

# Proteomics Analysis of Co-Purifying Cellular Proteins Associated with rAAV Vectors

Biao Dong<sup>1</sup>\*, Xunbao Duan<sup>1</sup>\*, Hoi Yee Chow<sup>2</sup>, Lingxia Chen<sup>1</sup>, Hui Lu<sup>1</sup>, Wenman Wu<sup>1</sup>, Bernd Hauck<sup>3</sup>, Fraser Wright<sup>3</sup>, Philipp Kapranov<sup>4\*</sup>, Weidong Xiao<sup>1\*</sup>

**1** Department of Microbiology and Immunology, Sol Sherry Thrombosis Research Center, Temple University, Philadelphia, Pennsylvania, United States of America, **2** Cancer Biology Program, Fox Chase Cancer Center, Philadelphia, Pennsylvania, United States of America, **3** Department of Pathology and Laboratory Medicine, University of Pennsylvania School of Medicine, and Center for Cellular and Molecular Therapeutics, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, United States of America, **4** St. Laurent Institute, Cambridge, Massachusetts, United States of America

## Abstract

Recombinant adeno-associated vectors (rAAV) are commonly purified by either chromatography or equilibrium CsCl gradient. Nevertheless, even after purification various cellular proteins often associate with rAAV vector capsids. Such co-purifying cellular proteins may raise concern about safety of gene therapy. Here we report identification and characterization of the co-purifying cellular protein in the vector preparations by using a combination of two proteomics approaches, GeLC-MS (gel electrophoresis liquid chromatography-mass spectrometry) and 2DE (two-dimensional gel electrophoresis). Most prominent bands revealed by Coomassie Blue staining were mostly similar to the AAV capsid proteins. Posttranslational modifications of capsid proteins were detected by the proteomics analysis. A total of 13 cellular proteins were identified in the rAAV vectors purified by two rounds of cesium chloride gradient centrifugation, including 9 by the GeLC-MS analysis and 4 by the 2DE analysis. Selected cellular proteins were verified by western blot. Furthermore, the cellular proteins could be consistently found associated with different AAV serotypes and carrying different transgenes. Yet, the proteins were not integral components of the viral capsid since a stringent washing procedure by column purification could remove them. These co-purified proteins in AAV vector preparations may have a role in various stages of the AAV life cycle.

**Citation:** Dong B, Duan X, Chow HY, Chen L, Lu H, et al. (2014) Proteomics Analysis of Co-Purifying Cellular Proteins Associated with rAAV Vectors. PLoS ONE 9(2): e86453. doi:10.1371/journal.pone.0086453

**Editor:** Jianming Qiu, University of Kansas Medical Center, United States of America

**Received:** November 8, 2013; **Accepted:** December 11, 2013; **Published:** February 3, 2014

**Copyright:** © 2014 Dong et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This study was supported by National Institutes of Health (NIH) grants R01HL080789 and R01HL084381. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: wxiao@temple.edu (WX); philippk08@gmail.com (PK)

† These authors contributed equally to this work.

## Introduction

Adeno-associated virus (AAV) is a nonpathogenic, replication-deficient parvovirus. It contains a single-stranded DNA genome of 4.7 kb, where two open reading frames (ORFs), Rep and Cap, are flanked by two inverted terminal repeats (ITR). Transcription initiation from internal promoters followed by splicing of the RNA transcripts generates four Rep proteins (Rep78, Rep68, Rep52 and Rep40) and three Cap proteins (VP1, VP2 and VP3) that assemble the virus particle at a ratio of 1:1:10. Up to now, more than 100 serotypes of AAV have been identified including the most studied serotype AAV2 [1].

Recombinant AAV (rAAV) has been used in clinical trials for a decade and is believed to represent one of the most promising vectors for gene therapy [2–5]. Currently, triple plasmid transfection is the most popular method to make rAAV vectors in which three individual plasmids supply the ITR-flanked transgene, the Ad-helper genes, and Rep/Cap genes [6]. Afterwards, the rAAV vectors can be purified by two successive rounds of CsCl gradient centrifugation to levels sufficient for most pre-clinical studies [7,8]. To reach higher purity, the vectors can be further refined by chromatography techniques [8,9]. Contam-

inants in the vector preparations, such as plasmid DNA, replication-competent AAV, and co-purifying cellular proteins, could decrease the efficiency of gene therapy and raise concerns about its safety [10–12].

On the other hand, knowledge of the co-purifying proteins can also improve our knowledge of cellular factors involved in viral biology and potentially lead to better vectors and production systems. Protein identification has benefited from advancements in sensitivity and accuracy of mass spectrometry (MS) that led to its broad use in proteomics [13–15]. GeLC-MS and 2DE are two popular approaches in proteomics due to the robust nature of the conventional SDS-PAGE and the resolving power of LC-MS/MS. To take full advantage of each method, we used a combination of both to detect cellular proteins in the rAAV vector preparations.

## Results

### Identification of proteins associated with rAAV vector by GeLC-MS

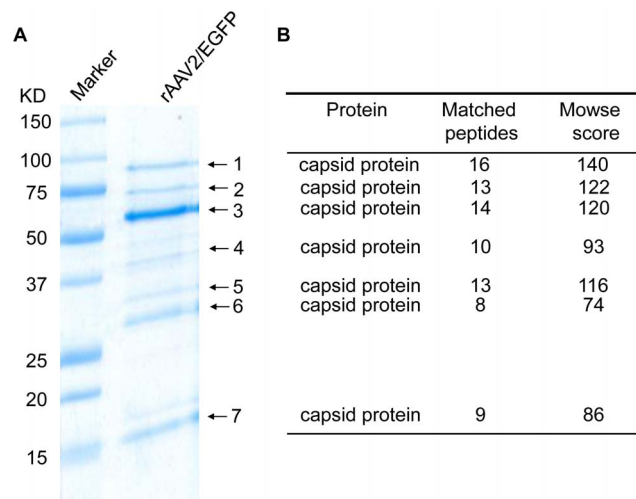
The AAV vector, AAV2-dsEGFP, was produced by triple plasmid transfection and purified by two-rounds of CsCl gradient centrifugation [16]. After 1D SDS/PAGE gel electrophoresis,

Coomassie Blue staining of the gel revealed major protein components of the preparation (Fig 1). The AAV capsid proteins, VP1, VP2 and VP3, were detected as three major bands (band 1–3). In addition, several other bands with lower molecular weights than VP3, such as bands 4–7, were also observed. To search the identity of these proteins, bands 4–7 were excised out individually and analyzed by MALDI-TOF. A PMF database search suggested that all these proteins were capsid-related proteins with high confidence. Figure S1 shows the protein fingerprint of the band 7 as an example of MALDI-TOF identification. Of note, posttranslational modifications of protein, carbamidomethyl of cysteine and oxidation of methionine, were detected in two fragments in this analysis (Figure S1).

For the GeLC-MS analysis, the capsid protein components were resolved on 1D SDS/PAGE gel, stained with Coomassie Blue, and sliced to three parts according to the molecular weight, namely, > 62 kD, 30–62 kD and <30 kD (Fig. 1). The proteins in the sliced gel fragments were then processed for MS/MS identification. Again, the major bands are AAV capsid related. Figure S2 shows the protein fingerprint of the band 7 as an example of MS/MS identification. In addition to viral capsid proteins, nine cellular proteins were detected and identified, which were showed in Table 1. Specifically, nucleolin, protein SET and splicing factor arginine/serine-rich 1 were identified in two fractions with different molecular weights, suggesting the alternatively spliced and/or modified forms of these proteins have strong affinity for the AAV capsids.

### Identification of proteins associated with rAAV vector by 2-DE

As a complementary approach, 2DE was used to identify vector-associated proteins potentially missed by GeLC-MS. Figure 2 shows an example of protein distribution in a range of isoelectric point (IP) 3–10. The marked protein spots were extracted and processed for MS analysis. Seven different proteins were identified, including four cellular proteins that were missed



**Figure 1. Identification of cellular proteins co-purified with recombinant AAV vector.** A, 1DE-SDS PAGE of recombinant AAV vector. The seven marked bands were excised and processed for MALDI-TOF analysis. B, All seven bands were identified as capsid-related proteins by the sequence homology. The Mowse score and the number of peptides matched were from the matching results against capsid protein VP1.

doi:10.1371/journal.pone.0086453.g001

by GeLC-MS (Table 2). Interestingly, distribution of each capsid protein as a series of distinct spots with similar molecular weights but different isoelectric points suggested presence of post-translational modifications in these proteins.

### Confirmation of the proteins identified in rAAV vectors

Western blotting was used to confirm the identities of the co-purified proteins in the rAAV vector with 293 cell lysates as the positive control. All four selected proteins detected in the AAV vector preparation could be detected in the 293 cell lysates, suggesting these proteins came from 293 cells during rAAV purification (Fig 3).

During rAAV purification, empty and full virus particles are separated according to their buoyant densities, 1.32 g/cm<sup>3</sup> and 1.41 g/cm<sup>3</sup>, respectively. To test whether these co-purifying proteins distinguish between the types of viral particles, a range of different fractions during purification were used for western blotting by using antibodies against SET as an example. Equal amount of virus particles from each fraction were loaded in each lane as confirmed by silver staining of capsid proteins (Fig 4A). The results showed that SET was detected in the fractions containing full particles, but not in the fraction containing empty particles (Fig 4B).

To find other factors affecting the co-purifying cellular proteins of AAV vector, rAAV vectors prepared from different virus serotypes, carrying different transgenes and produced by different purification schemes were tested. Amount of protein from all vectors was confirmed by silver staining (Fig 5A) before the western blotting (Fig 5B). Three findings were revealed by the western blotting with the anti-SET antibody. First, the presence of this co-purified protein is serotype independent. The vector serotypes included the most popular ones 2, 5, 6, 8 and 9 and the protein could be detected in all of them. Second, its presence was transgene independent: the protein could be identified in the AAV vectors containing three different transgenes, EGFP, Cre and FIX. Third, the presence of the protein was dependent on the vector purification procedures. Vectors highly purified by chromatography (AAV2-EGFP/C and AAV2-FIX/C) did not contain the protein, unlike the same kind of vectors prepared by the CsCl gradient purification (Fig 5).

### Discussion

The purity is a major criterion for an AAV vector preparation used for human gene therapy in clinical applications. It is also conceivable that many cellular proteins take part in the AAV life cycle. In this study, we use two highly sensitive mass spectrometry approaches to identify viral and cellular proteins that are associated in the rAAV vector purified by CsCl gradient centrifugation. By using GeLC-MS and 2DE, we demonstrated for the first time that major bands smaller than VP3 are polypeptides which share similar protein fingerprints with the AAV capsid proteins. It suggests that those AAV related capsid fragments have a high affinity to the capsid proteins (Figure 1) and may contribute to the immunogenicity but not transgene expression.

Interestingly, a combination of GeLC-MS and 2DE identified thirteen cellular proteins (Table 1 and 2), including nucleolin [17,18] and nucleophosmin [17,19] which have been reported to bind to the capsid during AAV packaging or infection process. However, the remaining 11 appear to represent novel cellular AAV capsid-binding proteins as evidenced by one such protein SET. The latter could be consistently detected in multiple batches of rAAV vectors containing different transgenes and based on

**Table 1.** List of cellular proteins identified using GeLC-MS analysis of adeno-associated virus vectors.

Molecular Weight	Swiss-Prot Accession #	Protein Name	Mowse Score	Peptides detected
>62 KD	P19338	Nucleolin	101	4
	P11279	Lysosome-associated membrane glycoprotein 1	82	2
	O00592	Podocalyxin-like protein 1	67	4
30–62 KD	Q01105	Protein SET	377	19
	Q07955	Splicing factor, arginine/serine-rich 1(SF2)	244	11
	P19338	Nucleolin	174	5
	P39687	Acidic leucine-rich nuclear phosphoprotein 32 family member A	90	2
	Q08170	Splicing factor, arginine/serine-rich 4	78	2
<30 KD	P06748	Nucleophosmin	185	12
	Q01105	Protein SET	149	10
	Q04837	Single-stranded DNA-binding protein, mitochondrial	144	4
	Q07955	Splicing factor, arginine/serine-rich 1(SF2)	135	5

doi:10.1371/journal.pone.0086453.t001

various serotypes unlike other cellular proteins that randomly appeared in different vector fractions. It is possible that these proteins could bind to the common domains shared by the serotype capsids.

Function wise, some of these proteins may be involved in endosomal processing and nuclear transportation of the vector since packaged vector released from cells can re-infect other cells and continue normal vector transduction pathways. It is known that the rAAV capsid structure undergoes modifications in the endosomal processing and nuclear transportation, and some kinases such as PI3 kinase family play important roles in these processes [20,21]. In this study, we identified 14-3-3 protein, which may interact with the kinases RAF1 and AKT1, two well-known downstream regulator of PI3 kinase. Of note, connection of several identified cellular proteins to protein phosphorylation suggests that phosphorylation of viral capsid proteins or cellular factors could be important for the vector packaging. Phosphorylation of capsid proteins has been reported to play a role in the infection of rAAV particles [22–25]. In addition, specific association of SET polypeptide with full particles but not the empty ones is especially interesting and further suggests a role in packaging. Additional in-depth studies will be necessary to confirm this hypothesis.

Interestingly, “dirty” vectors may actually lead to better transduction performance [26]. Identification of SET protein may provide hypothetical support for such notion. This protein inhibits acetylation of nucleosomes and has a role in enhancing DNA replication of the adenovirus genome. It has two isoforms which are complexed with adenovirus viral core proteins. In addition, SET protein is a potent inhibitor of protein phosphatase 2A and considering the importance of phosphorylation in AAV transduction performance, it is tempting to hypothesize that this co-purifying protein has an intriguing role related in the AAV life cycle.

Confirmation of four selected cellular proteins by western blot suggests that the detection of co-purifying cellular proteins by proteomics analysis could be a fast, reliable methodology to control the quality of AAV vectors. Furthermore, using protein SET as an example, the presence of the cellular proteins is independent of transgenes and AAV serotypes, but dependent on the purification procedures. In this respect, the absence of the co-purified proteins from column-purified AAV vectors suggests that

these proteins are not integral components of the AAV capsids. Also, since we do not have additional direct evidence to prove that proteins physically bind to AAV capsids, there is a possibility that these co-purifying proteins represent contaminants “co-sedimenting” or “co-migrating” with viral capsids after the regular 2 rounds of CsCl -purification. However, they do not appear as mere non-specific contaminants either because of the following reasons. First, we are not aware of any protein with a buoyant density of 1.4 g/ml representing the heavy viral particle with high DNA contents. Second, the association of the proteins with the capsids survived two rounds of CsCl-purification. This is a very strong indication that the proteins interact with the capsid. Third, at this stringency, we could identify only 13 such AAV-associated proteins out of thousands of cellular proteins. Of those, two are known and characterized cellular factors that play a role in the AAV life cycle. All this makes it likely that the novel co-purifying proteins could also function in the AAV biology and their exact roles warrant further characterization.

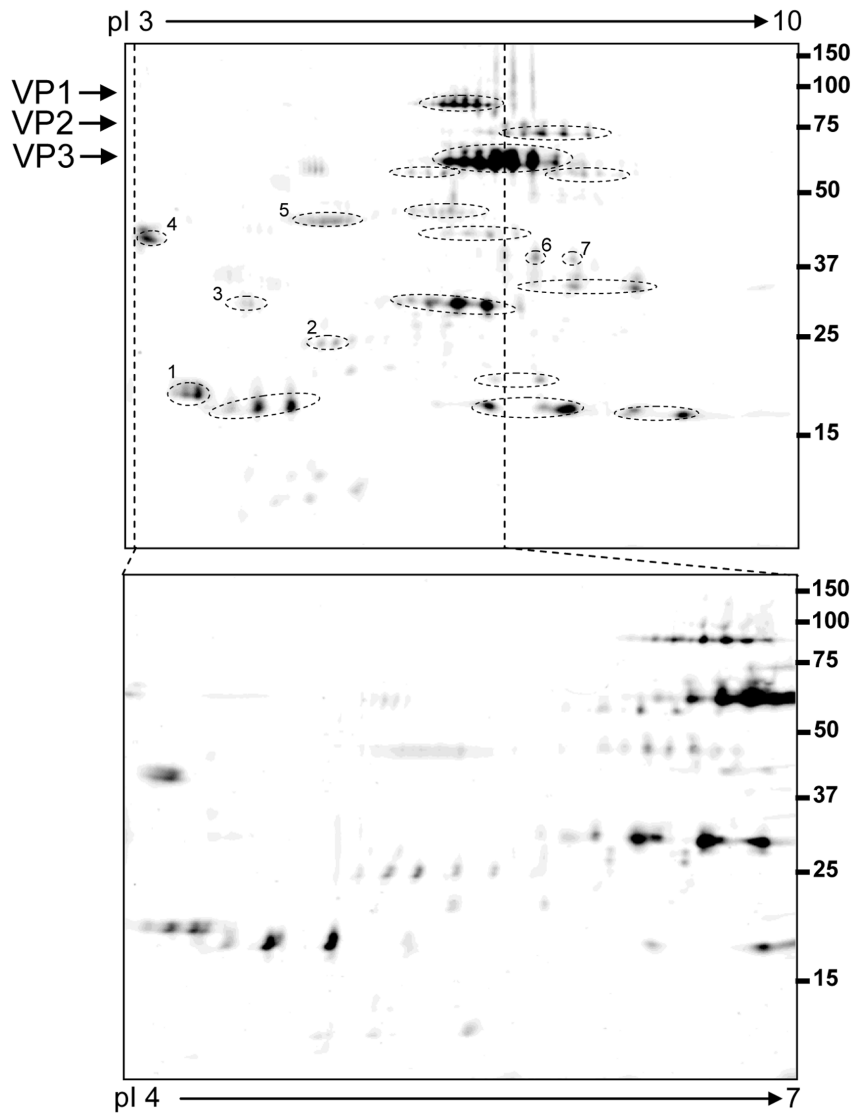
## Materials and Methods

### Production of AAV vector

A triple plasmid co-transfection method was used to produce the rAAV vectors used in this study (9). Briefly, one vector plasmid, one AAV helper plasmid, and one mini adenovirus function helper plasmid pFΔ6, were co-transfected into HEK293 cells cultured in roller bottles at a ratio of 1:1:2. The transfected cells were harvested 3 days later. rAAVs were then purified by two rounds of cesium chloride-gradient ultracentrifuge. After extensive buffer exchange against phosphate-buffered saline with 5% D-sorbitol, the peak fractions of purified virus were pooled and stored at –80°C before administration. Two rAAV vectors purified by ion exchange chromatography: AAV2-FIX were produced by GMP facility, Children’s Hospital of Philadelphia (Philadelphia, PA) and AAV2-EGFP were kindly provided by Dr. Haifeng Chen, Virovek, Inc. (Hayward, CA).

### 1-DE Separation

$10^{12}$  virus particles of AAV2-dsEGFP was solubilized in Tris-buffered saline solution (pH 8.0) containing 1% SDS. The protein concentration was then measured with Bradford-based BioRad protein assay using BSA as standard. 20 μg proteins were diluted



**Figure 2. 2DE map of recombinant AAV vector.** The top 2DE map was obtained using a wide pH range of IPG strip pH 3–10, and the bottom one using a narrow pH range of IPG strip pH 4–7. The circled and numbered spots were identified as human proteins (Table 2), the spots only circled were identified as AAV capsid proteins.  
doi:10.1371/journal.pone.0086453.g002

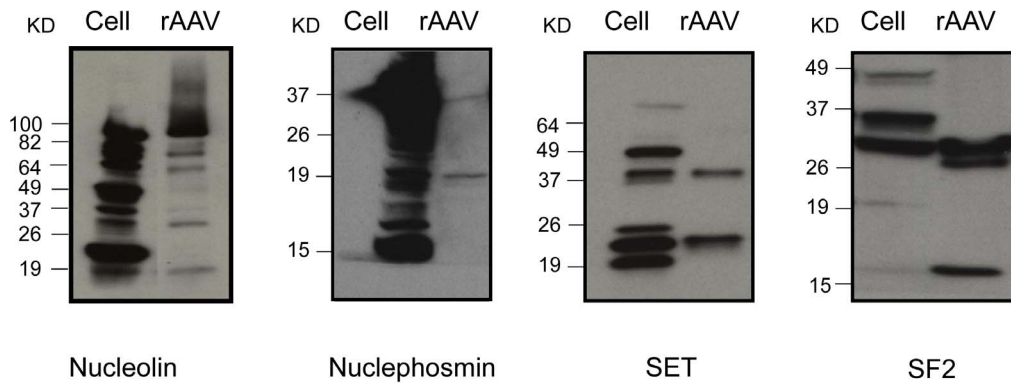
with Laemmli sample buffer (BioRad) containing 5%  $\beta$ -mercaptoethanol. The mixture was heated for 10 min at 90°C and loaded onto 10–14% polyacrylamide gel. The separation was performed

using a mini Protean II system (BioRad) at 200 V for 45 min. The resolved protein bands were visualized with SimplyBlue SafeStain (Invitrogen).

**Table 2.** List of human cellular proteins identified from 2DE analysis via nano-LC IT MS.

Spot	Swiss-Prot Accession#	Protein Name	Mowse Score	Peptides detected
1	P06748	Nucleophosmin	195	4
2	Q01105	Protein SET	379	9
3	P63104	14-3-3 protein zeta/delta	243	4
4	Q01105	Protein SET	436	11
5	P60709	Actin, cytoplasmic 1	308	8
6	P07355	Annexin A2	308	6
7	P04083	Annexin A1	159	2

doi:10.1371/journal.pone.0086453.t002



**Figure 3. Western blot analysis of co-purified proteins in 293 un-infected cell lysates and purified AAV2-dsEGFP vector particles.** 10  $\mu$ g of cell lysate protein and protein from  $1 \times 10^{10}$  viral particles were used for the western blot. doi:10.1371/journal.pone.0086453.g003

### 2-DE Gel

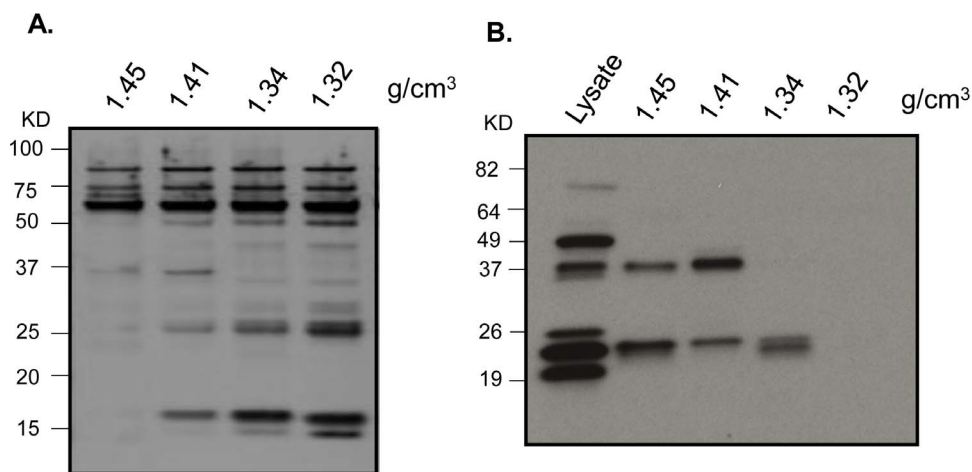
2-DE was performed as described previously [27]. The proteins solubilized in TBS containing 1% SDS were precipitated with 80% acetone at  $-20^{\circ}\text{C}$  overnight. The protein pellet after centrifugation was washed three times with 80% acetone to minimize SDS residue, and then solubilized in a buffer consisting of 7 M urea, 2 M thiourea, 40 mM Tris, and 65 mM DTT. Total 40  $\mu$ g of sample protein was diluted in 125  $\mu$ l with rehydration buffer and loaded onto an immobilized pH gradient strip (pH 3–10) by overnight passive in-gel rehydration. The rehydration buffer contains 8 M urea, 2% CHAPS, 0.2% carrier ampholytes and 10 mM DTT. Isoelectric focusing (IEF) within the strips was performed at  $20^{\circ}\text{C}$  with a MultiPhor II system (Amersham Biosciences Corp., Piscataway, NJ) using a total of 12,000 V h with a maximum of 4,000 V.

For 2nd dimensional separation, the IPG strips removed from the MultiPhor II chamber were reduced in 10 ml of an equilibration buffer (8 M urea, 30% glycerol, 2% SDS, 1% DTT, and 0.05 M pH 8.8 Tris) for 10 min, alkylated in 10 ml of a second equilibration buffer (with 2.5% iodoacetamide substituted for 1% DTT) for 15 min, and positioned against 10–14% SDS

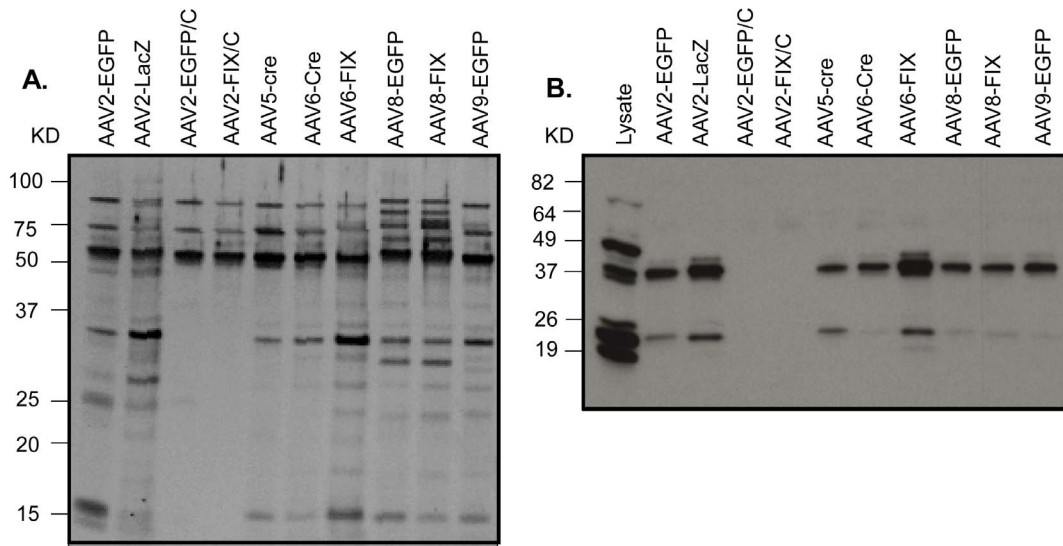
polyacrylamide gels in a BioRad Mini-PROTEAN<sup>®</sup> 3 System at 200 V for 45 min. Polyacrylamide gels were then fixed twice, each time for 30 min using 50% methanol, 7% acetic acid, balance water. Protein spots in the gels were revealed by staining with Sypro-Ruby fluorescence total protein stain. Fluorescence images of the gels were captured with FLA-5000 Fluor Imager (Fuji Photo Film Co, Ltd., Tokyo, Japan) and saved in TIFF format.

### In-Gel Trypsin Digestion

The visualized protein spots, protein bands, or gel sections were excised and then sliced into  $\sim 1$  mm gel cubes. Destaining of the sliced gel pieces was performed by two 40 min washes with 50% acetonitrile containing 50 mM ammonium bicarbonate. Following dehydration with 100% acetonitrile, 10  $\mu$ l 12.5 ng/ $\mu$ l sequencing grade trypsin (Promega, Madison, WI) in 50 mM ammonium bicarbonate was added to the protein spot excised from 2DE gel, and 30  $\mu$ l 12.5 ng/ $\mu$ l trypsin was added to gel band or gel section. The gel pieces with trypsin buffer were incubated overnight at  $37^{\circ}\text{C}$ . Resulting tryptic peptides were extracted twice with 15  $\mu$ l extraction buffer (5% formic acid, 50% acetonitrile, balance water)



**Figure 4. Detection of protein SET in different fractions of the AAV2-dsEGFP preparation.** The vector was produced by triple plasmid transfection and purified by two rounds of cesium chloride ultra-centrifugation. The fractions with different densities were collected after the second round of ultra-centrifugation. Protein from  $1 \times 10^{10}$  viral particles was resolved on 10% SDS/PAGE. The full AAV vector particles have buoyant densities in CsCl from 1.41 to 1.45  $\text{g}/\text{cm}^3$  while empty particles have the density of 1.32  $\text{g}/\text{cm}^3$ . A, silver staining; and B, western blot. doi:10.1371/journal.pone.0086453.g004



**Figure 5. The effect of AAV serotypes, transgenes and purification methods on the presence of SET proteins in the rAAV vector preparations.** Protein from  $1 \times 10^{10}$  viral particles was resolved on 10% SDS/PAGE. AAV2-EGFP/C and AAV2-FIX/C were purified by ion exchange chromatography. A, silver staining; and B, western blot. doi:10.1371/journal.pone.0086453.g005

for 20 min and the pooled extracts were processed with desalting ZipTips according to the manufacturer's instructions.

#### MALDI-TOF- MS Analysis

The desalted peptides from protein band or 2DE spot were mixed 1:1 with matrix solution (1%  $\alpha$ -cyano-4-hydroxy cinnamic acid in 50% acetonitrile and 50% 0.1% trifluoroacetic acid) and 1.0  $\mu$ l volumes were applied to wells of an AnchorChip<sup>TM</sup> sample target plate used for the Bruker Auto-flex MALDI-TOF-TOF instrument. Peptide mass fingerprints were obtained using the reflective and positive ion mode. Mass spectra were collected from the sum of 100–400 laser shots and mono-isotopic peaks were generated by FlexAnalysis<sup>TM</sup> software with signal-to-noise ratio of 2:1. Mass peak value calculations used two trypsin auto-digestion peptides with M+H values 842.509 and 2211.104 as internal standards. Protein were identified by matching the calibrated peptide mass values within NCBI protein databases using an in-house version of Mascot Server 2.2 imbedded in Bruker's Biotool software. Species viruses or homo sapiens was selected for taxonomy. Match variances allowed were a mass tolerance of 40 ppm, one missed trypsin cleavage, fixed modification of carbamidomethyl cysteine, and variable modification of methionine oxidation.

#### Nano-LC-ESI-IT MS Analysis

*Nano-LC-ESI-IT MS* was carried out as described previously [28]. The desalted peptides were dried in a vacuum centrifuge and re-solubilized in 20  $\mu$ l of 0.1% (vol/vol) trifluoroacetic acid. The tryptic peptide sample was loaded onto a 2  $\mu$ g capacity peptide trap (CapTrap; Michrom Bioresources) separated by a C18 capillary column (15 cm 75  $\mu$ m, Agilent) at 300 nl/min delivered by an Agilent 1100 LC pump. A mobile-phase gradient was run using mobile phase A (1% acetonitrile/0.1% formic acid) and B (80% acetonitrile/0.1% formic acid) from 0 to 10 min with 0–15% B followed by 10–50 min with 15–60% B and 50–55 min with 60–100% B. ESI was delivered using end-coated spray Silica tip (ID 20  $\mu$ m, tip inner ID 10  $\mu$ m, New Objective) at a spray voltage of  $-1300$  V. Using an automatic switching between MS and MS/

MS modes, MS/MS fragmentation was performed on the two most abundant ions on each spectrum using collision -induced dissociation. The complete system is fully controlled by HyStar 3.1 software. Mass spectra processing was performed with Data Analysis 3.3 and generated peak lists were submitted to an in-house Mascot server 2.2. The search was carried out against NCBI database. Mascot search parameter are set as follows: species viruses or homo sapiens as taxonomy, trypsin enzyme with maximal 1 missed cleavage, cysteine carboxymethylation as fixed modification, methionine oxidation as variable modification, and 0.30 Da mass tolerance for MS and 0.4 Da for MS/MS. A protein was considered to be positively identified if at least two different peptides with ion scores 33 or higher (95% confidence threshold) in the Mascot MS/MS ion search report.

#### Silver staining

Purified virus ( $\sim 1 \times 10^{10}$  particles) was mixed with protein loading dye and heated at 95°C for 5 minutes. The denatured samples were then separated on 10% SDS-PAGE gel. After electrophoresis, the gel was stained by Pierce<sup>®</sup> Silver Stain Kit (Thermo Scientific, Rockford, IL) according to the manufacturer's instruction.

#### Western blot

rAAV vectors ( $\sim 1 \times 10^{10}$  particles) were mixed with protein loading dye and heated at 95°C for 5 minutes. The denatured samples were then separated on 10% SDS-PAGE gel and transferred to nitrocellulose membrane (Bio-Rad, Hercules, CA). After blocking the membrane with 5% nonfat dry milk in TBST buffer (25 mM Tris·HCl at pH 8.0, 150 mM NaCl and 0.1% Tween 20), the membranes were incubated with the primary antibodies, Nucleolin (Abcam, Cambridge, MA, 1: 1,000), Nucleophosmin (Abcam, Cambridge, MA, 1:500), SET(I2PPA E-15, Santa Cruz, CA, 1:200) or SF2 (Invitrogen, Carlsbad, CA, 1:500), at 4°C overnight. The membrane was washed and incubated with the second antibodies, a HRP-conjugated sheep anti-mouse IgG antibody (Sigma, St Louis, MI, 1:2,000) or anti-rabbit (Abcam, Cambridge, MA, 1:3,000). The membrane was

developed using an enhanced chemiluminescent substrate (Amersham-Pharmacia Biotech, Piscataway, NJ).

## Supporting Information

**Figure S1 Identification of Band 7 as capsid protein by MALDI-TOF analysis.** The peptide mass fingerprinting was internally calibrated with autolytic peaks, resulting in high mass accuracy and high Mowse score. The search results showed that 8 matched peptides were from two different regions within the sequence of capsid protein VP1.  
(TIF)

**Figure S2 Identification of Band 7 as capsid protein VP1 by nano-LC-IT MS.** Ten unique peptides were identified based on sequence information with a cumulative Mowse score of 601. The left panel lists all the peptides identified including the location of each peptide. The right panel is an example of an MS/MS

spectrum for one peptide K.RWNPEIQYTSNYNK.S. The matched y ions and b ions are labeled. It should be noted that these 10 peptides are in the two different regions of the protein sequence similar to the peptides obtained from MALDI-TOF analysis.

(TIF)

## Acknowledgments

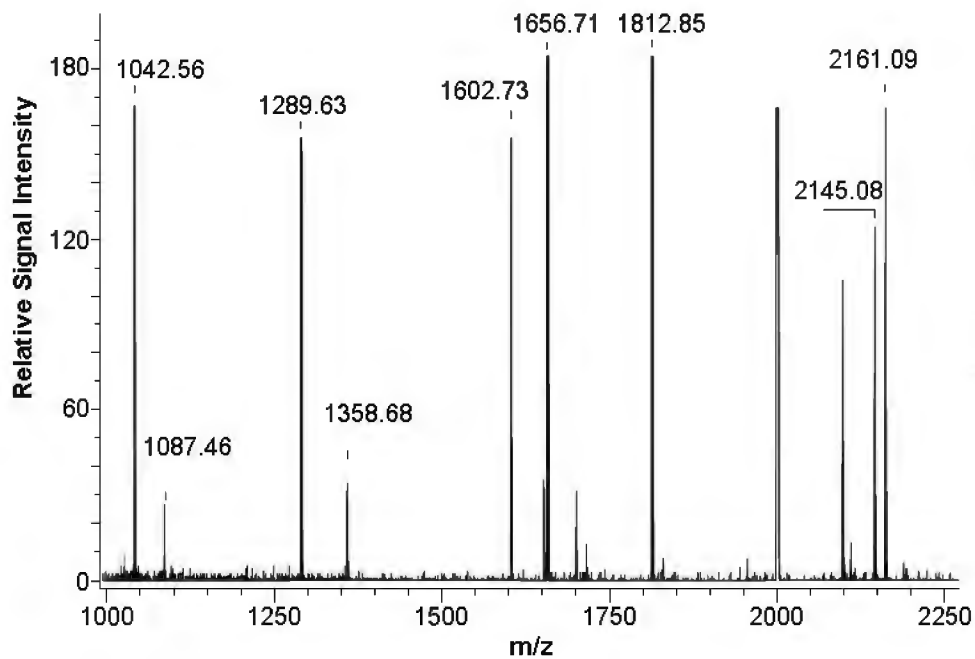
We thank Dr. Haifeng Chen (Virovek, Hayward, CA) for providing us AAV2-EGFP vectors.

## Author Contributions

Conceived and designed the experiments: BD XD PK WX. Performed the experiments: BD XD HYC LC HL WW. Analyzed the data: BD XD PK WX. Contributed reagents/materials/analysis tools: BH FW. Wrote the paper: BD XD PK WX.

## References

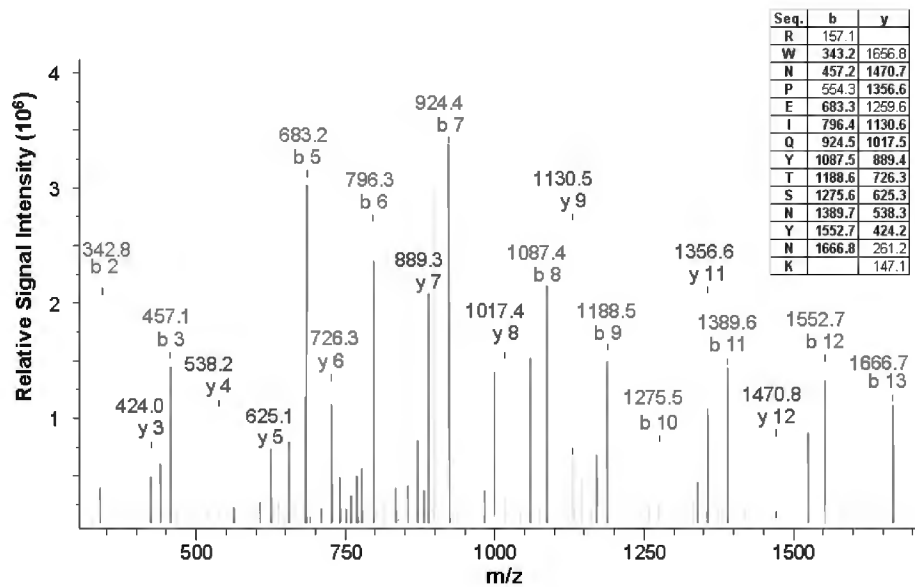
- Muzyczka N, Berns K (2001) Parvoviridae: the viruses and their replication. *Fields virology* 2: 2327–2359.
- Flotte T, Carter B, Conrad C, Guggino W, Reynolds T, et al. (1996) A phase I study of an adeno-associated virus-CFTR gene vector in adult CF patients with mild lung disease. *Hum Gene Ther* 7: 1145–1159.
- Maguire AM, Simonelli F, Pierce EA, Pugh EN Jr, Mingozzi F, et al. (2008) Safety and efficacy of gene transfer for Leber's congenital amaurosis. *N Engl J Med* 358: 2240–2248.
- Manno CS, Pierce GF, Arruda VR, Glader B, Ragni M, et al. (2006) Successful transduction of liver in hemophilia by AAV-Factor IX and limitations imposed by the host immune response. *Nat Med* 12: 342–347.
- Nathwani AC, Tuddenham EG, Rangarajan S, Rosales C, McIntosh J, et al. (2011) Adenovirus-associated virus vector-mediated gene transfer in hemophilia B. *N Engl J Med* 365: 2357–2365.
- Gray SJ, Choi VW, Asokan A, Haberman RA, McCown TJ, et al. (2011) Production of recombinant adeno-associated viral vectors and use in in vitro and in vivo administration. *Curr Protoc Neurosci Chapter 4: Unit 4 17*.
- Grieger JC, Choi VW, Samulski RJ (2006) Production and characterization of adeno-associated viral vectors. *Nat Protoc* 1: 1412–1428.
- Allay JA, Sleep S, Long S, Tillman DM, Clark R, et al. (2011) Good manufacturing practice production of self-complementary serotype 8 adeno-associated viral vector for a hemophilia B clinical trial. *Hum Gene Ther* 22: 595–604.
- Lock M, Alvira MR, Wilson JM (2012) Analysis of particle content of recombinant adeno-associated virus serotype 8 vectors by ion-exchange chromatography. *Hum Gene Ther Methods* 23: 56–64.
- Wright JF (2009) Transient transfection methods for clinical adeno-associated viral vector production. *Hum Gene Ther* 20: 698–706.
- Ayuso E, Mingozzi F, Montane J, Leon X, Anguela XM, et al. (2010) High AAV vector purity results in serotype- and tissue-independent enhancement of transduction efficiency. *Gene Ther* 17: 503–510.
- Wright JF (2011) Adeno-associated viral vector manufacturing: keeping pace with accelerating clinical development. *Hum Gene Ther* 22: 913–914.
- Murray S, Nilsson CL, Hare JT, Emmett MR, Korostelev A, et al. (2006) Characterization of the capsid protein glycosylation of adeno-associated virus type 2 by high-resolution mass spectrometry. *J Virol* 80: 6171–6176.
- Nash K, Chen W, Salganik M, Muzyczka N (2009) Identification of cellular proteins that interact with the adeno-associated virus rep protein. *J Virol* 83: 454–469.
- Van Vliet K, Mohiuddin Y, McClung S, Blouin V, Rolling F, et al. (2009) Adeno-associated virus capsid serotype identification: Analytical methods development and application. *J Virol Methods* 159: 167–177.
- Wang J, Xie J, Lu H, Chen L, Hauck B, et al. (2007) Existence of transient functional double-stranded DNA intermediates during recombinant AAV transduction. *Proc Natl Acad Sci U S A* 104: 13104–13109.
- Johnson JS, Samulski RJ (2009) Enhancement of adeno-associated virus infection by mobilizing capsids into and out of the nucleolus. *J Virol* 83: 2632–2644.
- Qiu J, Brown KE (1999) A 110-kDa nuclear shuttle protein, nucleolin, specifically binds to adeno-associated virus type 2 (AAV-2) capsid. *Virology* 257: 373–382.
- Bevington JM, Needham PG, Verrill KC, Collaco RF, Basrur V, et al. (2007) Adeno-associated virus interactions with B23/Nucleophosmin: identification of sub-nucleolar virion regions. *Virology* 357: 102–113.
- Sanlioglu S, Benson PK, Yang J, Atkinson EM, Reynolds T, et al. (2000) Endocytosis and nuclear trafficking of adeno-associated virus type 2 are controlled by *rac1* and phosphatidylinositol-3 kinase activation. *J Virol* 74: 9184–9196.
- Hansen J, Qing K, Srivastava A (2001) Adeno-associated virus type 2-mediated gene transfer: altered endocytic processing enhances transduction efficiency in murine fibroblasts. *J Virol* 75: 4080–4090.
- Aslanidi GV, Rivers AE, Ortiz L, Govindasamy L, Ling C, et al. (2012) High-efficiency transduction of human monocyte-derived dendritic cells by capsid-modified recombinant AAV2 vectors. *Vaccine* 30: 3908–3917.
- Zhong L, Li B, Jayandharan G, Mah CS, Govindasamy L, et al. (2008) Tyrosine-phosphorylation of AAV2 vectors and its consequences on viral intracellular trafficking and transgene expression. *Virology* 381: 194–202.
- Zhong L, Li B, Mah CS, Govindasamy L, Agbandje-McKenna M, et al. (2008) Next generation of adeno-associated virus 2 vectors: point mutations in tyrosines lead to high-efficiency transduction at lower doses. *Proc Natl Acad Sci U S A* 105: 7827–7832.
- Jayandharan GR, Zhong L, Li B, Kachniarz B, Srivastava A (2008) Strategies for improving the transduction efficiency of single-stranded adeno-associated virus vectors in vitro and in vivo. *Gene Ther* 15: 1287–1293.
- Shen WY, Lai YK, Lai CM, Rakoczy PE (2004) Impurity of recombinant adeno-associated virus type 2 affects the transduction characteristics following subretinal injection in the rat. *Vision Res* 44: 339–348.
- Duan X, Kelsen SG, Merali S (2008) Proteomic analysis of oxidative stress-responsive proteins in human pneumocytes: insight into the regulation of DJ-1 expression. *J Proteome Res* 7: 4955–4961.
- Duan X, Kelsen SG, Clarkson AB Jr, Ji R, Merali S (2010) SILAC analysis of oxidative stress-mediated proteins in human pneumocytes: new role for treacle. *Proteomics* 10: 2165–2174.



Meas. m/z	Cal. MH <sup>+</sup>	Range	Sequence
1042.56	1042.64	640 - 648	HPPPQILIK
1087.46	1087.49	286 - 293	FHC*HFSPR
1289.63	1289.68	609 - 619	DVYLQGPIWAK
1358.68	1358.73	298 - 308	LINNNGFRPK
1602.73	1602.81	649 - 664	NTPVPANPSTTFSAAK
1656.71	1656.77	693 - 705	WNPEIQYTSNYNK
1812.85	1812.87	692 - 705	RWNPEIQYTSNYNK
2145.08	2145.07	620 - 639	IPHTDGHFHPSPLMGGFGLK
2161.09	2161.07	620 - 639	IPHTDGHFHPSPLM*GGFGLK

C\* -Carbamidomethyl of cysteine M\* -Oxidation of methionine

Figure S1.



Range	Sequence	Ion Score
245 - 257	R.TWALPTYNNHLYK.Q	48
314 - 320	K.LFNIQVK.E	32
527 - 543	K.DDEEKFFPQSGVLIFGK.Q	64
549 - 555	K.TNVDIEK.V	50
556 - 565	K.VMITDEEEIR.T	75
609 - 619	R.DVYLQGPWAK.I	61
640 - 648	K.HPPPQILIK.N	38
649 - 664	K.NTPVPANPSTTFSAAK.F	70
692 - 705	K.RWNPEIQYTSNYNK.S	86
693 - 705	R.WNPEIQYTSNYNK.S	77

Figure S2.