Nanoscale

PAPER

Cite this: *Nanoscale*, 2020, **12**, 23721

Adeno-associated virus characterization for cargo discrimination through nanopore responsiveness†

Buddini Iroshika Karawdeniya, ‡[a] Y. M. Nuwan D. Y. Bandara, ‡[a] Aminul Islam Khan,‡[b] Wei Tong Chen,[c] Hoang-Anh Vu,[d] Adnan Morshed, [b] Junghae Suh,[c,d,e,f] Prashanta Dutta[b] and Min Jun Kim *[a]

Solid-state nanopore (SSN)-based analytical methods have found abundant use in genomics and proteomics with fledgling contributions to virology – a clinically critical field with emphasis on both infectious and designer-drug carriers. Here we demonstrate the ability of SSN to successfully discriminate adeno-associated viruses (AAVs) based on their genetic cargo [double-stranded DNA (AAV_{dsDNA}), single-stranded DNA (AAV_{ssDNA}) or none (AAV_{empty})], devoid of digestion steps, through nanopore-induced electro-deformation (characterized by relative current change; $\Delta I/I_0$). The deformation order was found to be $AAV_{empty} > AAV_{ssDNA} > AAV_{dsDNA}$. A deep learning algorithm was developed by integrating support vector machine with an existing neural network, which successfully classified AAVs from SSN resistive-pulses (characteristic of genetic cargo) with >95% accuracy – a potential tool for clinical and biomedical applications. Subsequently, the presence of AAV_{empty} in spiked AAV_{dsDNA} was flagged using the $\Delta I/I_0$ distribution characteristics of the two types for mixtures composed of ~75 : 25% and ~40 : 60% (in concentration) $AAV_{empty} : AAV_{dsDNA}$.

Received 28th July 2020,
Accepted 12th November 2020

DOI: 10.1039/d0nr05605g

rsc.li/nanoscale

Introduction

With over 200 clinical studies globally and the recent FDA approval of Luxturna – the first approved gene therapy in the United States to treat hereditary blindness[1,2] – adeno-associated virus (AAV) vectors are gaining substantial traction in viral gene therapy. One considerable challenge in the translation of AAV vectors, once produced, is the difficulty of characterizing the vectors based on their transgene packaging. Key characterization metrics of AAV include titer (capsid and genome titers), exact genomic content [single-stranded (ssDNA) *versus* self-complementary double-stranded (dsDNA) and overall genome length], and heterogeneity of a vector preparation (empty *versus* full capsids). To obtain these metrics, a combination of multiple assays has to be performed, including quantitative polymerase chain reaction (qPCR) or droplet digital PCR (ddPCR) for genomic titer,[3] enzyme-linked immunosorbent assay (ELISA) for capsid titer, and analytical ultracentrifugation for vector prep composition.[4] Alarmingly, the variability associated with vector characterization assays was revealed through a blind study in which AAV samples were sent to several groups for vector titering using qPCR and ELISA. The mean and standard deviation (SD) for genomic and capsid titer were $3.82 \times 10^{10} \pm 2.97 \times 10^{10}$ and $9.43 \times 10^{11} \pm 3.19 \times 10^{11}$, respectively.[5] These are highly concerning outcomes especially for dose-dependent therapeutics where overdosing, for example through underestimation of the empty capsids, could trigger unexpected immune responses. The genomic content could also be analyzed through alkaline gel assay or a Southern blot which, however, requires overnight runs and are semi-quantitative at best. In this study, we demonstrate the use of solid-state nanopores (SSNs) – low-cost, ostensibly simple, low-sample requiring, high sensing throughput, label-free single-molecule sensor class – to characterize each vector type based on electro-deformation, discriminate between the vectors, and flag the presence of empty capsid from a mixture (a critical development step towards quality assurance of AAV preps).

A solitary nano-scale aperture that spans an impervious membrane (biological or solid-state) separating two electrolyte reservoirs – a nanopore – has been used to characterize a

---

[a] *Department of Mechanical Engineering, Southern Methodist University, Dallas, TX, 75275, USA. E-mail: mjkim@lyle.smu.edu*
[b] *School of Mechanical and Materials Engineering, Washington State University, Pullman, WA, 99164, USA*
[c] *Department of Chemical and Biomolecular Engineering, Rice University, Houston, TX, 77005, USA*
[d] *Department of Bioengineering, Rice University, Houston, TX, 77005, USA*
[e] *Department of Biosciences, Rice University, Houston, TX, 77005, USA*
[f] *Systems, Synthetic, and Physical Biology Program, Rice University, Houston, TX, 77005, USA*

† Electronic supplementary information (ESI) available. See DOI: 10.1039/d0nr05605g

‡ These authors contributed equally.

myriad of biomolecules[6–10] and particles,[11–14] nanoparticles[15,16] and synthetic polymers[17] using a multitude of molecular level features[6,18,19] and membrane mechanical properties (*i.e.*, stiffness, deformability).[11,12] However, compared to the plethora of DNA and protein studies, the virus footprint in the nanopore community is surprisingly meager perhaps because nanopores were mostly recognized for small molecule analysis (driven by potential commercial interests) while studies on filamentous,[20] rod-like,[21,22] and spherical viruses[12,23] have redrawn the boundaries of nanopore technology. With the emergence of new viral threats challenging the very fabric of human existence, the importance of developing low-cost, high throughput, portable technologies for diagnostic purposes has gained substantial focus with the dawn of 2020. While biological nanopores have been used to sequence the genome of viruses such as Zaire Ebola[24] – emphasizing the clinical importance of this technology – our proposed method using SSN would analyze the virus particles, devoid of digestions steps, and would eventually pave the way for rapid assessment of the genetic cargo and the purity of an AAV prep. In this work, we designed and tested a silicon nitride (Si_xN_y) based SSN device to characterize three AAV vector types – empty (AAV_{empty}), AAV with ssDNA (AAV_{ssDNA}) and AAV with dsDNA (AAV_{dsDNA}) – using the demonstrated ability of SSN to estimate the electro-deformation of soft nanoparticles in response to an electric field[11,12] and numerical predictions to quantify the deformation. In addition, we used a deep convolutional neural network to classify AAVs based on their cargo from the resistive pulse data. Machine learning approaches have been shown to distinguish biomolecules using ionic current-time waveforms.[25–27] The deep neural network used here was developed by modifying an existing residual neural network (ResNet50) with a support vector machine. The electro-deformation was apparent through the voltage trend of the relative-current change ($\Delta I/I_0$) originating from particle transit – a departure from the linear (Ohmic) nature was observed. The extent to which a given particle deforms is a function of its spring constant for which both membrane mechanical properties and intra-particle properties such as transgene packaging are paramount. For example, an AAV_{empty} is expected to deform more than a cargo-carrying counterpart of the same serotype. Therefore, the deformation characteristics of the three AAV types are expected to be fundamentally different and we intend to use this property to discriminate each type. The expected deformation order is $AAV_{empty} > AAV_{ssDNA} > AAV_{dsDNA}$, which as will be shown later, agrees well with our obtained results. Such discrimination would be useful to flag the presence of AAV_{empty} in a sample consisting of genetic cargo-carrying AAV capsids. This is important as previous studies demonstrated that a high dose of AAV vector can cause severe toxicity which may be triggered by the high capsid dose or cargo expression.[28,29] Therefore, it is especially crucial in dose-dependent studies to know not only the AAV concentration but also the composition of the vector distribution to reduce the delivery of excessive AAV capsids.

Results and discussion

Basic operation principle of a nanopore is outlined in Fig. 1a where the analyte (AAV in this case) is added to the *cis* side and a voltage (negative for AAV) is applied to the *trans* side to drive the analyte across the nanopore from the *cis* to the *trans* side. This perturbs the open-pore ionic current stamping particle specific information. All experiments were conducted with <10 nM AAV – this minimal sample usage complements the tedious AAV preparation methods. Given the AAV size (∼25 nm in diameter), SSN is an obvious requirement since the narrowest constriction of ubiquitous biological nanopores such as α-hemolysin and MspA are not wide enough for such a particle to transit. A rich blend of fabrication techniques are accessible to us, such as controlled dielectric breakdown (CDB),[30,31] focused ion beam (FIB)[32,33] and transmission electron microscope (TEM).[34] Since it is difficult to fabricate larger diameter nanopores using CDB due to non-opening failure among other factors,[35] and preliminary studies with FIB fabricated pores produced poor event frequencies, we ultimately fabricated nanopores of ∼100 nm in diameter through ∼12 nm thick free-standing Si_xN_y on silicon using TEM (Fig. S1a†). Any pore showing significant current rectification was discarded and only those with rectification ∼1 was used (Fig. S1b†). Although the nanopore devices are low-cost and high throughput sensors (in general), the fabrication method (both the membrane and the pore) would, to a large extent, govern the overall cost associated with the device. Since, TEM-based nanopore fabrication methods are not as high throughput as other methods such as CDB, we note, the workflow could be limited by the pore fabrication step. However, if this limitation could be overcome, the overall throughput (combination of the fabrication and sensing time scales) and cost of the nanopore device could be improved significantly. The voltage polarity used for AAV translocations herein (−20 mV to −175 mV; Fig. 1b) has an added advantage of being immune to any DNA contaminations during AAV preparation (*i.e.*, any DNA that did not get encapsidated) since DNA would only respond to a positive voltage bias at this operational electrolyte chemistry (2 M LiCl buffered at pH ∼ 7). All experiments were in triplicate (unless otherwise noted) with unique nanopores of comparable size and each nanopore was discarded after running a given virus type to avoid any cross-contamination. On average, a minimum of 500 resistive pulses were recorded at −20 mV whereas a minimum of 1000 resistive pulses were recorded at subsequent voltages.

The question that intrigued us was, *can the resistive pulses (each resistive pulse corresponds to a single AAV particle translocating through the nanopore) shown in* Fig. 1b *be used to characterize and distinguish each AAV class*? For this, we initially used a deep neural network (DNN) framework which is based on an existing deep residual network (ResNet50 – an award-winning platform developed for the ImageNet database[36]) as shown in Fig. 1c. Since ResNet50 is not trained for virus detection, we had to make necessary changes to our network by modifying the last few steps from the ResNet50. Therefore, the

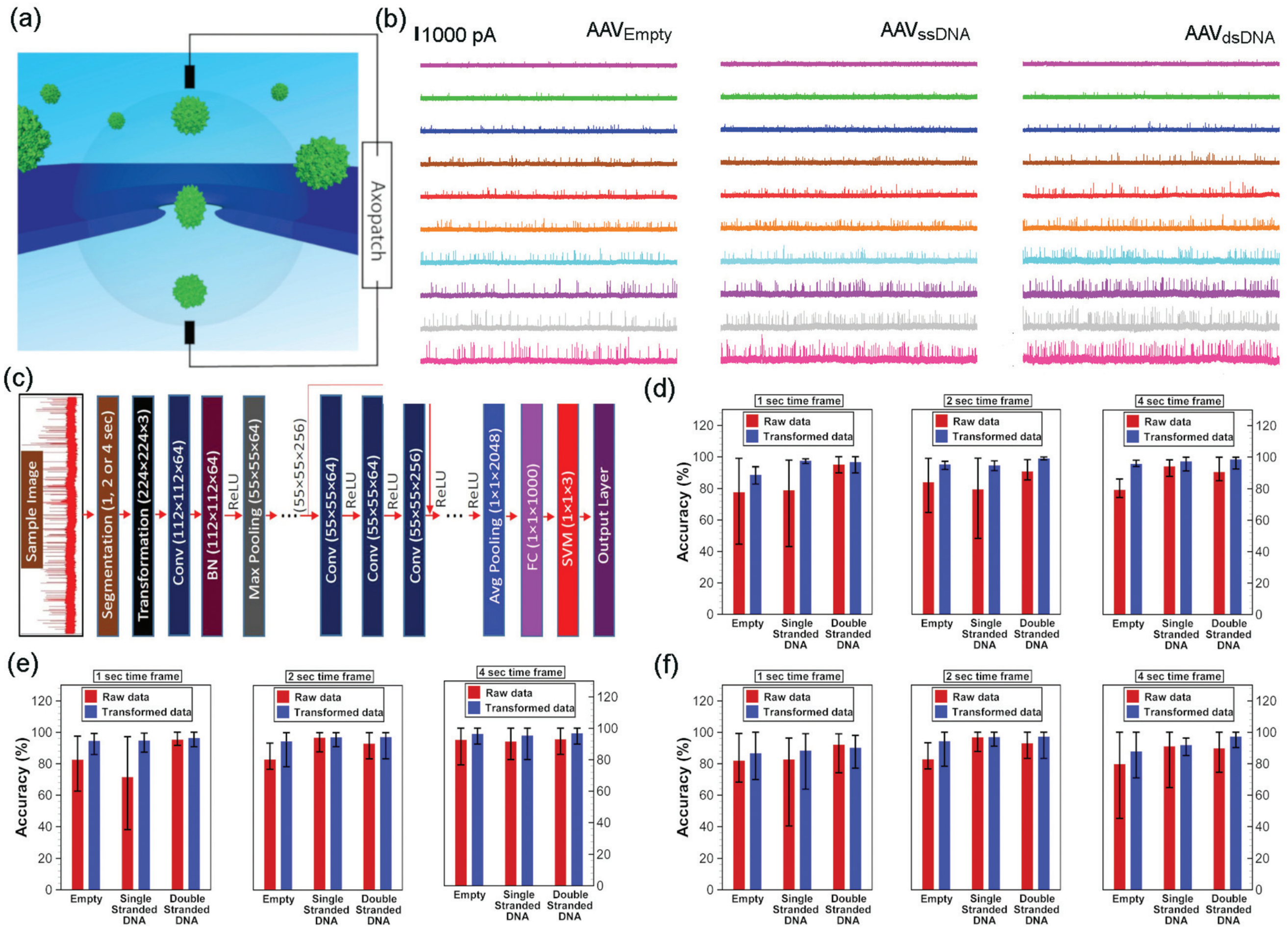

Fig. 1 (a) A schematic representation of a typical nanopore setup where the analyte is driven across the nanopore in response to an applied electric field stamping analyte-specific resistive-pulses as shown in (b). (b) Representative 10 seconds current traces corresponding to the translocation of "AAV_{empty} (left column)", "AAV_{ssDNA} (middle column)" and, "AAV_{dsDNA} (right column)" at −20 mV (magenta), −30 mV (green), −40 mV (blue), −50 mV (brown), −60 mV (red), −75 mV (orange), −100 mV (cyan), −125 mV (purple), −150 mV (gray) and −175 mV (pink). (c) Architecture of the deep convolutional neural network used for the classification of AAVs based on the genetic cargo. All current signals (images) obtained from solid-state nanopore experiments were segmented based on time frames (1, 2, or 4 s) and resized to have the dimensions of 224 × 224 × 3 to comply with ResNet50 requirements before those images are inputted into the network. Both raw data as well as transformed data was used for the classification, where the Conv.: convolutional block; BN: batch normalization; ReLU: rectified linear activation unit; Max Pooling: maximum pooling; FC: fully connected layer; SVM: support vector machine; Avg pooling: average pooling. Efficacy of deep convolutional neural networks in classification of AAVs from nanopore experimental data was obtained for the applied bias voltages of (d) −175 mV, (e) −150 mV and (f) −100 mV using 1 s, 2 s and 4 s frames. For a particular class, 80% of the data were randomly selected from all images of that class for training the network, while the rest of the images of that class were used for validation. Mean and error bars were obtained from training and validation of the model by randomly selecting the training and validation data set for 10 times.

output features of fully connected layer (1000 1 × 1) of the ResNet50 are fed into a support vector machine (SVM) as shown in Fig. 1c. Thus, based on the extracted features of the fully connected layer (fc1000) of ResNet50, we trained a multi-class SVM using the one-*versus*-one method[37] for three different classes (see Deep neural network for classification of virus cargoes from electrical signal section under Methods for more details). Classification results obtained from our deep learning model are shown in Fig. 1d–f for AAVs translocation data at voltage −175 mV, −150 mV and −100 mV (see ESI Table S1† for the number of images used in each class for different time frames). Our results show that accuracy (Fig. 1d–f) can be improved by transforming (see Methods for transformation of data) the experimental data. Although maximum accuracy is sometimes higher for raw data than transformed data, the mean accuracy for any class (or any time frame) is always lower for raw data than that of transformed data (Fig. 1d–f). The transformed data has always outperformed the raw data because distinct features of signals are more preserved in transformed data. Unlike classical machine learning problem, we used a few experimental data for training (for a given applied electric field), but segmentation of each experimental signal helped us to attain our desired goal in the data-driven classification. The accuracy of the prediction is improved significantly for 4 seconds time frame data, even though the number of images used to train the support vector machine is much smaller for 4 seconds case compared to 1 or 2 seconds cases. Further increase in the time frame window would probably help us to get a better prediction, but one must be mindful of the reduction in the training data with an

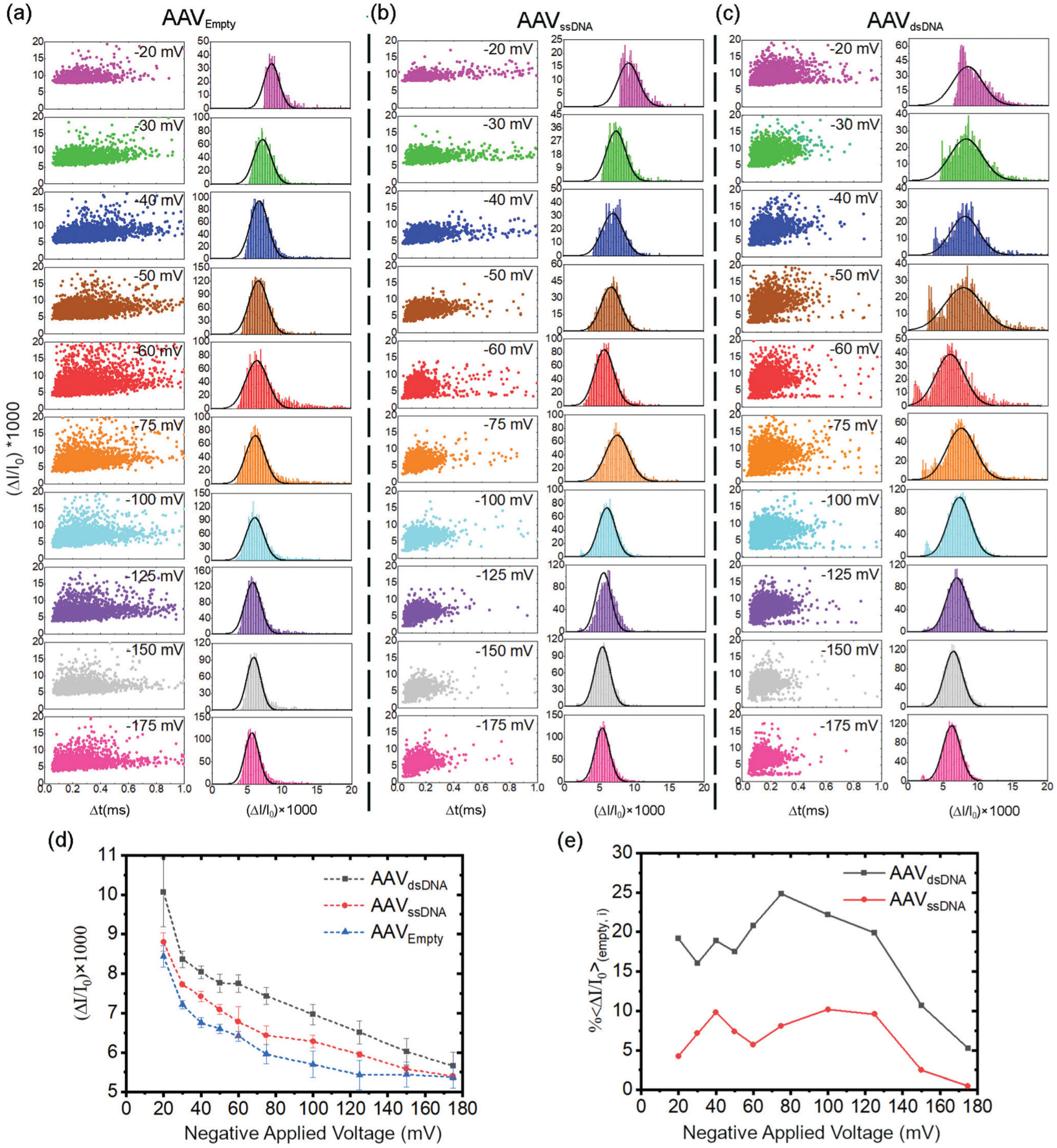

Fig. 2 Scatter plots of relative current change ($\Delta I/I_0$) *versus* translocation time (left column) and, histograms corresponding to the $\Delta I/I_0$ (right column) of (a) $\text{AAV}_{\text{empty}}$, (b) $\text{AAV}_{\text{ssDNA}}$, (c) $\text{AAV}_{\text{dsDNA}}$ at −20 mV (magenta), −30 mV (green), −40 mV (blue), −50 mV (brown), −60 mV (red), −75 mV (orange), −100 mV (cyan), −125 mV (purple), −150 mV (gray) and −175 mV (pink). The histograms were fitted with a single Gaussian function (see ESI† section 1 for the histogram and fitting details). (d) $\Delta I/I_0$ *vs.* voltage corresponding to $\text{AAV}_{\text{dsDNA}}$ (black), $\text{AAV}_{\text{ssDNA}}$ (red) and $\text{AAV}_{\text{empty}}$ (blue). (e) Percentage $\Delta I/I_0$ difference of AAV encapsulating dsDNA (black) and ssDNA (red) with respect to $\text{AAV}_{\text{empty}}$ ($\%\langle\Delta I/I_0\rangle_{(\text{empty},i)}$, calculated using eqn (S2)†). All experiments were in triplicate with unique nanopores and performed using ~98 ± 5 nm diameter nanopores in 2 M LiCl (buffered at pH ~ 7 using 10 mM Tris buffer).

increase in the time frame. Our model results show that as long as our deep learning algorithm is trained with the appropriate data, we can get accurate prediction despite the fact that data-based techniques such as machine learning is never 100% accurate. The reason for no false positive (or negative) in our proposed method is due to flexibility of testing multiple frames from a single experiment by segmenting resistive-pulse data into hundreds of smaller time frames. This comprehensive analysis indicates that our approach to identify AAVs based on electric current signal is robust, and this algorithm can be used to detect viruses quickly from SSN experiments.

After successfully identifying each AAV class using our DNN model, we then investigated the possibility of using relative current change ($\Delta I/I_0$) to discriminate each AAV type as this metric is dependent on the membrane rigidity.[12,38] The scatter plots of $\Delta I/I_0$ *versus* translocation time and histograms corresponding to $\Delta I/I_0$ are shown in Fig. 2a–c. Each of the histograms were fitted with a single Gaussian function (see ESI section 1† for the histogram and fitting details). The behaviour of $\Delta I/I_0$ with voltage for each AAV class as shown in Fig. 2d, is indicative of electro-deformation due to deviation from the Ohmic linear scaling (see the associated discussion of ESI section 5† for more details). A single AAV can only house a single DNA molecule.[39,40] Iodixanol ultracentrifugation can effectively separate the empty capsids formed during production,[41,42] thus reducing the complexity arising from the number of DNA copies inside an AAV – either a given particle will have a single copy of the intended genome package or it will not. The deformation profiles ($\Delta I/I_0$ *versus* applied voltage) as seen in Fig. 2d indicate that electro-deformation follow an inverse relationship with cargo content: lesser the void within the AAV (*i.e.*, higher the volume occupied by the cargo material), lesser would be the deformation ($\text{AAV}_{\text{empty}} > \text{AAV}_{\text{ssDNA}} > \text{AAV}_{\text{dsDNA}}$). At higher voltages, we see that the differences in the deformation profiles become less, and almost within error at ∼−175 mV. If the trends of the deformation profiles seen in Fig. 2d continue, one would pragmatically not expect to see any discernible differences between each virus type (based on the genetic cargo) at voltages >−175 mV. We then plotted the percentage difference of the $\Delta I/I_0$ of each AAV type referenced to $\text{AAV}_{\text{empty}}$ ($\%\langle\Delta I/I_0\rangle_{(\text{empty},i)}$, see ESI section 2† – for the definition). The trends of Fig. 2e show a sharp drop in $\%\langle\Delta I/I_0\rangle_{(\text{empty},i)}$ at voltages above −125 mV. The $\text{AAV}_{\text{dsDNA}}$ and $\text{AAV}_{\text{ssDNA}}$ showcased an averaged $\%\langle\Delta I/I_0\rangle_{(\text{empty},i)}$ of ∼19.9 ± 2.7% and ∼7.8 ± 2.1%, respectively up to −125 mV. It is not surprising to see $\text{AAV}_{\text{dsDNA}}$ having the greatest difference with respect to $\text{AAV}_{\text{empty}}$ (*i.e.*, highest $\%\langle\Delta I/I_0\rangle_{(\text{empty},i)}$) as it is expected to deform the least. It typically takes ∼2 hours in total to acquire the minimum event count for all voltages noted previously (at least 500 events for −20 mV and 1000 events for the rest). Among these voltages, all AAV types showed >1 resistive pulses per s at ≥−50 mV. Taking the useful voltage regime for electro-deformation-based discrimination (≤−125 mV) and appreciable event frequency (≥−50 mV), one can bracket −50 mV to −125 mV as the optimized voltage range for this study. Consequently, it merely takes ∼30 minutes in total to acquire ∼1000 events for all voltages in this optimized voltage range – a testament to the sensing throughput of the nanopore platform.

To numerically model the deformation of each AAV type, we used an immersed interface method (IIM), which has been developed and validated earlier.[43–45] The IIM can estimate the electric potential distribution inside the nanopore geometry, which is used to calculate electric current at any particular section using $I = -\sigma A(\nabla\phi)\cdot\hat{\eta}$, where σ is the local conductivity, A is the cross-sectional area, and $\hat{\eta}$ is the direction vector normal to the particular section. To quantify the extent of deformation, we defined aspect ratio (α) of the virus as the ratio of its equatorial (along the electric field) length over the polar (perpendicular to the electric field) length. For a circular shape, α = 1.0, and it is greater than unity when the virus is deformed in the direction of the applied field.[44] As shown in Fig. 3a–c, when the viruses are allowed to deform with an increasing electric field, we observe consistent nonlinear behavior in the $\Delta I/I_0$ due to competing electrostatic and electrophoretic forces on the virus capsid.[38,45] For a properly chosen

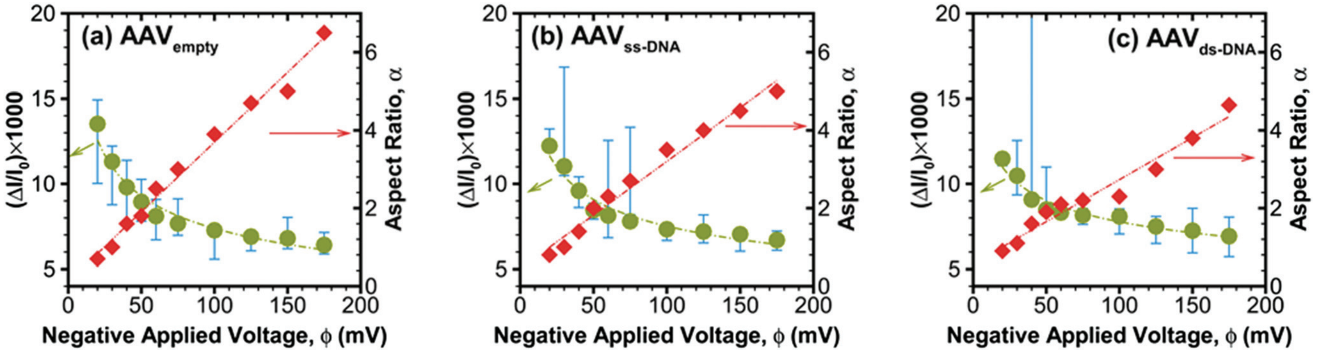

Fig. 3 Nondimensional relative current change ($\Delta I/I_0$) and corresponding aspect ratio (α) with negative applied voltage (ϕ) corresponding to (a) $\text{AAV}_{\text{empty}}$, (b) $\text{AAV}_{\text{ssDNA}}$ and (c) $\text{AAV}_{\text{dsDNA}}$. The numerically calculated relative current drop data (green circle) are fitted with a power-law curve drawn in green lines while the estimated aspect ratio data (red diamonds) are fitted with a straight line (red dashed lines). Three sets of independent experimental data are used to calculate the bounding upper and lower values (blue bounding bars) for $\Delta I/I_0$.

set of conductivity ratios, the numerical predictions of $\Delta I/I_0$ (green circles) fall within the experimental bounds (blue limits) and reveal an interesting power-law like behavior (green dashed line, Fig. 3a–c) in all three cases. The corresponding change in the virus shape with increasing potential is presented in terms of the aspect ratio (red diamonds), which

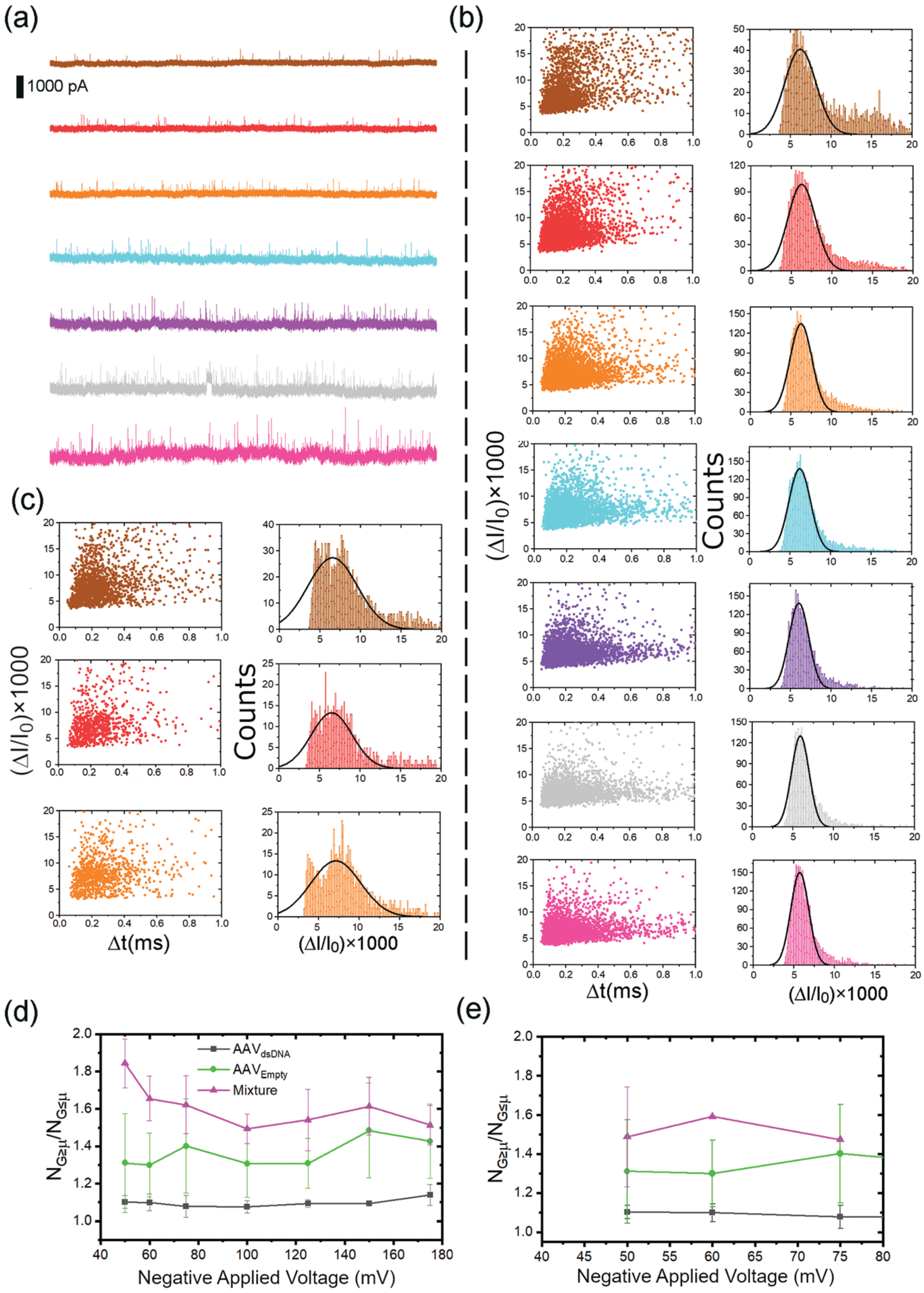

Fig. 4 (a) 5-Second representative current traces corresponding to a 75 : 25% molar mixture of $\text{AAV}_{\text{empty}}$ and $\text{AAV}_{\text{dsDNA}}$ in response to −50 mV (brown), −60 mV (red), −75 mV (orange), −100 mV (cyan), −125 mV (purple), −150 mV (gray) and −175 mV (pink). (b) The corresponding scattered plots and histograms of (a). (c) The scattered plots and histograms of 40 : 60% molar mixture of $\text{AAV}_{\text{empty}}$ and $\text{AAV}_{\text{dsDNA}}$ in response to −50 mV, −60 mV, −75 mV. (d)-(e) The ratio of the population above and below the mean of each Gaussian fit ($N_{G\geq\mu}/N_{G\leq\mu}$) of $\Delta I/I_0$ histograms with applied voltage, corresponding to $\text{AAV}_{\text{dsDNA}}$ (black), $\text{AAV}_{\text{empty}}$ (green) and mixtures (magenta) of (d) 75 : 25% and (e) 40 : 60% (in molar concentration) of $\text{AAV}_{\text{empty}}$ and $\text{AAV}_{\text{dsDNA}}$. Each mixture was measured in duplicate using unique nanopores.

shows a linearly increasing behavior with increasing electric field. The slope of this aspect ratio *vs.* applied voltage plot was found to be decreasing with increasing conductivity ratio (λ) (red dashed line with slopes of 0.0358, 0.0276, and 0.0216 mV^{-1} for $\text{AAV}_{\text{empty}}$, $\text{AAV}_{\text{ssDNA}}$, and $\text{AAV}_{\text{dsDNA}}$, respectively). The decrease in the slope also corresponds to a higher degree of deformation of the AAV samples. Hence, one can use the slope of the aspect ratio *vs.* applied voltage plot as a characteristic identifier of each virus type with its signature inner conductivity and deformation attributes.

We then ventured to mimic a sample of $\text{AAV}_{\text{dsDNA}}$ contaminated with $\text{AAV}_{\text{empty}}$ by spiking an $\text{AAV}_{\text{dsDNA}}$ aliquot with a significant amount of $\text{AAV}_{\text{empty}}$ (~75% $\text{AAV}_{\text{empty}}$ and ~25% $\text{AAV}_{\text{dsDNA}}$; $\text{AAV}_{75:25\%}$). Identification of $\text{AAV}_{\text{empty}}$ in vector batches is especially important for clinical studies to minimize adverse immune responses in patients. Since, in a real-world sample-scenario, the operator would not have pre-knowledge of such contamination, to stay true to such a situation, we fitted each $\Delta I/I_0$ profile with a single Gaussian rather than two or more Gaussians. Unlike the $\Delta I/I_0$ histograms of $\text{AAV}_{\text{dsDNA}}$ (Fig. 2c), spiked mixtures (Fig. 4b and c) cannot be well fitted with a single Gaussian – a clear population outside the Gaussian fit exist at higher $\Delta I/I_0$. The existence of an apparent outlying population compared to the one residing within the Gaussian fit may also serve as a visual and qualitative metric to qualify the presence of a significant $\text{AAV}_{\text{empty}}$ population in the sample. Thus, one can use the ratio of the population higher and lower than the mean of the Gaussian fit ($N_{G\geq\mu}/N_{G\leq\mu}$) as a metric to flag the presence of $\text{AAV}_{\text{empty}}$ in each sample: a perfect fit would have a value of 1 for $N_{G\geq\mu}/N_{G\leq\mu}$. Looking at Fig. 2a, it is evident that $\text{AAV}_{\text{empty}}$ has a broader distribution along the $\Delta I/I_0$ axis. The tail along the $\Delta I/I_0$ may indicate the presence of a secondary population, although not as prominent as the lower ($\Delta I/I_0$) and denser population (*i.e.*, the $\Delta I/I_0$ distribution can still be fitted with a single Gaussian function). However, such a tail along the $\Delta I/I_0$ axis is absent in its cargo-carrying counterparts. It could mean, the deformation is more restrictive in the presence of a cargo whereas it is more diverse in the absence of a cargo. Thus, it is not surprising to see $\text{AAV}_{\text{dsDNA}}$ having a $N_{G\geq\mu}/N_{G\leq\mu}$ value closer to one (Fig. 4d and e) whereas $\text{AAV}_{\text{empty}}$ deviating somewhat away from the ideal value. The mixture significantly deviated from the ideal value, which could be inextricably linked to the presence of populations corresponding to both $\text{AAV}_{\text{dsDNA}}$ and $\text{AAV}_{\text{empty}}$. It is evident from Fig. 4d, the profiles corresponding to $\text{AAV}_{\text{dsDNA}}$ and $\text{AAV}_{75:25\%}$ are well separated and indicative of a departure from a $\text{AAV}_{\text{dsDNA}}$ sample (*i.e.*, presence of a contaminant). We were also able to flag the presence of $\text{AAV}_{\text{empty}}$ in the mixture using deformation profiles ($\Delta I/I_0$ *vs.* voltage) as evident by Fig. S4† (see ESI section 6† for more details). Using the $N_{G\geq\mu}/N_{G\leq\mu}$ metric, we were able to flag the presence of $\text{AAV}_{\text{empty}}$ in a mixture of ~40% $\text{AAV}_{\text{empty}}$ and ~60% $\text{AAV}_{\text{dsDNA}}$ (Fig. 4e) through the visual separation of the mixture similar to above. The second mixture was deliberately limited to three voltages (−50 to −75 mV) as these three yielded the greatest separation of $N_{G\geq\mu}/N_{G\leq\mu}$ in the ~75 : 25% mixture evident from Fig. 4d. It is worthwhile noting, the error associated with the profile of $\text{AAV}_{\text{ds-DNA}}$ is much lower compared to the rest, which could also serve as a visual clue to the purity of the sample under investigation. One could potentially expand this study to cover a range of $\text{AAV}_{\text{dsDNA}}:\text{AAV}_{\text{empty}}$ ratios and develop a correlation between the $\text{AAV}_{\text{empty}}$ percentage and $N_{G\geq\mu}/N_{G\leq\mu}$ as a function of voltage.

Conclusions

We have demonstrated the ability of solid-state nanopores of ~100 nm diameter, fabricated using TEM through nominally ~12 nm thick Si_xN_y membranes, to discriminate AAV based on their genetic cargo (*i.e.*, single-stranded DNA, self-complementary DNA or none). All experiments were conducted using negative voltages and translocations were recorded from ~−20 mV to ~−175 mV in sufficiently small voltage increments. A deep neural network platform, developed based on ResNet50 with appropriate modifications by support vector machine, was used to identify the current profiles of each AAV type. The accuracy of the machine learning prediction can be improved significantly by segmenting each experimental resistive-pulse signal into hundreds of data and running the model for tens of data sets from each experiment. More importantly, the prediction accuracy increases with the length of the time frame (1 s *versus* 2 s *versus* 4 s) of experimental data. For transformed data, the mean accuracy of the network was always 90% or higher for any class regardless of the voltage bias or time frame. The electro-deformation was numerically modelled using an immersed interface approach. The model results indicated a power-law behavior for the nondimensional current drop ($\Delta I/I_0$) with applied potential for all three cases. Interestingly, the $\Delta I/I_0$ profiles with voltage clearly showed distinct deformation patterns for each AAV type with deformation being more prominent as the internal cavity of AAV is less occupied by its cargo: $\text{AAV}_{\text{empty}} > \text{AAV}_{\text{ssDNA}} > \text{AAV}_{\text{dsDNA}}$. The average percentage $\Delta I/I_0$ with respect to $\text{AAV}_{\text{empty}}$ ($(\%\langle\Delta I/I_0\rangle_{(\text{empty},i)})$ was, as expected, higher for $\text{AAV}_{\text{dsDNA}}$ than $\text{AAV}_{\text{ssDNA}}$ with the two having an averaged value of ~19.9 ± 2.7% and ~7.8 ± 2.1% respectively, up to −125 mV. Since $\text{AAV}_{\text{dsDNA}}$ displayed the highest difference, we ventured to see if $\text{AAV}_{\text{empty}}$ could be flagged from a mixture of $\text{AAV}_{\text{dsDNA}}$ and $\text{AAV}_{\text{empty}}$. Other than the difference associated with $\%\langle\Delta I/I_0\rangle_{(\text{empty},i)}$ another significant difference in the distribution of $\Delta I/I_0$ is $\text{AAV}_{\text{dsDNA}}$ being more Gaussian than $\text{AAV}_{\text{empty}}$ with the latter having a tail along higher $\Delta I/I_0$ values. This feature was used to successfully flag the presence of $\text{AAV}_{\text{empty}}$ in mixtures of ~75% $\text{AAV}_{\text{empty}}$ and ~25% $\text{AAV}_{\text{dsDNA}}$ and ~40% $\text{AAV}_{\text{empty}}$ and ~60% $\text{AAV}_{\text{dsDNA}}$. Taken together, SSN platforms along with their advantages such as low cost and sample requirement, rapid analysis, user friendliness with minimal training requirement (as seen with other nanopore technologies) could potentially transform the method discussed herein to a widely accessible tool to profile and discriminate each AAV class and to flag the presence of AAV_{empty} which could be crucial for minimizing safety issues with human gene therapy.

Methods

AAV production

AAV particles were produced using HEK293T cells (ATCC) using 25 kDa linear polyethylenimine (PEI, Thermo) mediated triple transfection.[46] Briefly, HEK293T cells were cultured to 70% confluency on 15 cm poly-L-lysine coated cell-culture plates using DMEM (LONZA) with 10% FBS (Atlanta Biologics) and 1% penicillin–streptomycin (Gibco). Adenovirus helper genes (pXX6-80), AAV9 rep-cap (pAAV2/9), and a transgene cassette plasmid (self-complement or single-stranded GFP) were mixed in a 1 : 1 : 1 molar ratio with the PEI transfection mix and allowed to incubate at room temperature for 30 minutes before adding to cells. The cell pellet was harvested 48 hours after transfection and underwent three cycles of freeze–thaw followed by benzonase treatment before purification using iodixanol (OptiPrep) step gradient (15%, 25%, 40%, 54%) ultracentrifugation. The 40% fraction was extracted, followed by concentration and buffer exchange using Amicon 150 kDa MWCO filtration unit (Millipore-Sigma) into GB-buffer (50 mM Tris, pH 7.6, 150 mM NaCl, 10 mM $MgCl_2$). Concentration of virus particles was established using qPCR using primers against cytomegalovirus (CMV) promoter (forward: TCACGGGGATTTCCAAGTCTC, reverse: AATGGGGCGGAGTTGTTACGAC) on the transgene cassette. The empty capsids were collected from the layer between the 25% and 40% fraction of the iodixanol column, and concentration was measured using western blot with B1 antibodies against a standard of AAV9 particles.

Nanopore electrical measurements

All electrical measurements were conducted using Ag/AgCl electrodes connected to an Axopatch 200B (Molecular Devices LLC, USA). The data were acquired at 250 kHz (except *I–V* measurements which were done at 10 kHz), filtered using the inbuilt Bessel low-pass filter at 10 kHz setting and digitized either using a PCIe-6321 connected to a BNC 2110 connector block (National Instruments, USA) or 1440A Digitizer (Molecular Devices LLC, USA). For pore diameter measurements, the former was used and for other temporal acquisitions, the latter was used. When the PCIe-6321/BNC 2110 was used, the instrument control was done using custom-coded LabVIEW scripts whereas pClamp (version 10.6, Molecular Devices LLC., USA) was used otherwise. Before each measurement, the *pipette offset* setting of the Axopatch 200B was used to nullify the zero-voltage current. The electrodes were prepared in the following manner: a ∼2-inch-long Ag wire was sanded to remove any oxide residuals and contaminants on the surface. Then it was dipped in a bleach solution (425044, Sigma Aldrich) for at least one hour (preferably overnight) until the electrode turns black. It was then soldered to a TE connectivity contact gold pin and connected to the head stage of the Axopatch 200B system. The electrodes were checked after each experiment to see whether it has retained its color or whether it has turned white. The latter indicates that the electrode needs to be sanded down and put in the bleach solution for it to function as a reversible electrode.

Nanopore fabrication

Nanopores were fabricated through as supplied silicon nitride chips (NBPX5001Z-HR, Norcada, Canada) that are nominally ∼12 nm thick using TEM (JEM-2100F, JEOL, Japan) at 200 keV as described previously (see Fig. S1† for a representative TEM image of a pore and its current–voltage curve).[47] The size was initially validated through TEM as shown in Fig. S1† and subsequently crosschecked with eqn (1).

Nanopore characterization

The fabricated nanopore chips were mounted between two Teflon half cells using PDMS gaskets to be watertight. Each chamber can hold ∼450 µL of electrolyte. The schematic diagram of the cell is shown in Fig. S2.† The chambers were initially filled with ethanol (A4094, Fisher Scientific), placed in a vacuum desiccator and connected to a mechanical pump to remove any air bubbles along the channel connecting the nanopore chip and the electrolyte reservoir. Upon the appearance of bubbles from both the channels, the pump was disconnected, and the system was equilibriated to atmospheric pressure gently to avoid re-entry of air bubbles. The content was then thoroughly exchanged with ultra-pure water followed by 1 M KCl (P9333, Sigma-Aldrich, USA) buffered at pH ∼ 7 (phosphate buffered saline, P5493, Sigma-Aldrich, USA). A voltage ramp of +200 mV to −200 mV was then applied to acquire a current–voltage (*I–V*) curve. The *I–V* curve was then linearly fitted and the slope (G) was used to estimate the nanopore size using,

$$G = \sigma\left[\frac{4L}{\pi D^2} + \frac{1}{D}\right]^{-1} \tag{1}$$

where G, σ, L and D are the ionic conductance, electrolyte conductivity, nanopore length, and diameter, respectively. If the pore is not properly wet, the *I–V* curve would showcase a G value significantly less than expected. Thus, all pores, before usage were subjected to a 2-second +8 V pulse to ensure proper-wetting.

Nanopore electrolyte preparation

All electrolytes including LiCl (L4408, Sigma-Aldrich, USA) and KCl were dissolved in ultra-pure water (ARS-102 Aries high purity water systems) with ∼18 MΩ cm resistivity. Each solution contains 10 mM of either Tris buffer (J61036, Fisher Scientific, USA) or phosphate buffered saline (P5493, Sigma-Aldrich, USA). The former was used for translocation experiments whereas the latter was used to acquire current–voltage (*I–V*) curves for pore-diameter estimation. The solutions were then filtered using a filtration system with a Polyethersulfone membrane (S2VPU02RE, Fisher Scientific). *Caution: dissolving LiCl in water is an exothermic process*. After the electrolyte solution reached the room temperature, the pH was adjusted by adding HCl (H1758, Sigma-Aldrich, USA) or KOH (306568, Sigma-Aldrich, USA) dropwise while gently stirring the electrolyte solution continuously. *Caution: these are concentrated solutions and should only be opened inside a properly functioning fume hood.* Both pH and conductivity of the electrolyte solutions were measured using an Orion Star™ pH meter. Typically, a 2 M LiCl solution at pH ∼ 7 would have a conductivity of ∼12 S m^{-1} whereas a 1 M KCl solution at pH ∼ 7 would have a conductivity of ∼11 S m^{-1}.

Event characterization

A custom MATLAB (version 9.4, USA) script was used, where events were characterized as perturbations at least 5 times the standard deviation of the open-pore current. In brief, the code scans through the open-pore current using custom moving windows. This ensures any subtle variations in the open-pore current of a given window is independent of the rest. The window size is typically set as $1/10^{\text{th}}$ the acquisition frequency (100 ms long window). Although larger windows can be used, we have observed the translocation times are mostly <1 ms, thus justifying the moving window size. This is also evident from the scatter plots shown in Fig. 2. The average of the data points in the window is then used to calculate a preliminary baseline, and any perturbation that is 5 times the standard deviation of the baseline is flagged and assigned temporarily to the value of the baseline. Then using the new values, a secondary baseline is calculated and used as the open-pore baseline of that window. After detecting an event (i.e., perturbations at least 5 times the standard deviation of the open-pore current), its duration (Δt), maximum current drop (ΔI) and the local baseline (I_0) are extracted to perform analysis shown in the manuscript.

Image preparation for deep neural network. Due to the unavailability of large training data (from experiments), we have segmented the electrical (resistive-pulse) signals of each experiment into $4N$, $2N$, and N number of images (graphs) depending on the time frames (1, 2, or 4 s). While it is possible to maintain the x-axis length constant in each image for a particular time frame, keeping the same scale range (the difference between the upper and lower bound of current) was challenging for the y-axis (Fig. 1b) during the auto plotting of graphs. Thus, we have trained and validated the model with two sets of images. The first set of images are plotted (aka raw data) automatically, while the y-axis of the second set was transformed as

$$I_{\text{up}} = I_{\text{low}} + \Delta I \tag{2}$$

where I_{up} and I_{low} are upper bound and lower bound of current values, respectively and ΔI is the current change. In other words, the second set of graphs (termed as transformed data) are plotted by considering a fixed current change (ΔI) in the vertical axis for all three classes.

Deep neural network for classification of virus cargoes from electrical signal

We have developed a deep neural network algorithm by modifying last couple of layers of the ResNet50 – a residual deep neural network developed by Microsoft research team. The ResNet50 has been trained for 1000 different classes with 13 000 000 natural images, and it requires a 224 × 224 × 3 color image as input for proper identification within its database. However, for our classification problem, we have only three classes based on the cargo inside AAVs: empty, single stranded DNA, and double stranded DNA. Thus, from the extracted features of the fully connected layer (fc1000) of ResNet50, we have trained a multiclass support vector machine (SVM) using the one-*versus*-one method[37] for three different classes. For the three class scenario, the one-*versus*-one method yields three binary classifiers where each one is trained on data from two classes. For example, to train data from the ith and the jth classes, we solved an optimization problem as

$$\min_{w^{ij},\, b^{ij},\, \xi^{ij}} \frac{1}{2}\left(w^{ij}\right)^{\mathrm{T}} w^{ij} + C\sum_{\mathrm{t}} \xi_{\mathrm{t}}\}^{ij} \tag{3}$$

where w, b, ξ, and C are the weight, bias, slack variable, and the penalty parameter, respectively. Eqn (3) is subjected to the following constraints

$$\begin{aligned} &\left(w^{ij}\right)^{\mathrm{T}}\varphi(x_{\mathrm{t}}) + b^{ij} \geq 1 - \xi_{\mathrm{t}}\}^{ij}, \quad \text{if} \quad y_{\mathrm{t}} = i \\ &\left(w^{ij}\right)^{\mathrm{T}}\varphi(x_{\mathrm{t}}) + b^{ij} \leq -1 + \xi_{\mathrm{t}}\}^{ij}, \quad \text{if} \quad y_{\mathrm{t}} = j \\ &\qquad \xi_{\mathrm{t}}{}^{ij} \geq 0 \end{aligned} \tag{4}$$

where x_{t} is the training data and y_{t} is the class of x_{t}. The function φ maps the training data x_{t} to a higher dimensional space. In eqn (3), the penalty (second) term $C\sum_{\mathrm{t}} \xi_{\mathrm{t}}\}^{ij}$ is used to reduce the number of training errors in case the data are not linearly separable, while optimization of the regularization (first) term $\frac{1}{2}\left(w^{ij}\right)^{\mathrm{T}} w^{ij}$ provides the maximum margin between two classes of data. Thus, the basic concept behind SVM is to find a balance between the regularization term and the training errors.

Based on the optimized weight and bias, scores are calculated for each class from an unseen test/validation image, and the highest score is used for classification of that image. If f_{ij} is the classifier to distinguish a pair of classes i (positive examples) and j (negative examples), the classification criteria for a new image x

$$f(x) = \underset{i}{\arg\max}\left[\sum_j f_{ij}(x)\right]. \tag{5}$$

Software's used for figure construction

Fig. 1: OriginLab 2018b, TecPlot 9.0, GIMP 2.10.8
Fig. 2: OriginLab 2018b, GIMP 2.10.8
Fig. 3: TecPlot 9.0
Fig. 4: OriginLab 2018b, GIMP 2.10.8

Software's used for machine learning

MATLAB 2018b.

Conflicts of interest

J. S. is an employee of Biogen as of August 2019.

W. T. C. is an employee of Biogen as of March 2020.

B. I. K is an employee of the Australian National University as of March 2020.

Y. M. N. D. Y. B is affiliated with the Australian National University as of September 2020.

An invention disclosure is filled for the work discussed herein.

Acknowledgements

This material is based upon work supported by a National Science Foundation Fellowship to W. C. (2018253392). We acknowledge the University of North Carolina at Chapel Hill Gene Therapy Center Vector Core for providing us with pXX6-80 and scAAV2-CMV-GFP, and the University of Pennsylvania Vector Core for providing us with pAAV2/9. The work was partly funded by NSF CMMI 1712069, NIH R03EB022759, and NIH R21GM134544. We would like to thank Prof. Moon J. Kim and Mr Qingxiao Wang at the University of Texas at Dallas and Mr Jung Soo Lee at Southern Methodist University for fabricating solid-state nanopores.

References

1 S. Russell, J. Bennett, J. A. Wellman, D. C. Chung, Z. F. Yu, A. Tillman, J. Wittes, J. Pappas, O. Elci, S. McCague, D. Cross, K. A. Marshall, J. Walshire, T. L. Kehoe, H. Reichert, M. Davis, L. Raffini, L. A. George, F. P. Hudson, L. Dingfield, X. Zhu, J. A. Haller, E. H. Sohn, V. B. Mahajan, W. Pfeifer, M. Weckmann, C. Johnson, D. Gewaily, A. Drack, E. Stone, K. Wachtel, F. Simonelli, B. P. Leroy, J. F. Wright, K. A. High and A. M. Maguire, *Lancet*, 2017, **390**, 849–860.

2 H. Ledford, *Nature*, 2017, **550**, 314.

3 M. Lock, M. R. Alvira, S.-J. Chen and J. M. Wilson, *Hum. Gene Ther: Methods.*, 2013, **25**, 115–125.

4 B. Burnham, S. Nass, E. Kong, M. Mattingly, D. Woodcock, A. Song, S. Wadsworth, S. H. Cheng, A. Scaria and C. R. O'Riordan, *Hum. Gene Ther: Methods.*, 2015, **26**, 228–242.

5 M. Lock, S. McGorray, A. Auricchio, E. Ayuso, E. J. Beecham, V. Blouin-Tavel, F. Bosch, M. Bose, B. J. Byrne, T. Cation, J. A. Chiorini, A. Chrarto, K. R. Clark, T. Conlon, C. Darmon, M. Doria, A. Douar, T. R. Flotte, J. D. Francis, A. Francois, M. Giacca, M. T. Korn, I. Korytov, X. Leon, B. Leuchs, G. Lux, C. Melas, H. Mizukami, P. Moullier, M. Müller, K. Ozawa, T. Philipsberg, K. Poulard, C. Raupp, C. Rivière, S. D. Roosendaal, R. J. Samulski, S. M. Soltys, R. Surosky, L. Tenenbaum, D. L. Thomas, B. van Montfort, G. Veres, J. F. Wright, Y. Xu, O. Zelenaia, L. Zentilin and R. O. Snyder, *Hum. Gene Ther.*, 2010, **21**, 1273–1285.

6 J. Saharia, Y. M. N. D. Y. Bandara, G. Goyal, J. S. Lee, B. I. Karawdeniya and M. J. Kim, *ACS Nano*, 2019, **13**, 4246–4254.

7 B. I. Karawdeniya, Y. M. N. D. Y. Bandara, J. W. Nichols, R. B. Chevalier and J. R. Dwyer, *Nat. Commun.*, 2018, **9**, 3278.

8 K. J. Freedman, M. Jürgens, A. Prabhu, C. W. Ahn, P. Jemth, J. B. Edel and M. J. Kim, *Anal. Chem.*, 2011, **83**, 5137–5144.

9 C. Plesa, D. Verschueren, S. Pud, J. van der Torre, J. W. Ruitenberg, M. J. Witteveen, M. P. Jonsson, A. Y. Grosberg, Y. Rabin and C. Dekker, *Nat. Nanotechnol.*, 2016, **11**, 1093–1097.

10 J. T. Hagan, B. S. Sheetz, Y. M. N. D. Y. Bandara, B. I. Karawdeniya, M. A. Morris, R. B. Chevalier and J. R. Dwyer, *Anal. Bioanal. Chem.*, 2020, **6**, 10.

11 J. S. Lee, J. Saharia, Y. M. N. D. Y. Bandara, B. I. Karawdeniya, G. Goyal, A. Darvish, Q. Wang, M. J. Kim and M. J. Kim, *Electrophoresis*, 2019, **40**(9), 1337–1344.

12 A. Darvish, J. S. Lee, B. Peng, J. Saharia, R. V. K. Sundaram, G. Goyal, N. Bandara, C. W. Ahn, J. Kim, P. Dutta, I. Chaiken and M. J. Kim, *Electrophoresis*, 2019, **40**(5), 776–783.

13 M. Tsutsui, T. Yoshida, K. Yokota, H. Yasaki, T. Yasui, A. Arima, W. Tonomura, K. Nagashima, T. Yanagida, N. Kaji, M. Taniguchi, T. Washio, Y. Baba and T. Kawai, *Sci. Rep.*, 2017, **7**, 1–9.

14 M. Tsutsui, M. Tanaka, T. Marui, K. Yokota, T. Yoshida, A. Arima, W. Tonomura, M. Taniguchi, T. Washio, M. Okochi and T. Kawai, *Anal. Chem.*, 2018, **90**, 1511–1515.

15 J. S. Lee, B. Peng, A. C. Sabuncu, S. Nam, C. Ahn, M. J. Kim and M. Kim, *Electrophoresis*, 2018, **39**, 833–843.

16 A. Darvish, G. Goyal, R. Aneja, R. V. Sundaram, K. Lee, C. W. Ahn, K.-B. Kim, P. M. Vlahovska and M. J. Kim, *Nanoscale*, 2016, **8**, 14420–14431.

17 J. W. Robertson, C. G. Rodrigues, V. M. Stanford, K. A. Rubinson, O. V. Krasilnikov and J. J. Kasianowicz, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 8207–8211.

18 Y. M. N. D. Y. Bandara, J. Tang, J. Saharia, L. W. Rogowski, C. W. Ahn and M. J. Kim, *Anal. Chem.*, 2019, **91**, 13665–13674.

19 B. Hornblower, A. Coombs, R. D. Whitaker, A. Kolomeisky, S. J. Picone, A. Meller and M. Akeson, *Nat. Methods*, 2007, **4**, 315.

20 A. McMullen, H. W. De Haan, J. X. Tang and D. Stein, *Nat. Commun.*, 2014, **5**, 4171.

21 H. Wu, Y. Chen, Q. Zhou, R. Wang, B. Xia, D. Ma, K. Luo and Q. Liu, *Anal. Chem.*, 2016, **88**, 2502–2510.

22 L. Liu, H. Wu, J. Kong and Q. Liu, *Sci. Adv. Mater.*, 2013, **5**, 2039–2047.

23 K. Zhou, L. Li, Z. Tan, A. Zlotnick and S. C. Jacobson, *J. Am. Chem. Soc.*, 2011, **133**, 1618–1621.

24 T. Hoenen, A. Groseth, K. Rosenke, R. J. Fischer, A. Hoenen, S. D. Judson, C. Martellaro, D. Falzarano, A. Marzi and R. B. Squires, *Emerging Infect. Dis.*, 2016, **22**, 331.

25 M. Taniguchi, *ACS Omega*, 2020, **5**, 959–964.

26 A. Arima, M. Tsutsui, I. H. Harlisa, T. Yoshida, M. Tanaka, K. Yokota, W. Tonomura, M. Taniguchi, M. Okochi and T. Washio, *Sci. Rep.*, 2018, **8**, 1–7.

27 A. Arima, I. H. Harlisa, T. Yoshida, M. Tsutsui, M. Tanaka, K. Yokota, W. Tonomura, J. Yasuda, M. Taniguchi, T. Washio, M. Okochi and T. Kawai, *J. Am. Chem. Soc.*, 2018, **140**, 16834–16841.

28 C. Hinderer, N. Katz, E. L. Buza, C. Dyer, T. Goode, P. Bell, L. K. Richman and J. M. Wilson, *Hum. Gene Ther.*, 2018, **29**, 285–298.

29 J. Hordeaux, Q. Wang, N. Katz, E. L. Buza, P. Bell and J. M. Wilson, *Mol. Ther.*, 2018, **26**, 664–668.

30 Y. M. N. D. Y. Bandara, J. Saharia, B. I. Karawdeniya, J. T. Hagan, J. R. Dwyer and M. J. Kim, *Nanotechnology*, 2020, **31**, 335707.

31 H. Kwok, K. Briggs and V. Tabard-Cossa, *PLoS One*, 2014, **9**, e92880.

32 A. S. Prabhu, T. Z. N. Jubery, K. J. Freedman, R. Mulero, P. Dutta and M. J. Kim, *J. Phys.: Condens. Matter*, 2010, **22**, 454107.

33 A. S. Prabhu, K. J. Freedman, J. W. Robertson, Z. Nikolov, J. J. Kasianowicz and M. J. Kim, *Nanotechnology*, 2011, **22**, 425302.

34 M. J. Kim, M. Wanunu, D. C. Bell and A. Meller, *Adv. Mater.*, 2006, **18**, 3149–3153.

35 I. Yanagi, R. Akahori and K.-i. Takeda, *Sci. Rep.*, 2019, **9**, 1–15.

36 K. M. He, X. Y. Zhang, S. Q. Ren and J. Sun, Deep residual learning for image recognition, in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, USA, 2016, pp. 770–778.

37 C. W. Hsu and C. J. Lin, *IEEE Trans. Neural. Netw.*, 2002, **13**, 415–425.

38 A. Morshed, B. I. Karawdeniya, Y. Bandara, M. J. Kim and P. Dutta, *Electrophoresis*, 2020, **41**, 449–470.

39 M. D. Weitzman and R. M. Linden, in *Adeno-Associated Virus*, Springer, 2012, pp. 1–23.

40 E. D. Horowitz, K. S. Rahman, B. D. Bower, D. J. Dismuke, M. R. Falvo, J. D. Griffith, S. C. Harvey and A. Asokan, *J. Virol.*, 2013, **87**, 2994–3002.

41 T. Kimura, B. Ferran, Y. Tsukahara, Q. Shang, S. Desai, A. Fedoce, D. R. Pimentel, I. Luptak, T. Adachi, Y. Ido, R. Matsui and M. M. Bachschmid, *Sci. Rep.*, 2019, **9**, 1–13.

42 M. Lock, M. Alvira, L. H. Vandenberghe, A. Samanta, J. Toelen, Z. Debyser and J. M. Wilson, *Hum. Gene Ther.*, 2010, **21**, 1259–1271.

43 M. R. Hossan, R. Dillon and P. Dutta, *J. Comput. Phys.*, 2014, **270**, 640–659.

44 A. Morshed, P. Dutta, M. R. Hossan and R. Dillon, *Phys. Rev. Fluids*, 2018, **3**(103702), 1–18.

45 A. Morshed, P. Dutta and M. J. Kim, *Electrophoresis*, 2019, **40**, 2584–2591.

46 J. Judd, F. Wei, P. Q. Nguyen, L. J. Tartaglia, M. Agbandje-McKenna, J. J. Silberg and J. Suh, *Mol. Ther.–Nucleic Acids*, 2012, **1**, e54.

47 J. Saharia, Y. M. N. D. Y. Bandara, J. S. Lee, Q. Wang, M. J. Kim and M. J. Kim, *Electrophoresis*, 2020, **41**, 630–637.