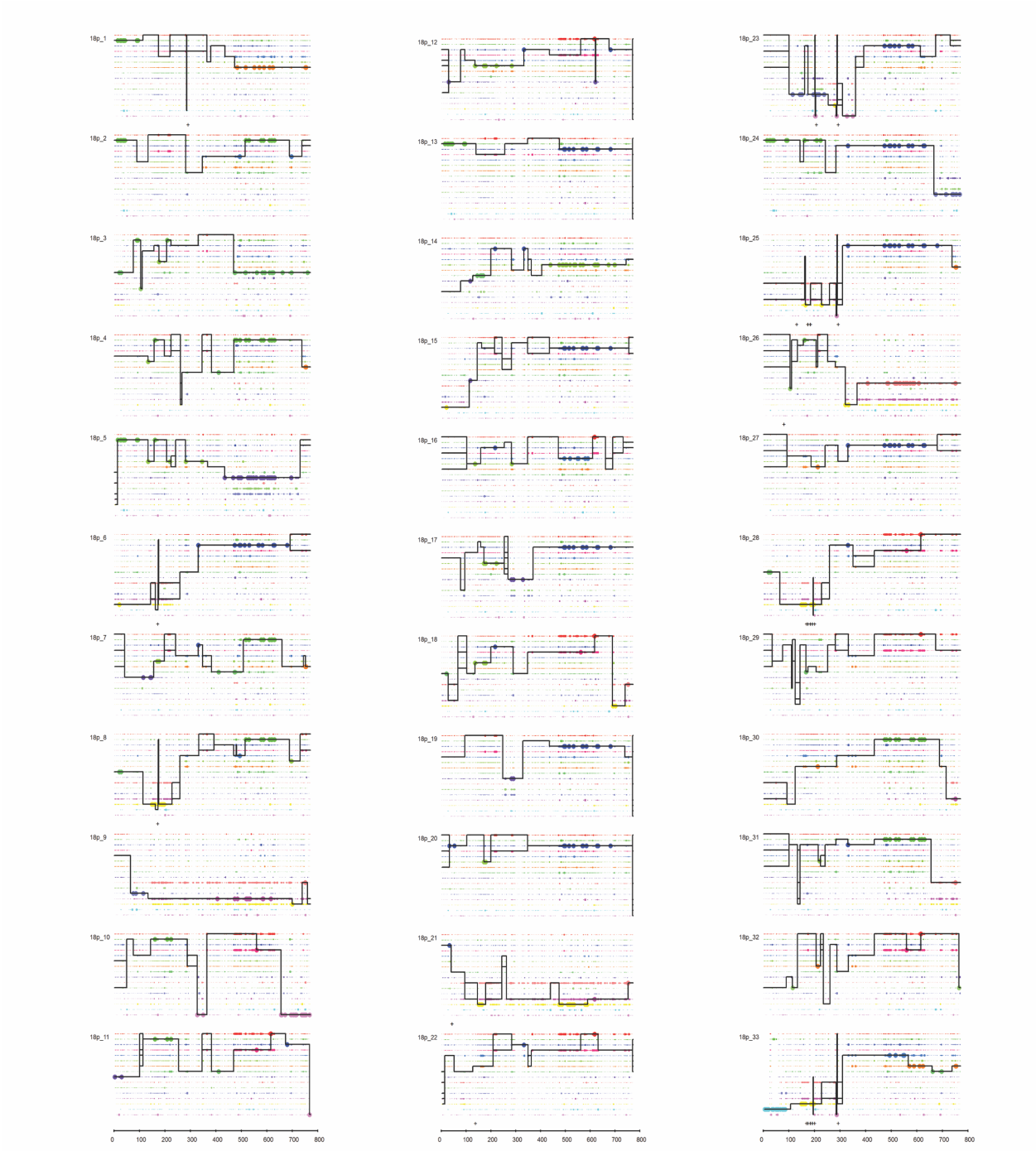
1. Stemmer WP. DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. *Proc Natl Acad Sci U S A.* 1994;91(22):10747-51.
2. Nair GG, et al. Recapitulating endocrine cell clustering in culture promotes maturation of human stem-cell-derived beta cells. *Nat Cell Biol.* 2019;21(2):263-74.
3. Charville GW, et al. Ex Vivo Expansion and In Vivo Self-Renewal of Human Muscle Stem Cells. *Stem Cell Reports.* 2015;5(4):621-32.
4. Boukamp P, et al. Normal keratinization in a spontaneously immortalized aneuploid human keratinocyte cell line. *J Cell Biol.* 1988;106(3):761-71.
5. Mays LE, et al. Mapping the structural determinants responsible for enhanced T cell activation to the immunogenic adeno-associated virus capsid from isolate rhesus 32.33. *J Virol.* 2013;87(17):9473-85.



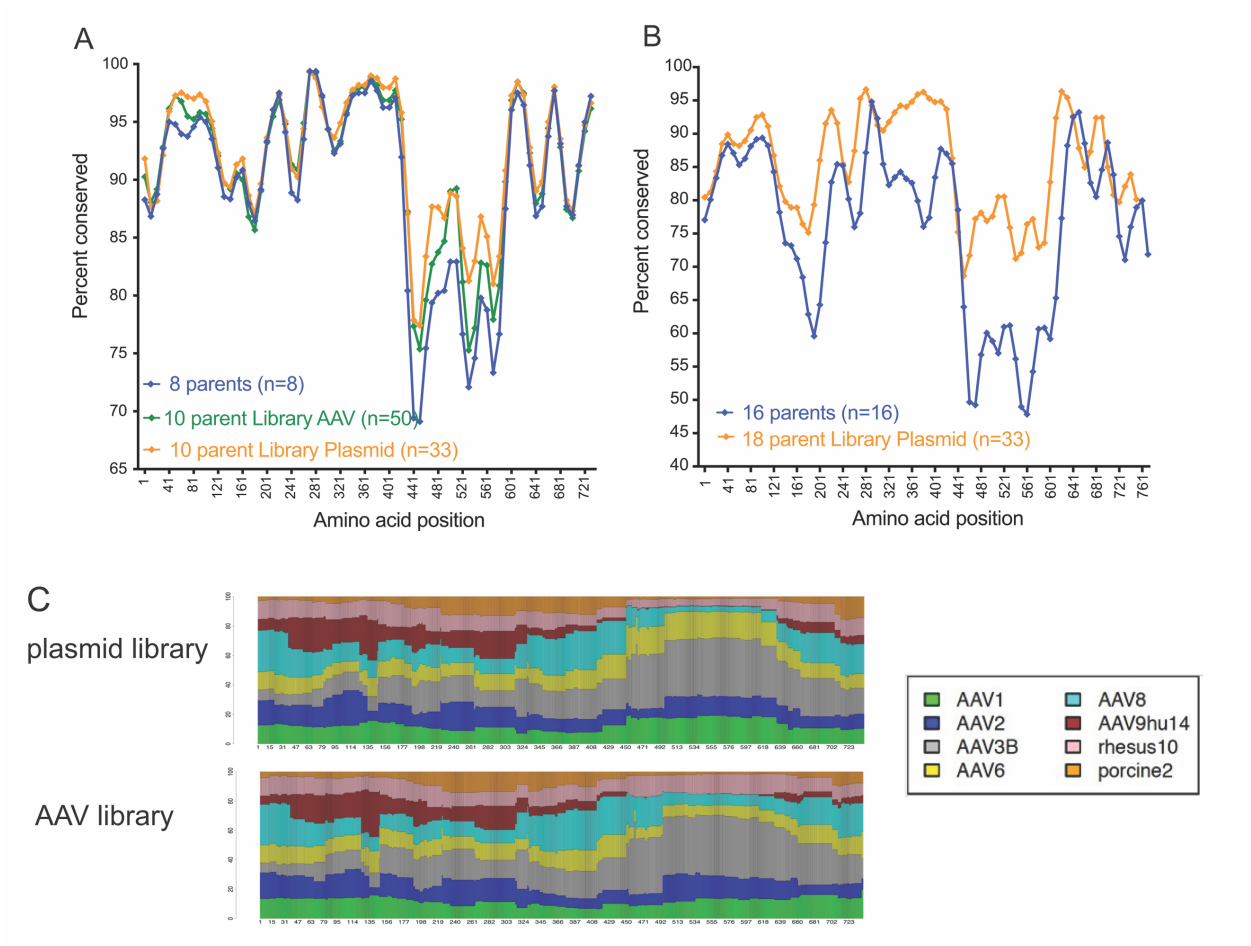
Supplemental Figure 1: Parental capsids used for library generation. (A) Phylogenetic relationship of the 18 parental capsids on the amino acid level. The parental capsids used for generation of the 10 parent library are shown in red. The neighbor-joining tree was constructed using Genious 6.0.6. **(B)** Crossover analysis of chimeric capsids DJ and LK03 with parental contributions indicated. Large dots represent 100% parental match (i.e. the position in question matches only one parent) and small dots represent more than one parental match (i.e. the position matches more than one parent) at each position. The solid line for each chimera represents the library parents identified within the sequence between crossovers. For simple visualization, the same parental capsid sequences are shown for both chimeras although they were derived from different libraries with different parental compositions. AAV-DJ was derived from a library that contained AAV2, AAV4, AAV5, AAV8, AAV9hu14, AAV-bovine, AAV-avian, and AAV-goat1 capsid sequences, AAV- LK03 was derived from a library that contained AAV1, AAV2, AAV3B, AAV4, AAV5, AAV6, AAV8, AAV9hu14, AAV-bovine, AAV-avian, and AAV-goat1 sequences. **(C)** Crossover analysis of AAP for AAV-DJ and AAV- LK03 with parental contributions indicated.



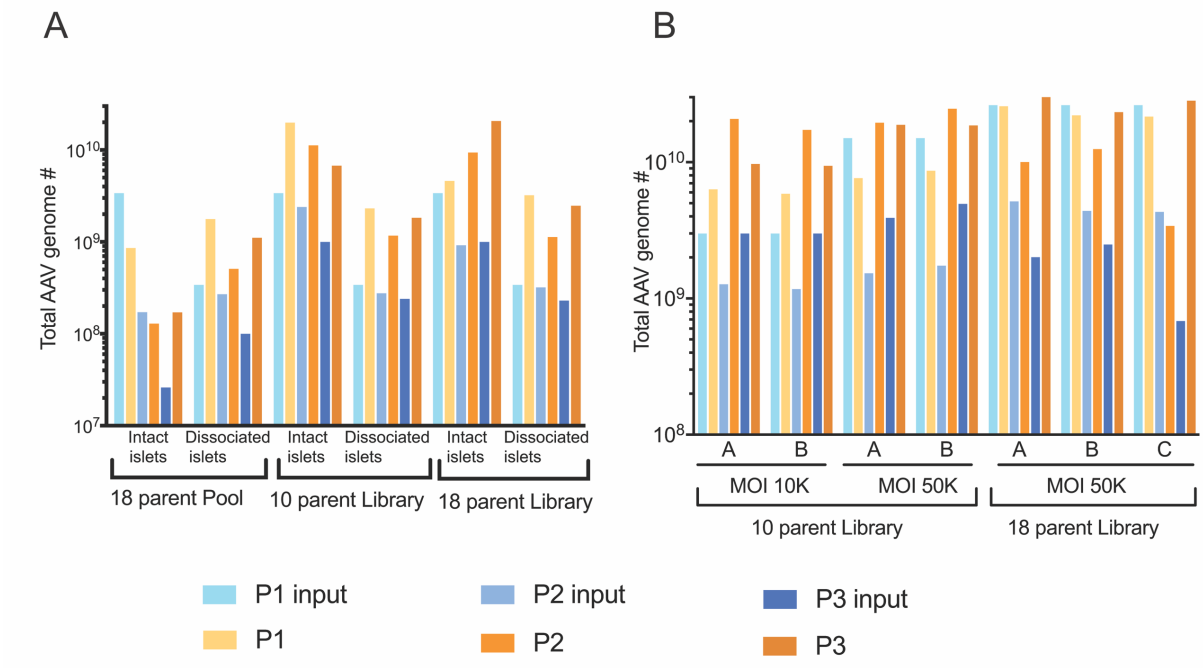
Supplemental Figure 2: Analysis of the 10 parent library by Sanger sequencing. Crossover analysis of amino acid sequences of several shuffled capsids derived from the plasmid library (left) as well as the AAV library (right). For consistency LK03 and DJ are not shown as individual parents. Positions that are marked with a cross are not de novo mutations but were derived from AAV-LK03 that contains a short stretch from AAV4. Parental capsids are shown in the following order (top to bottom): AAV1, AAV2, AAV3B, AAV6, AAV8, AAV9hu14, AAV-rhesus10, AAV-porcine2.



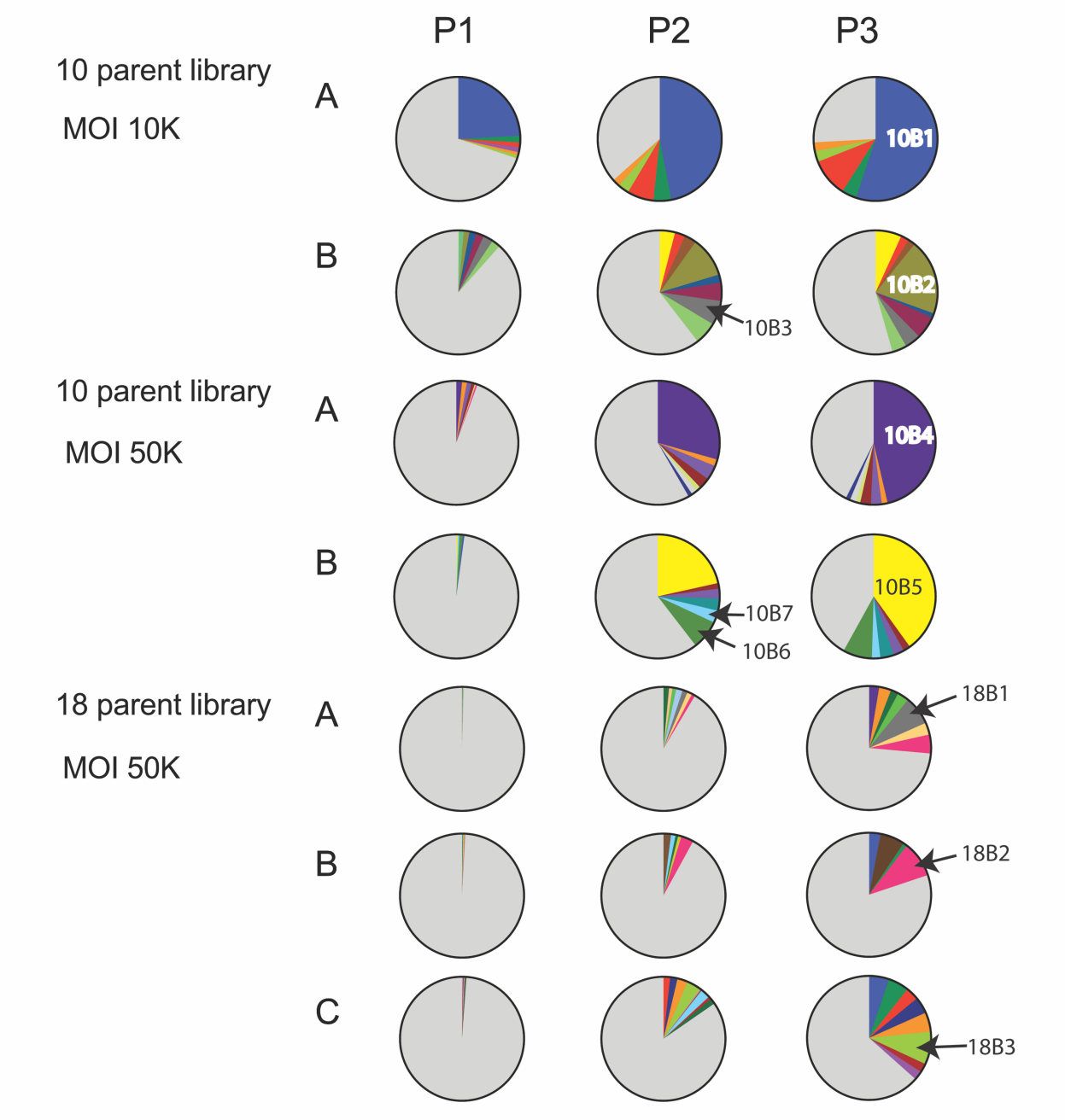
Supplemental Figure 3: Analysis of the 18 parent library by Sanger sequencing. Crossover analysis of amino acid sequences of several shuffled capsids derived from the plasmid library. Parental capsids are shown in the following order (top to bottom): AAV1, AAV2, AAV3B, AAV6, AAV8, AAV9hu14, AAV-rhesus10, AAV-porcine2, AAV4, AAV5, AAV12, AAV-bovine, AAV-goat1, AAV-porcine1, AAV-mouse1, and AAV-avian.



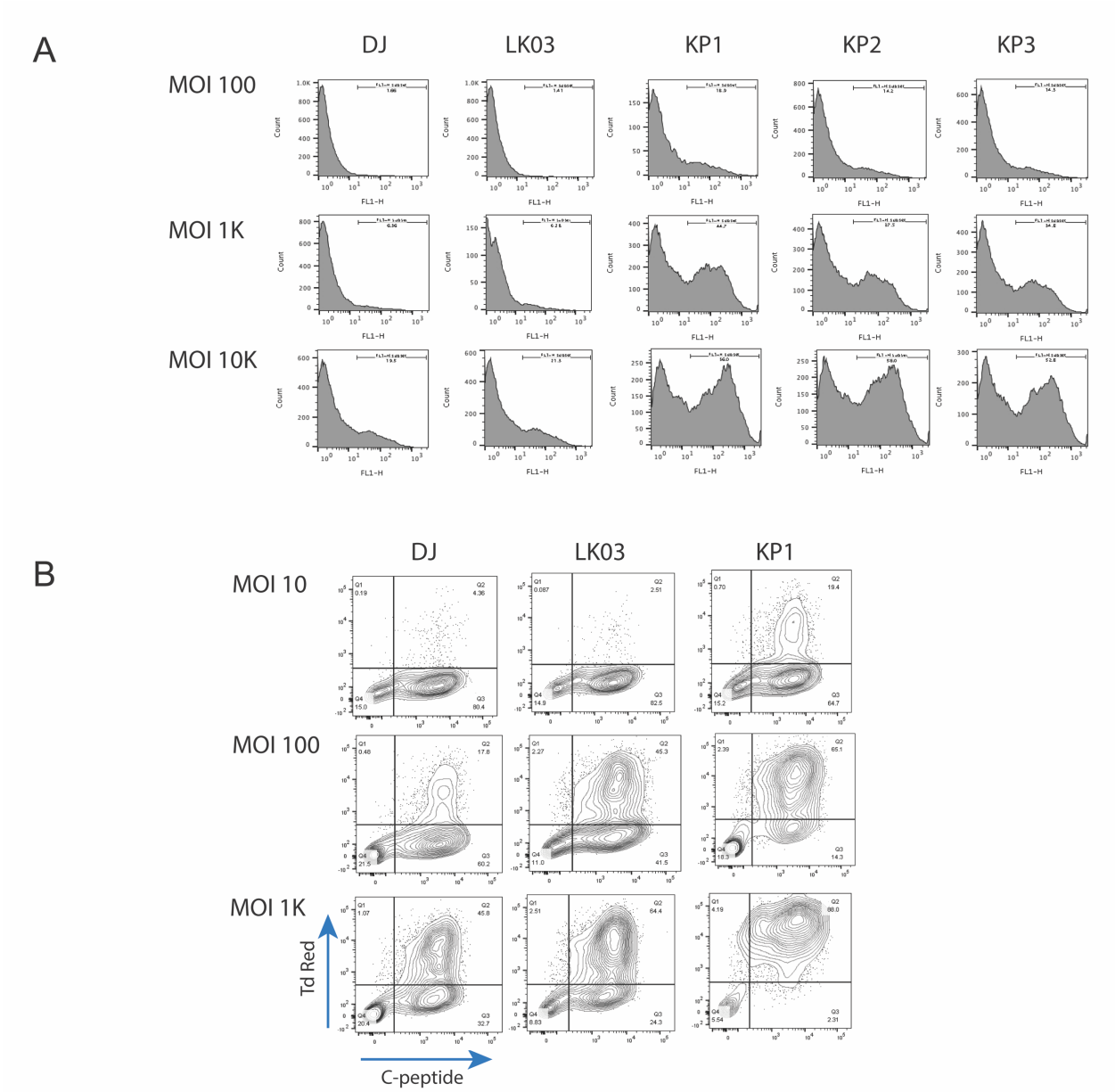
Supplemental Figure 4: Sequence analysis of the capsid shuffled libraries. (A) Conservation analysis along the capsid sequences for the 8 input parental sequences and the 10 parent plasmid as well as rAAV libraries. (B) Conservation analysis along the capsid sequences for the input 16 parental sequences and the 18 parent plasmid library. Capsids DJ and LK03 are not listed as parents since they are chimeras consisting of fragments from the parentals used (**C**) Parental contribution analysis for the 10 parent library. Chimeric capsid sequences obtained from the 10 parent library on the plasmid level and on the AAV level were analyzed by PacBio sequencing and parental contribution was assigned for each amino acid position.



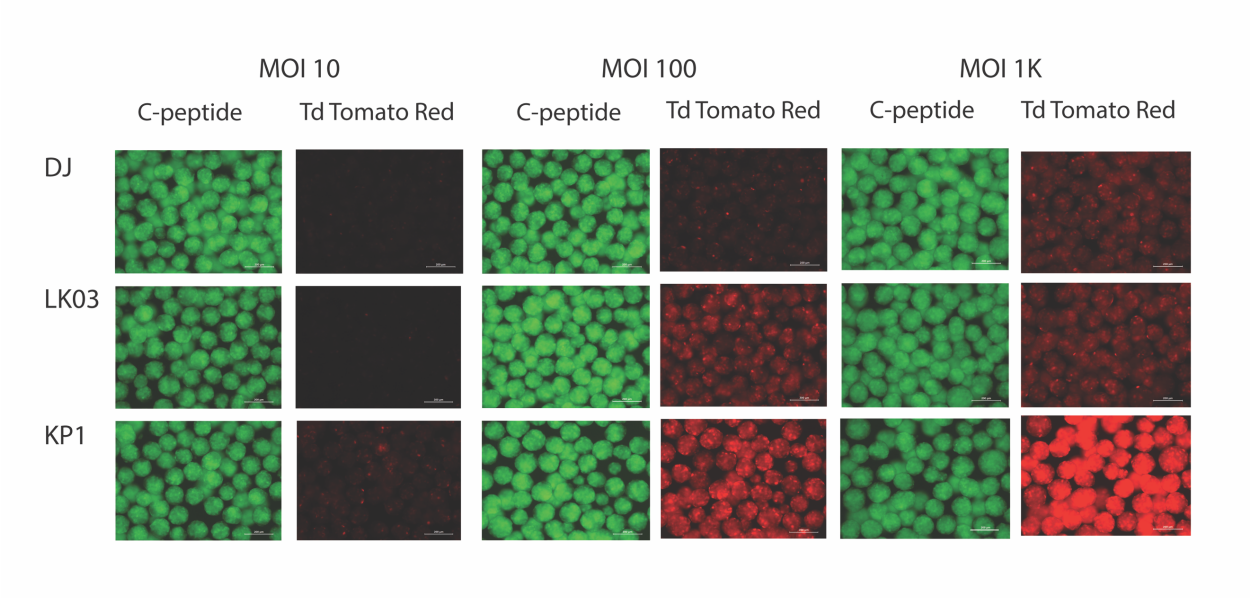
Supplemental Figure 5: Replication of AAV during 3 rounds of passaging. Results of the first screen are shown in **(A)**, results for the second screen are shown in **(B)**. For the first screen intact islets were infected with the indicated AAV preparations using a MOI of 20K, dissociated islet cells were infected using an MOI of 2K. For the second screen intact islets were infected in biological duplicates with the 10 parent library using an MOI of 10K or 50K. Infection with the 18 parent library was performed in biological triplicates using an MOI of 50K. Wildtype adenovirus 5 was used to replicate AAV and virus was harvested from the supernatant and cells after 4 days. Replication of viral species was determined by rep qPCR. For each round the total viral genome copies used for infection and the total viral genome copies retrieved are shown. Each sample was analyzed once.



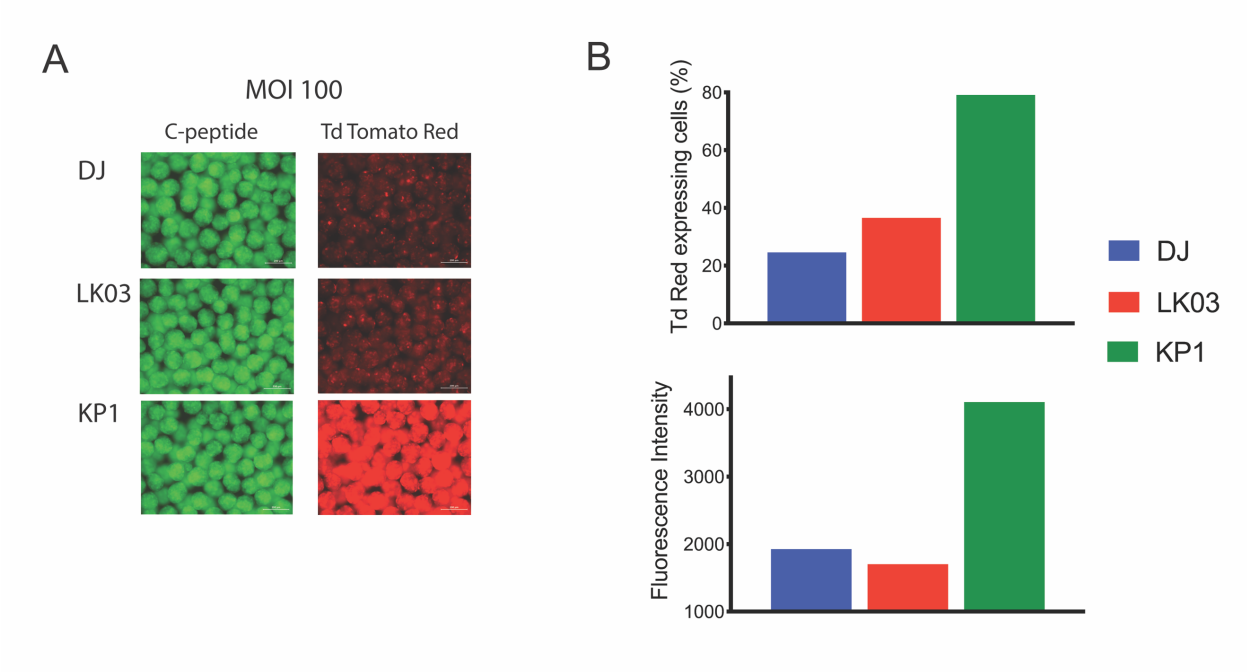
Supplemental Figure 6: Enrichment of distinct capsids during three rounds of passaging during the second library screen. Barcode sequences were amplified from viral genomes after each passage and were analyzed by high-throughput sequencing. Enriched variants are depicted in different colors while all other variants are shown in grey. Enrichment of AAV capsid variants used for vectorization is indicated (10B1, 10B2, 10B3, 10B4 10B5, 10B6, 10B7, 18B1, 18B2, 18B3).



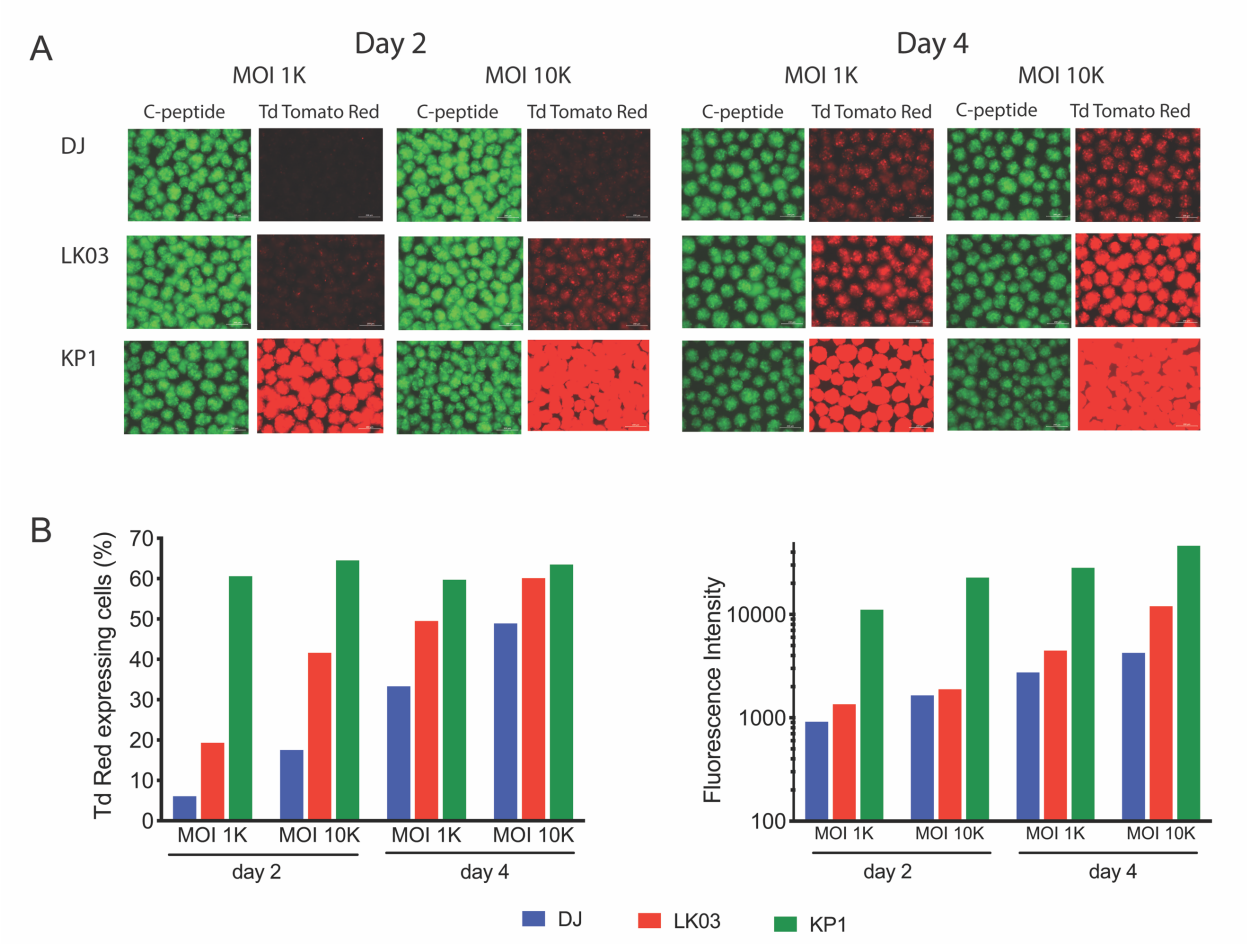
Supplemental Figure 7: Transduction efficiency of novel and parental AAV capsids on human islets and on hESC derived β-cells. (A) Dissociated islet cells were transduced with CsCl gradient purified scCAG-GFP rAAV preps generated with the two best parental capsids as well as the novel capsids that were the top transducers in the pre-screen. Three different MOIs were used for transduction. Transduction efficiency is measured as a function of the number of GFP expressing cells in conjunction with fluorescence intensity for each cell. Results of a representative experiment that was performed twice are shown. **(B)** DJ, LK03 and KP1 capsids were used to package a Tomato Red vector and hESC derived mature β-cells were transduced with the MOIs indicated. Intracellular staining for the β-cell marker C-peptide was performed on day 6 post transduction and cells were analyzed by flow cytometry. Transduction with MOI 100 was performed in a second independent experiment (see Supplemental Figure 9), transduction at the other MOIs was performed only once.



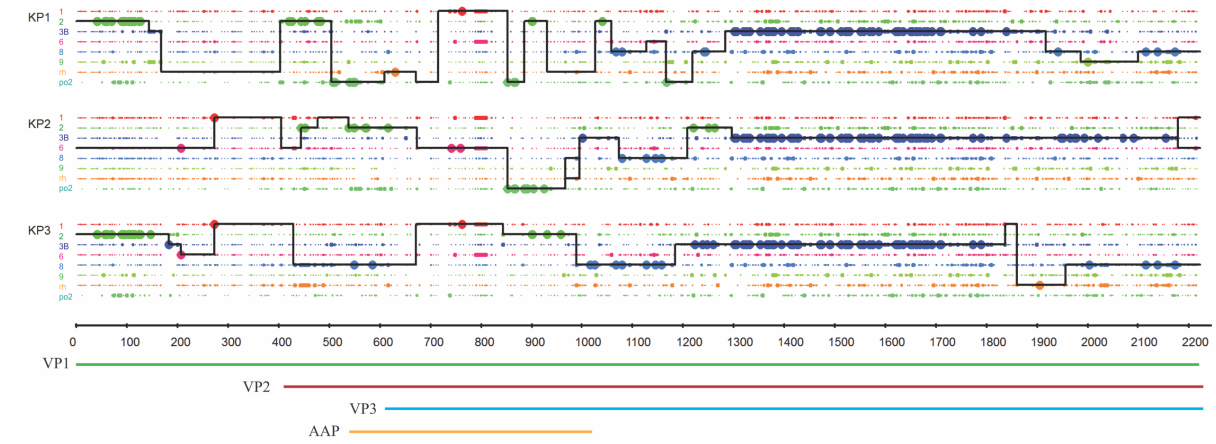
Supplemental Figure 8: Image analysis of hESC derived β-cells after rAAV transduction. DJ, LK03 and KP1 capsids were used to package a Tomato Red vector and hESC derived mature β-cells were transduced with the MOIs indicated. Intracellular staining for the β-cell marker C-peptide was performed on day 6 post transduction and cells were analyzed by live imaging. The size bar represents 200 um.



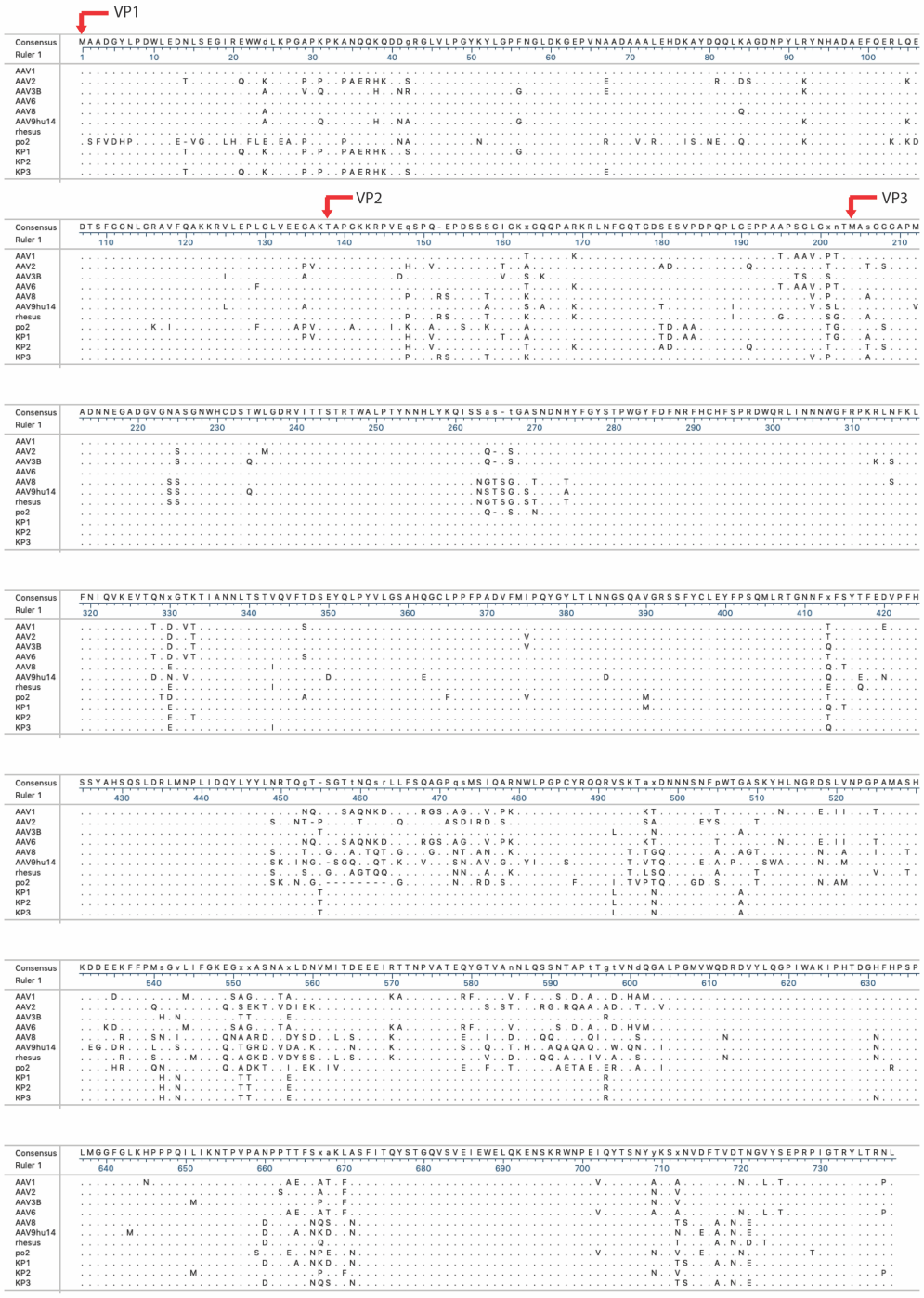
Supplemental Figure 9: Analysis of hESC derived β-cells after rAAV transduction. DJ, LK03 and KP1 capsids were used to package a Tomato Red vector and hESC derived mature β-cells were transduced with a MOI of 100. Transduction efficiency was assessed on day 6 post transduction by live imaging **(A)**. Intracellular staining for the β-cell marker C-peptide in those eBCs was performed and cells were analyzed by flow cytometry **(B)**. The size bar represents 200 um.



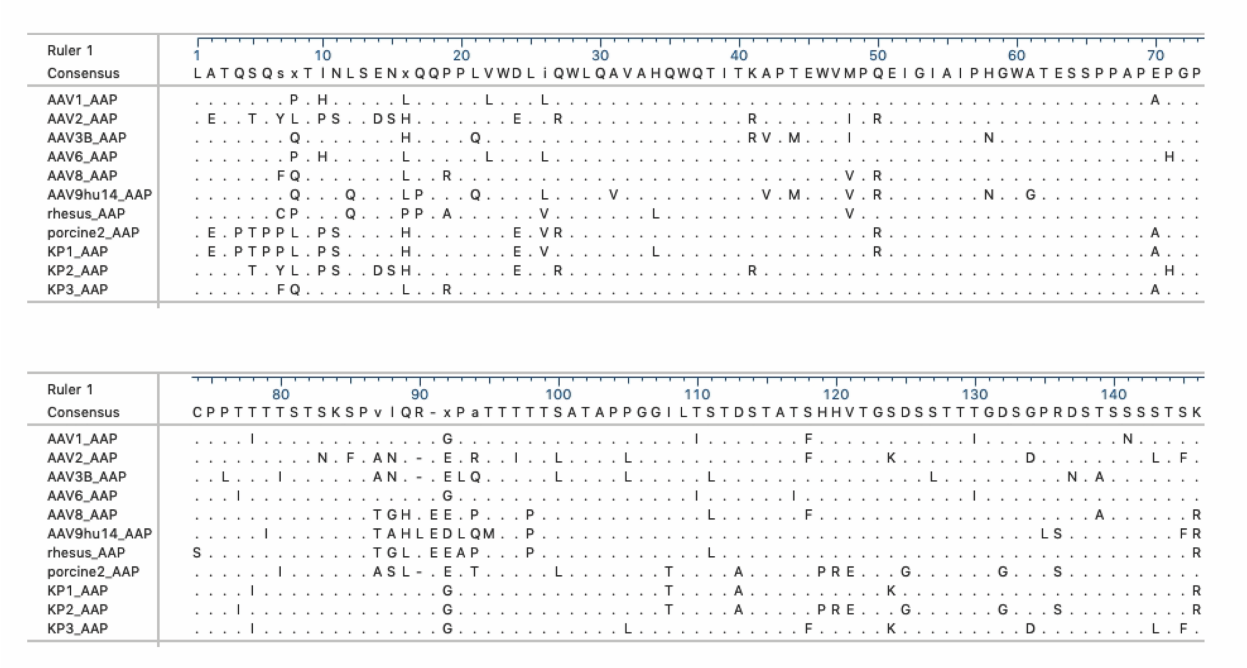
Supplemental Figure 10: Analysis of hESC derived β-like spheres after rAAV transduction. DJ, LK03 and KP1 capsids were used to package a Tomato Red vector and hESC derived immature β-like spheres were transduced with the MOIs indicated. Transduction efficiency was assessed on day 2 and on day 4 post transduction by live imaging **(A)**. Intracellular staining for the β-cell marker C-peptide in those β-like spheres was performed and cells were analyzed by flow cytometry **(B)**. The experiment was performed once. The size bar represents 200 um.



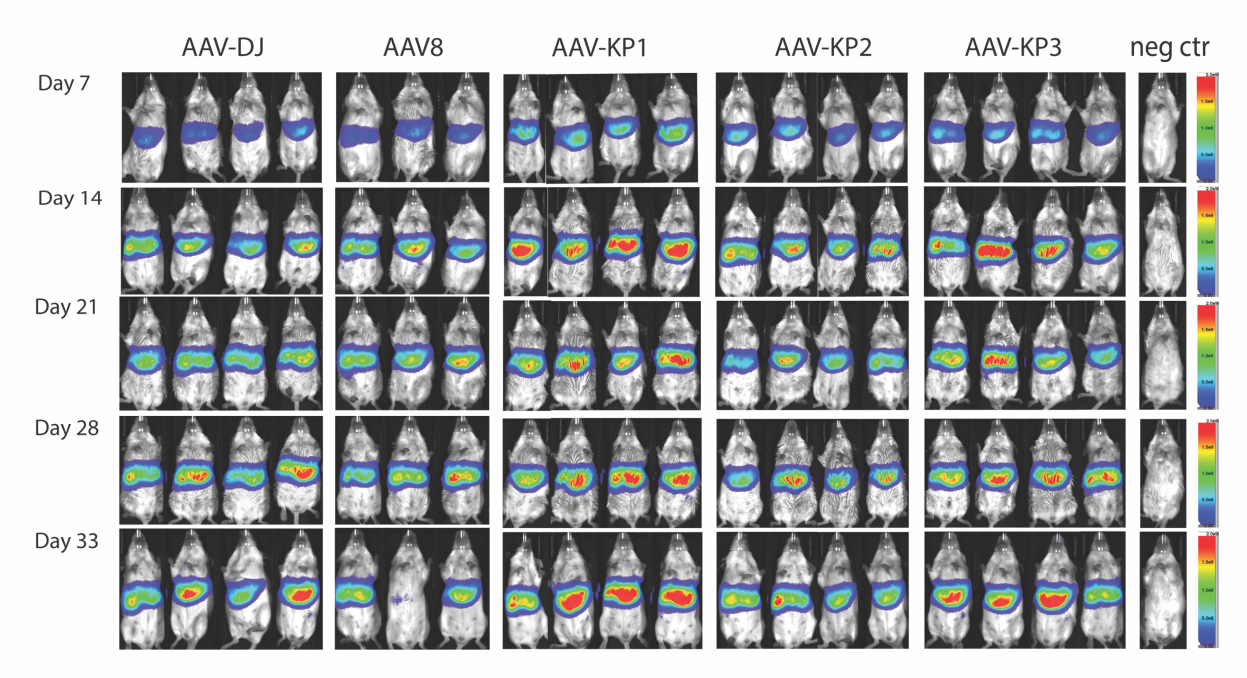
Supplemental Figure 11: Analysis of the novel capsids for parental contribution on the nucleotide level. VP1, VP2, VP3 and AAP coding sequences are indicated below the crossover analysis for KP1, KP2, and KP3 sequences.



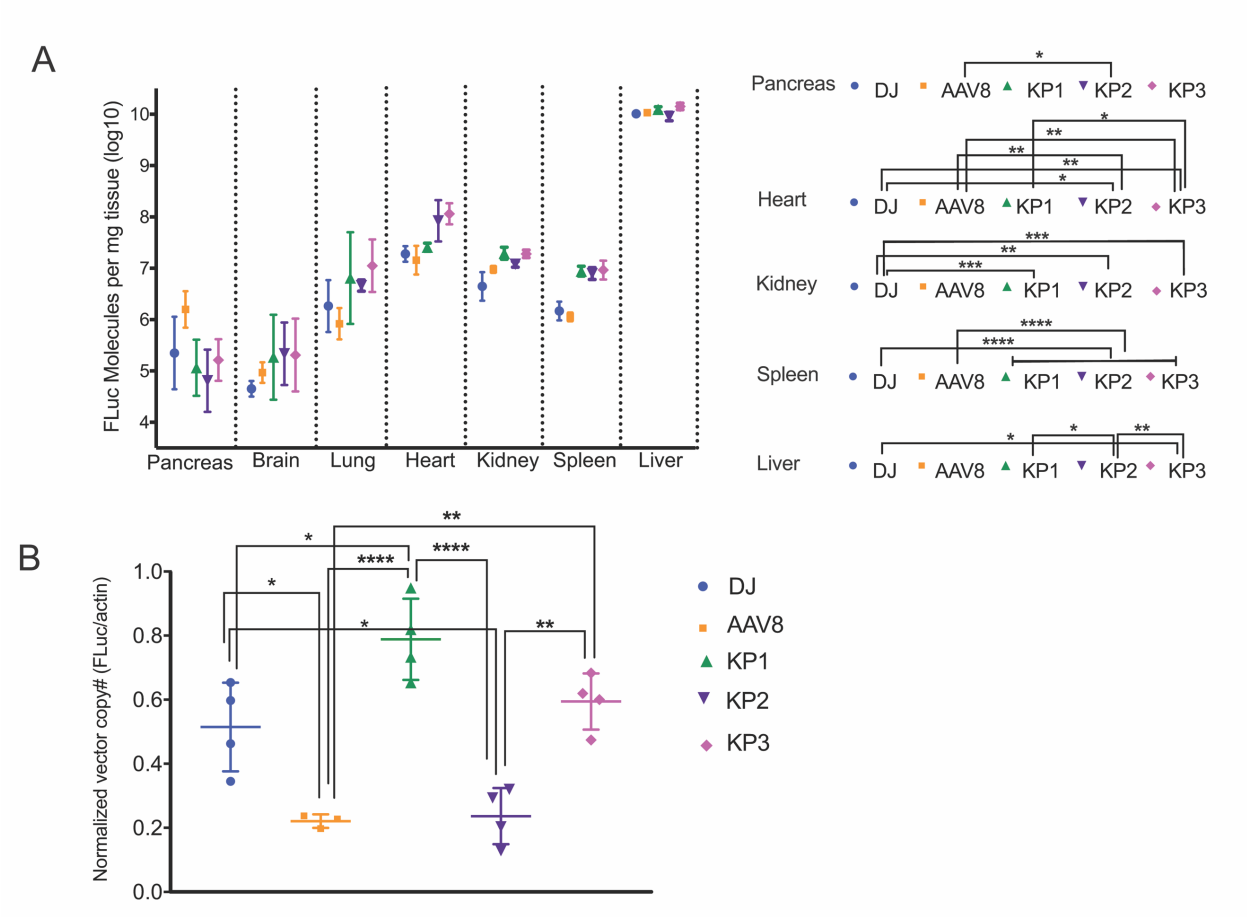
Supplemental Figure 12: Amino acid alignment of capsid sequences of the novel variants as well as the 8 parental serotypes. Only differences to the consensus sequence are shown. A dot indicates a residue that is identical to the consensus sequence, a hyphen indicates that there is no residue at this position. The start sites for VP1, VP2, and VP3 are shown.



Supplemental Figure 13: Amino acid alignment of the AAP sequences of the novel variants as well as the 8 parental serotypes. Only differences to the consensus sequence are shown. Only differences to the consensus sequence are shown. A dot indicates a residue that is identical to the consensus sequence, a hyphen indicates that there is no residue at this position.



Supplemental Figure 14: In vivo transduction efficiency of rAAVs packaged with the novel, AAV8, and DJ capsids. Balb/C SCID mice were injected via tail vein with 2E10 vg of each Firefly luciferase expressing rAAV and luciferase expression in the livers was monitored over several weeks using live imaging after i.p injection of luciferin substrate. Four animals were injected for each group with the exception of the AAV8 group which contained 3 animals. Maximum radiance was set at 2E09 with a threshold of 5E07.



Supplemental Figure 15: In vivo transduction efficiency of rAAVs packaged with the novel, AAV8, and DJ capsids. (A) 34 days post injection mice were sacrificed, and several organs were analyzed for luciferase expression ex vivo using the Promega Luciferase assay system. **(B)** Genomic DNA was isolated from mouse livers and analyzed for vector copy numbers by qPCR. Experimental values were assessed via two-way ANOVA using Tukey's multiple comparisons test. Only statistically significant differences between groups are indicated.
*: $p<0.05$, **: $p<0.01$, ***: $p<0.001$, ****: $p<0.0001$.

Supplemental Table 1: Composition of the 18 parental pool based on Illumina HTS of the barcode sequences. Data are given as percentage of total full-length reads with absolute read numbers in parenthesis.

AAV1	7.12 (68,241)	AAV8	6.16 (59,023)	porcine 1	5.82 (55,711)
AAV2	8.34 (79,931)	AAV9hu14	5.44 (52,087)	porcine 2	10.47 (100,288)
AAV3B	5.08 (48,660)	AAV12	0.35 (3,307)	mouse 1	5.29 (50,690)
AAV4	0.006 (62)	DJ	8.24 (78,973)	goat 1	1.58 (15,103)
AAV5	10.28 (98,474)	LK03	6.48 (62,091)	avian	10.05 (96,275)
AAV6	1.08 (10,320)	rhesus 10	8.22 (78,743)	bovine	0.00001 (2)

Supplemental Table 2: Barcode NGS data of the BC library, and the capsid shuffled libraries at the plasmid as well as the AAV level. The top identical sequences are shown, the number of times each sequence is represented within the sequence reads is shown in parenthesis.

Sample	BC library	10 parent plasmid library	10 parent AAV library	18 parent plasmid library	18 parent AAV library
Total reads (#)	1,249,614	1,256,634	1,141,870	1,804,277	1,000,494
Unique reads (#)	1,164,408	1,110,174	866,601	1,598,929	798,927
Unique reads (%)	93.18	88.35	75.89	88.62	79.85
Identical BC (# of repeats)	1 (6x)	7 (7x)	1 (287x)	1 (8x)	1 (232x)
Identical BC (# of repeats)	20 (5x)	30 (6x)	1 (195x)	8 (7x)	1 (177x)
Identical BC (# of repeats)	342 (4x)	220 (5x)	1 (182x)	63 (6x)	1 (173x)
Identical BC (# of repeats)	4,844 (3x)	1,809 (4x)	1 (177x)	365 (5x)	1 (141x)

Supplemental Table 3: Shared amino acid residues of the novel variants. Only residues that are different from AAV3B and are shared among at least two of the variants are shown. Amino acids shared among all three variant capsids are highlighted in bold. A blank field indicates a residue identical to that of AAV3B. A missing amino acid is indicated (-). Numbering is according to the alignment in Figure S10. Hypervariable regions (HVR) in VP3 are labeled according to (5). Surface exposed residues on the VP3 capsid protein are marked (*).

Position	Protein / HVR	AAV3B	KP1	KP2	KP3
14	VP1	N	T		T
21	VP1	E	Q		Q
24	VP1	A	K	D	K
29	VP1	V	P	A	P
31	VP1	Q	P	K	P
34	VP1	A	P		P
35	VP1	N	A		A
36	VP1	Q	E		E
37	VP1	Q	R		R
39	VP1	Q	K		K
41	VP1	N	**D**	**D**	**D**
42	VP1	R	S	G	S
56	VP1	G		F	F
67	VP1	E	A	A	
92	VP1	K	**R**	**R**	**R**
125	VP1	I	**V**	**V**	**V**
135	VP1	A	P	G	G
147	VP1, VP2	D	**E**	**E**	**E**

148	VP1, VP2	Q	H	H	P
151	VP1, VP2	Q	V	V	
160	VP1, VP2	V	T	I	I
165	VP1, VP2	K	**Q**	**Q**	**Q**
181	VP1, VP2	E	D	D	
197	VP1, VP2	T	**S**	**S**	**S**
198	VP1, VP2	S	**G**	**G**	**G**
201	VP1, VP2	S	T	T	P
206	VP1, VP2, VP3	S	A	T	A
225	VP1, VP2, VP3	S	**A**	**A**	**A**
234	VP1, VP2, VP3	Q	**T**	**T**	**T**
264*	VP1, VP2, VP3 (HVR I)	Q	**A**	**A**	**A**
266*	VP1, VP2, VP3 (HVR I)	-	**T**	**T**	**T**
313	VP1, VP2, VP3	K	**R**	**R**	**R**
315	VP1, VP2, VP3	S	**N**	**N**	**N**
330*	VP1, VP2, VP3 (HVR II)	D	**E**	**E**	**E**
333*	VP1, VP2, VP3 (HVR II)	T	K		K
375	VP1, VP2, VP3	V	**I**	**I**	**I**
651	VP1, VP2, VP3	M	L		L
660*	VP1, VP2, VP3	N	D		D
666*	VP1, VP2, VP3	S	N		N
670*	VP1, VP2, VP3	F	L		L

671*	VP1, VP2, VP3	A	N		N
709*	VP1, VP2, VP3 (HVR IX)	N	Y		Y
712*	VP1, VP2, VP3 (HVR IX)	V	T		T
713*	VP1, VP2, VP3 (HVR IX)	N	S		S
717*	VP1, VP2, VP3 (HVR IX)	T	A		A
719*	VP1, VP2, VP3 (HVR IX)	D	N		N
721*	VP1, VP2, VP3 (HVR IX)	N	E		E