

# Electrophoretic Deformation of Individual Transfer RNA Molecules Reveals Their Identity

Robert Y. Henley,<sup>†</sup> Brian Alan Ashcroft,<sup>∇</sup> Ian Farrell,<sup>§</sup> Barry S. Cooperman,<sup>||</sup> Stuart M. Lindsay,<sup>⊥,∇,¶</sup> and Meni Wanunu<sup>\*,†,‡</sup>

<sup>†</sup>Department of Physics, Northeastern University, Boston, Massachusetts 02115, United States

<sup>‡</sup>Department of Chemistry/Chemical Biology, Northeastern University, Boston, Massachusetts 02115, United States

<sup>§</sup>Anima Cell Metrology, Inc., Bernardsville, New Jersey 07924, United States

<sup>||</sup>Department of Chemistry, University of Pennsylvania, Philadelphia, Pennsylvania 19104, United States

<sup>⊥</sup>Department of Physics, Arizona State University, Tempe, Arizona 85281, United States

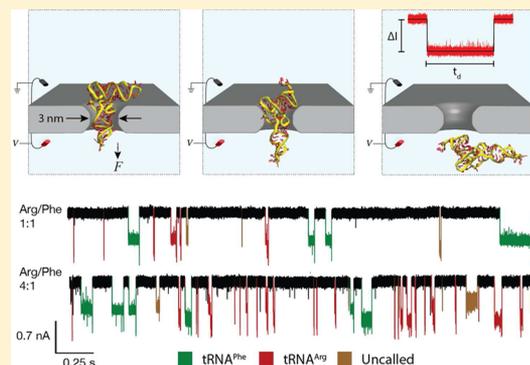
<sup>∇</sup>Biodesign Institute, Arizona State University, Tempe, Arizona 85281, United States

<sup>¶</sup>Department of Chemistry and Biochemistry, Arizona State University, Tempe, Arizona 85281, United States

## Supporting Information

**ABSTRACT:** It has been hypothesized that the ribosome gains additional fidelity during protein translation by probing structural differences in tRNA species. We measure the translocation kinetics of different tRNA species through  $\sim 3$  nm diameter synthetic nanopores. Each tRNA species varies in the time scale with which it is deformed from equilibrium, as in the translocation step of protein translation. Using machine-learning algorithms, we can differentiate among five tRNA species, analyze the ratios of tRNA binary mixtures, and distinguish tRNA isoacceptors.

**KEYWORDS:** nanopores, tRNA, single molecule, RNA, SVM



In the 1950s, Crick put forth the adaptor hypothesis to explain how genetic information contained in the cell is used to maintain and express cellular machinery. Though this hypothesis was made without any knowledge of the ribosome or other complex machinery involved in protein synthesis, it did correctly predict the existence of tRNA and its role as the molecule that links the genetic code to amino acids.<sup>1</sup> Crick's original hypothesis predicted that base-pairing interactions were responsible for all of the decoding carried out during what we now call the translation step. However, although it was confirmed that mRNA (mRNA) codes for a sequence of specific amino acids that are identified by complementary tRNA codons, the hypothesis that specificity is conferred solely by codon–anticodon pairing has been disproven.<sup>2</sup> Discoveries of the Hirsh repressor<sup>3</sup> and other mutations far from the anticodon region<sup>4–6</sup> have shown that additional factors are involved in tRNA recognition, although the nature of this added specificity is still debated. Possible sources for additional discrimination include tRNA-specific fluctuations<sup>7</sup> and flexibility,<sup>8</sup> post-transcriptional modifications,<sup>9</sup> variability in internal base-pairing within the tRNA,<sup>6</sup> and ternary interactions between tRNA, EF-Tu, and the ribosome.<sup>4,8,10–12</sup>

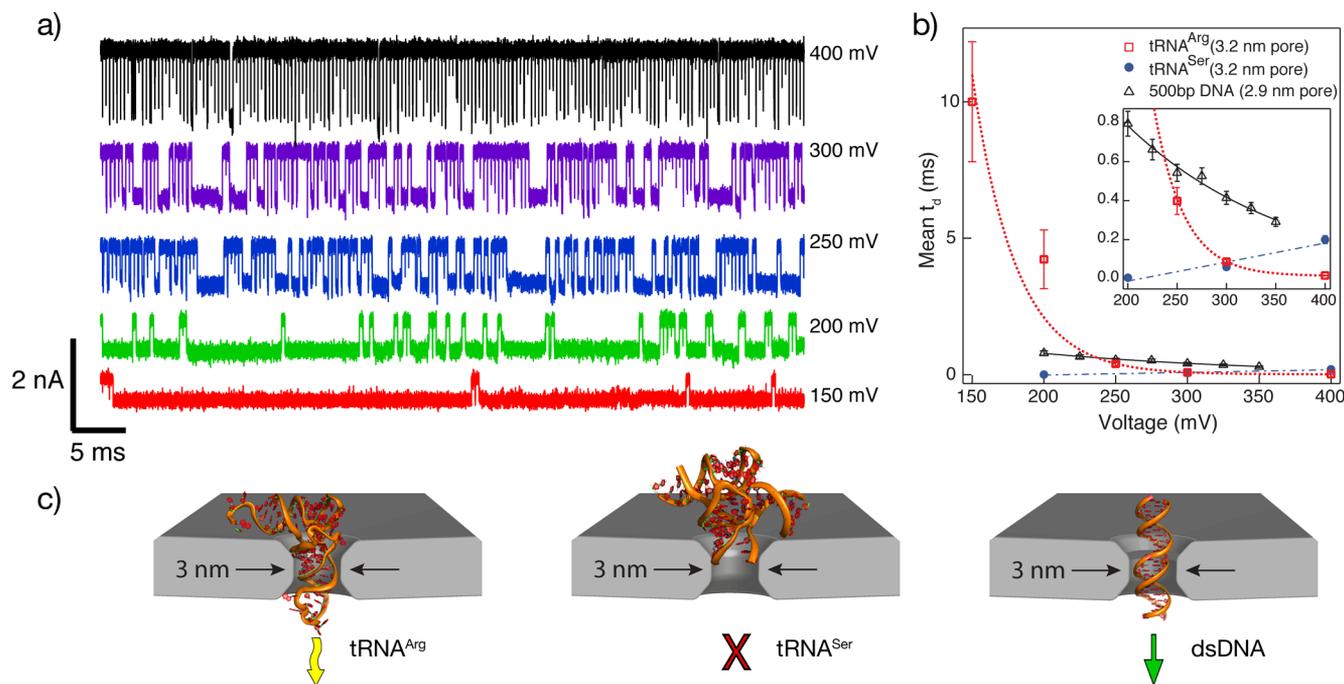
The sequence and post-transcriptional modifications of each tRNA are highly conserved within a biological species. Studies have been conducted to investigate the effects of these post-transcriptional modifications on tRNA and how these effects contribute to the fidelity of decoding.<sup>13–15</sup> A recent report suggests that each tRNA adopts a unique decoding conformation within the ribosome due to these factors.<sup>14</sup> The results of this study imply that molecular stiffness can affect the tRNA conformation and dynamics, which in turn may play a mechanistic role in enhancing recognition by the ribosome.

Although the traditional L-shaped secondary structure of tRNA is well conserved for most tRNAs, as indicated by the few available crystal structures,<sup>16–18</sup> very little is known about the mechanical properties of tRNA and how these properties vary for different tRNAs and with different post-translational modifications. Both solid-state and protein nanopores have been shown to be effective tools for studying the structural and mechanical properties of RNA molecules.<sup>19–28</sup> Nanopore-based discrimination of two tRNA species has recently been

**Received:** August 19, 2015

**Revised:** November 20, 2015

**Published:** November 26, 2015



**Figure 1.** (a) Concatenated event traces for  $\text{tRNA}^{\text{Arg}}$  displaying events at the indicated voltages. Each event is shown with  $\sim 250 \mu\text{s}$  of data before and after the event. (b) Mean dwell times as a function of applied voltage for  $\text{tRNA}^{\text{Arg}}$ . Data for  $\text{tRNA}^{\text{Ser}}$  and 500 bp DNA are included for comparison (see text). (c) Schematic showing the PDB cartoon structures of the three control molecules positioned inside a  $\sim 3 \text{ nm}$  diameter pore (yellow arrow = translocation with unbending, red cross = no translocation, green arrow = unhindered translocation).

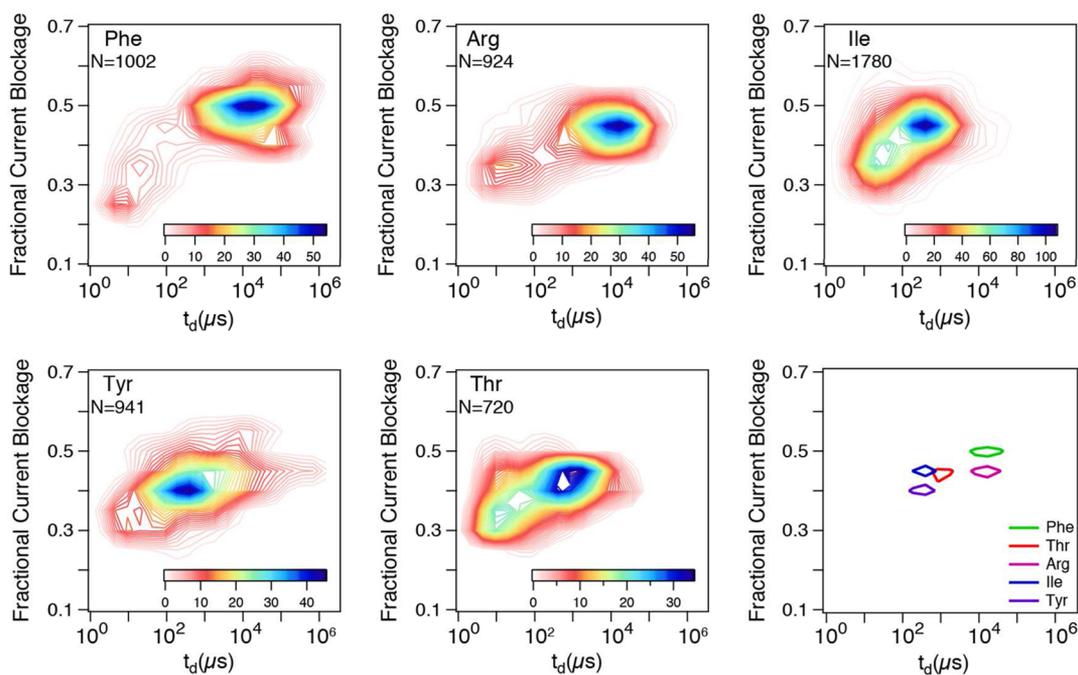
demonstrated by unzipping tRNA molecules that are ligated to a synthetic DNA adapter molecule.<sup>28</sup> In this study, a  $\phi 29$  DNA polymerase that binds to the DNA/tRNA hybrid serves as a barrier for unzipping the molecule as it is pulled through an  $\alpha$ -hemolysin nanopore. Though this technique can distinguish two tRNA species with  $\sim 87\%$  accuracy, the need for complete unzipping of the ligated tRNA molecules does not enable the study of tRNA tertiary structure in its natively folded state.

Though many works have used nanopores to study RNA, the vast majority of these did not probe natively folded RNA molecules, instead relying on linearizing the molecules in question. We have recently utilized small solid-state nanopores as a label-free single molecule method to recognize drug-induced unbending of the hepatitis C internal ribosome entry site (IRES) structural motif from its native confirmation, which disrupts its ability to bind to the ribosome.<sup>25</sup> In the current report, we perform high bandwidth (4.17 MHz sampling, 500 kHz low-pass filtering)<sup>29,30</sup> measurements of the ion current signals produced when single, natively folded tRNA molecules are electrophoretically deformed during their passage through a 3 nm pore. By employing a support-vector machine (SVM), we show that signals from tRNA translocation across the pore can be used to classify molecules with accuracies that greatly exceed random guesses. Further, the SVM can be retrained in order to obtain a similar tRNA identification quality for data acquired using different nanopores. Improvements to the SVM and experimental hardware may allow a larger set of tRNAs to be distinguished, potentially the entire complement of tRNA contained within an organism.

In order to study the flexibility of individual tRNA molecules, we fabricated nanopores with  $\sim 3 \text{ nm}$  diameters and 5–10 nm effective thickness (for full details please refer to the [Materials and Methods](#) section). This geometry is wide enough to easily accommodate the tRNA acceptor stem, but full passage

through the pore requires deformation at the tRNA elbow region. We employed a transmembrane voltage bias to electrophoretically drive tRNA molecules through the nanopore. Primary analysis of the nanopore data is performed by identifying the stochastic set of downward pulses that correspond to tRNA passage through the pore. A pulse can be described by its dwell time,  $t_d$ , and the mean blockade amplitude,  $\Delta I$ . We anticipate that the more structurally stable tRNAs will exhibit longer dwell times, as conformational changes are required to allow passage through the pore. However, aside from dwell time and current blockade amplitude, there are other indicators of molecular properties that may be contained within the signal during tRNA immobilization in the pore (and not specifically contained in  $t_d$  or  $\Delta I$ ). Therefore, we employ machine-learning techniques to identify additional, unique signal features for each tRNA pulse that include descriptors of the overall peak shape and portions of the noise spectrum within the blockade pulse. The methodology for implementation of these techniques/algorithms has been described fully elsewhere.<sup>31,32</sup>

**Device Characterization.** To illustrate the steep dependence of electrophoretic force on tRNA deformation kinetics, we first measured the dwell time dependence of  $\text{tRNA}^{\text{Arg}}$  on applied voltage in the range 150–400 mV (experiment performed in order of decreasing voltage). [Figure 1a](#) displays a concatenated representative set of pulses detected at different applied voltages. Clearly,  $\text{tRNA}^{\text{Arg}}$  dwell times drastically increase as voltage is decreased. We find that the tRNA dwell times fit log-normal distributions (see [Figure SI2](#) for distributions and fits), and we use these fits to find the most-probable  $t_d$  values, which are plotted in [Figure 1b](#). For comparison, we also plot data obtained for *S. cerevisiae*  $\text{tRNA}^{\text{Ser}}_{\text{GCU}}$  and a 500 bp sample of linear double-stranded DNA (dsDNA). As already previewed in [Figure 1a](#), for the



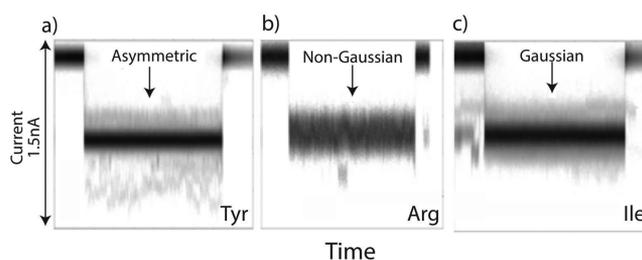
**Figure 2.** Contour plots depicting the distributions of fractional current blockage vs  $t_d$  for five tRNA samples electrophoresed through the same nanopore ( $\sim 3$  nm diameter,  $\sim 10$  nm thickness) using an applied voltage of 200 mV. The final panel shows the locations of the contour peaks for each tRNA.

tRNA<sup>Arg</sup> sample, we find a strong exponential dependence of  $t_d$  on voltage, indicative of a large free energy barrier associated with the bending of a tRNA molecule as it is electrophoretically forced through the pore.<sup>33</sup> In contrast, tRNA<sup>Ser</sup>, which possesses a long (14-nucleotide) variable loop,<sup>34</sup> exhibits dwell times that increase with applied voltage.<sup>35</sup> This behavior, reproduced several times in different pores with similar geometry, is characteristic of failed attempts to cross the pore due to steric constraints. Finally, as expected, dsDNA dwell times decrease with voltage, as shown in a prior study using a similar geometry pore.<sup>36</sup> However, the voltage dependence on dwell times for DNA is much less pronounced than for tRNA<sup>Arg</sup>, which suggests that tRNA transport is impeded by the free energy barrier associated with deformation. Figure 1c schematically depicts PDB-derived cartoons of the three molecules superimposed with the nanopore geometry (DNA length is truncated), indicating the proposed mechanisms that are suggested by our voltage dependence measurements.

**Measuring Stability of tRNA Species.** Next, we examined the transport of five different tRNAs using an applied voltage of 200 mV, at which the force is large enough to efficiently drive tRNAs into the nanopore but low enough to produce long dwell times that can aid in molecular identification (see Figure S13 for sample events). Two-dimensional contour plots of fractional current blockade ( $\Delta I/I_0$ , where  $I_0$  is the open pore current) vs dwell time are shown for the five different tRNAs, along with a graph that compares the most-probable position for each tRNA (see Figure 2). The plots show substantial differences in the most-probable dwell times for all tRNAs, which we attribute to different degrees of stability intrinsic to each tRNA, (tRNA<sup>Phe</sup> > tRNA<sup>Arg</sup> > tRNA<sup>Thr</sup> > tRNA<sup>Ile</sup>  $\approx$  tRNA<sup>Tyr</sup>). Further, the relative peak positions were consistent across multiple experiments using different pores at different voltages (see Figure S14 for example), with peak locations being independent of the order in which the molecules were tested.

However, though each of the tRNAs tested has its characteristic dwell time and current blockade level, the large levels of overlap in populations indicate that these two parameters are insufficient for accurate single pulse identification of individual tRNAs.

**SVM Based Single-Pulse tRNA Speciation.** To illustrate the limitations of two-parameter analysis, we present in Figure 3 sets of overlaid current pulses for three different tRNA



**Figure 3.** Noise characteristics for different tRNA species. Translocation events are overlaid (with up to 10% of time-rescaling to align the start and end of each event) and the gray scale density corresponds to the number of overlapping data points. (a) tRNA<sup>Tyr</sup> displaying asymmetric noise,  $n = 666$ ,  $\langle t_d \rangle = 0.5 \pm 0.05$  ms, (b) tRNA<sup>Arg</sup> displaying non-Gaussian noise,  $n = 551$ ,  $\langle t_d \rangle = 5 \pm 0.5$  ms, (c) tRNA<sup>Ile</sup> displaying Gaussian noise,  $n = 746$ ,  $\langle t_d \rangle = 0.7 \pm 0.07$  ms.

species, where signal intensity represents the relative frequency (time coordinates of pulses was rescaled to allow overlay). The blurry appearance of the images is due to incomplete overlap of data points in the sets. Clearly, the noise within the pulse is larger than baseline current noise, and is a function of the analyte type (asymmetric for tRNA<sup>Tyr</sup><sub>QUA</sub>, non-Gaussian for tRNA<sup>Arg</sup> and Gaussian for tRNA<sup>Ile</sup><sub>UAU</sub>). In addition to the visible differences among the tRNA signals in Figure 3, there are other parameters contained within the signals that are not necessarily distinguishable in raw data format. Therefore, we

applied a support vector machine (SVM) to the data, as previously reported for analysis of tunneling signals obtained from amino acids<sup>31</sup> and DNA bases.<sup>32</sup> The first step was to characterize the events for the machine analysis. A library of vectors (e.g., signal features) for describing peaks has been developed<sup>33</sup> for tunnel current signals that describe the shape of the peaks. These features include, for example, signal amplitudes, noise spectra, and Fourier components. Using a shotgun method, we determined the most powerful features for describing the differences in tRNAs, with the criterion that the distributions of a given feature vary little between different experiments on the same analyte, while varying significantly when compared to distributions obtained from other analytes (i.e.,  $\sigma_{\text{intragroup}}/\sigma_{\text{intergroup}} < 0.8$  where  $\sigma$  is the standard deviation). This criterion eliminates signal features that are subject to systematic error as well as features that are insensitive to analyte chemistry.

Once a number of candidate parameters have been selected, we trained the SVM using a representative fraction of the data sets and evaluated the training accuracy using the remainder of the data sets. We then employed adaptive feature selection<sup>37</sup> to determine which signal features contribute most to the correct assignment of an event to a specific tRNA. In this method, features are removed from the analysis a few at a time in order to determine the most essential combination of features.

We consistently found that a subset of the acquired data is common to all tRNA species and cannot be classified easily by the SVM. These data include noise, pore blockade events, and what are most likely unsuccessful attempts of the tRNA to cross the membrane. Thus, some level of data filtering is required, as follows: The first stage of filtering uses an overlapping one-class SVM to determine which events are similar to all the others. In this method, each analyte is used to train a one-class SVM, and then all the data are run through the entire set of one-class SVM's. Data that fall into nearly all of the classes, which comprises 30–60% of the candidate events, are suspect and removed. To create a second filter, a single one-class SVM algorithm was trained with all the known training data. This filter, which eliminates another 10% of the candidate events, makes it possible to identify events that are unknown with respect to the training set. Finally, we found that a major contribution to systematic errors came from time-dependent drift in the pore characteristics. Therefore, a least-squares regression algorithm was implemented for time correction. Figure S15 shows the distributions of two signal features (pulse amplitude and peak width) for tRNA<sup>Arg</sup> as measured over the course of an hour. The distributions noticeably drift with time; this is likely a result of slight changes in the pore geometry over time. Accordingly, we developed a linear interpolation algorithm to correct the distribution changes using data recorded from one analyte as a function of time. The feature selection routine was also made more sensitive to those parameters that vary with time, allowing removal of those that perform badly after the drift removal routine. A final and important factor to consider is that each pore produces its own noise signal in interaction with the target molecules. This framework is not general enough to allow the SVM training from one nanopore to be used on another nanopore. Each nanopore requires its own training and selection of features, and all experiments in this paper were handled in accordance with this limitation.

Table 1 details the accuracies with which the SVM can identify an analyte from a single pulse. Random calling would

**Table 1. Accuracies with Which Each Individual Event Is Assigned from the Pool of Five tRNAs, Both with and without Drift Correction.<sup>a</sup>**

tRNA species	uncorrected accuracy (%)	time-corrected accuracy (%)
all (average)	39	69
Arg	55	80
Tyr	18	61
Ile	60	84
Phe	2	39
Thr	13	51

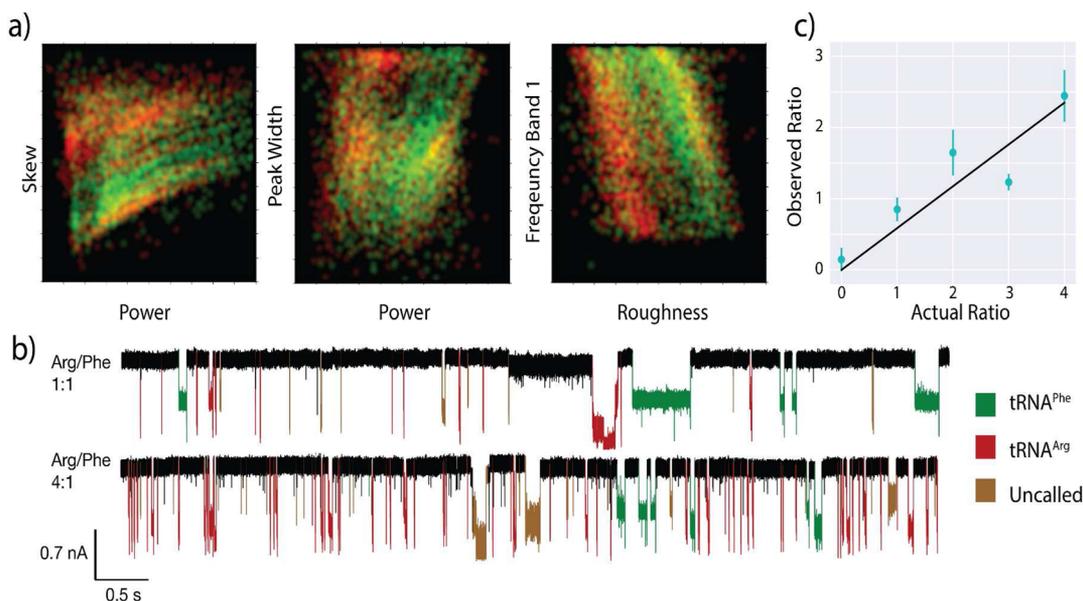
<sup>a</sup>The signal features used in this analysis are listed in Table S11.

result in 20% of the assignments being called correctly, but as can be seen here, the SVM can attain an average single-pulse accuracy (with drift correction) of 69%. This corresponds to an accuracy of 88% if analysis is constrained to pairs of tRNAs, comparable to the accuracies reported by Smith et al.<sup>28</sup> Thus, each single molecule event can be called with accuracy significantly better than random without resorting to statistical analysis of many events to distinguish between analytes. This is a powerful consequence of the use of machine learning.

**Determining Contents of a tRNA Mixture.** As a proof-of-principle application of this approach, we measured concentration ratios of binary tRNA mixtures. This experiment serves as validation of the SVM-based assignment of individual events. Though SVM studies of amino acids using tunneling signals have shown exceptionally high signal calling accuracies, quantitative mixture analysis has been difficult to achieve, likely due to intermolecular interactions resulting from high analyte concentrations.<sup>31</sup> To test the limits of our detection method, we selected two tRNAs with the highest degree of overlap in their dwell time and current blockade values. As a result, the two tRNA species are differentiated based upon signal components that, with rare exception,<sup>38</sup> have been largely ignored in previous nanopore experiments.

To perform discrimination in a binary tRNA mixture, tRNA<sup>Phe</sup> was first added to our standard buffer to a final concentration of ~40 nM. Next, tRNA<sup>Arg</sup> was titrated into the buffer in 40 nM increments up to a final concentration of 160 nM (corrected for dilution); data points were acquired with a bias voltage of 400 mV following each titration step. Then, machine-learning algorithms were applied to these traces to assign identity to each event in the trace. By calculating the assigned ratios of tRNA<sup>Arg</sup> to tRNA<sup>Phe</sup> events, apparent analyte molar ratios were measured. Presumably due to differences in capture rate efficiencies for the two molecules, apparent and actual molar ratios are unlikely to have 1:1 correspondence, but linear increases in molar ratio should result in linear increases in the apparent ratios. This makes it possible to obtain the molar ratio of unknown samples by calibrating the system with a mixture sample of a known ratio.

A standard comparison of dwell times and current blockades shows little to no difference between the two tRNAs (see Figure S16) as previously described. Clearly, SVM analysis is required for accurate single-molecule discrimination. The SVM was trained with pure samples presented to the nanopores both before and after the mixtures were run, and then it was tested on the mixtures. Examples of the different SVM parameters used and the separation they provide are shown in Figure 4a. These color plots show probability densities for two analytes (green tRNA<sup>Phe</sup>, red tRNA<sup>Arg</sup>) for two parameter values (listed on the axes of the plots and described in the caption). These

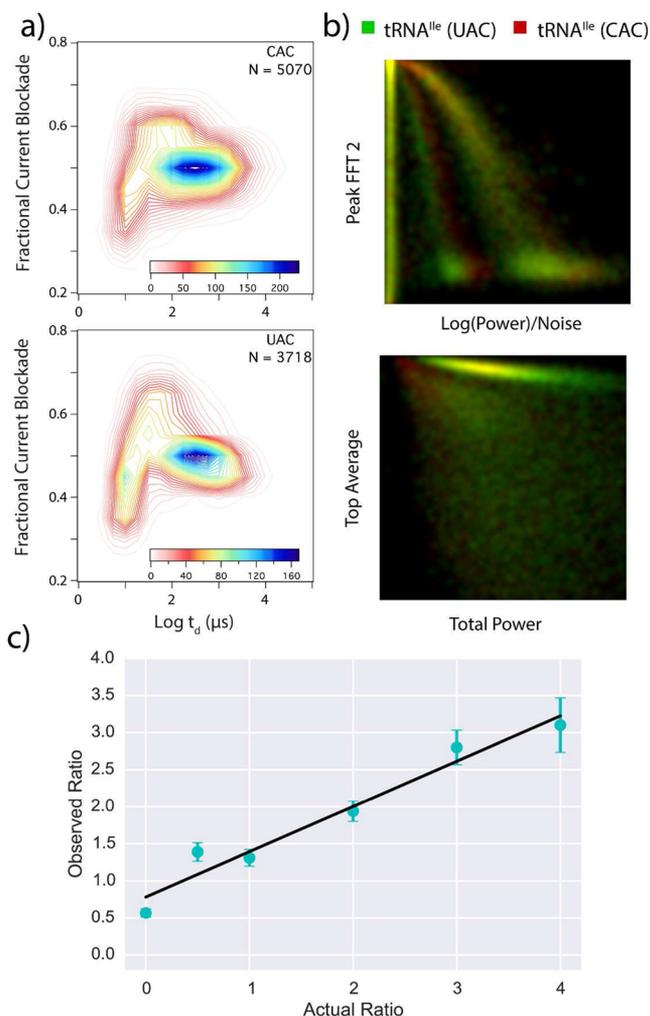


**Figure 4.** Detection of tRNA species in binary mixtures. (a) Several of the SVM signal features displayed in false-color 2D plots (see text and Table S12). All units are normalized and shown from 0 to 1. tRNA<sup>Phe</sup> is shown in red, tRNA<sup>Arg</sup> in green, and areas of overlap are shown in yellow. Signal feature definitions: skew (statistical skew of a peak), power (integrated area of event current), peak width (total time of event), frequency band 1 (noise spectrum between 0–208 kHz), roughness (standard deviation of current values within an event). (b) Current vs time traces for both the 1:1 and 4:1 tRNA<sup>Arg</sup>/tRNA<sup>Phe</sup> mixtures. Events are colored according to SVM calling. The number of tRNA<sup>Arg</sup> calls increases greatly with its relative concentration, whereas the number of tRNA<sup>Phe</sup> does not appreciably change. (c) Measured ratio of tRNA<sup>Arg</sup> to tRNA<sup>Phe</sup> signals vs the known molar ratio in solution. The ratios do not show 1:1 correspondence due to differences in the capture efficiency for the two molecules, but the linear correlation shows that the identity of single molecule events can be assigned with reasonable accuracy. Error bars were assigned by dividing each data set into many different subsets. The subsets were divided by time within the experiment to identify sampling errors, and then again using a number of different feature sets. The comparison between all the different assigned ratios determines the classification error, and controls for the standard problems of over- and underfitting of the SVM.

are measured distributions, where the probability density at any pair of values of the two parameters is indicated by the brightness of the point. The plots are color-mixed, so that nonoverlapped data appear as pure red or pure green, whereas overlapped data generate yellow points. The SVM operates with many parameters simultaneously, dividing up the corresponding multidimensional space into “red” or “green” regions using multidimensional partitions (the support vectors). In the analysis shown, seven features were used to attain an accuracy (true/positive rate) of 91% for tRNA<sup>Arg</sup> and 82% for tRNA<sup>Phe</sup> (see Table S12 for a full list of parameters). These 2D comparisons provide better separation between the two analytes than the traditional dwell time and current blockade. The SVM is able to separate these parameters in 7-dimensional space, leading to our exceptional accuracy numbers. Current traces with the resulting SVM calling identity can be seen in Figure 4b. The measured tRNA ratios correlate well with the known compositions of the mixtures (Figure 4c). As expected, the slope of the line is not equal to 1.0, reflecting differences in capture efficiency for the two molecules, though this slope may also be affected by pipetting errors.

**Differentiation of tRNA Isoacceptors.** As a final exercise the approach described here was tested for its ability to discriminate between tRNA isoacceptors. For this study, we collected translocation data from tRNA<sup>Val</sup><sub>UAC</sub> and tRNA<sup>Val</sup><sub>CAC</sub>, followed by translocation data from mixtures of the two analytes in various ratios as described previously. Data from the individual analytes are then taken once more to calibrate the time-correction algorithm of the SVM (examples of the time evolution of parameters are presented in Figure S17. Figure 5a shows data collected from each of the pure control samples.

The two molecules are virtually indistinguishable using the standard dwell time vs current blockade analysis. In contrast, Figure 5b shows examples of the separation attained by 2D projections of some of the parameters used in the SVM analysis. Though these projections still do not show perfect separation of the two tRNAs, particular regions where the tRNA identity can be distinguished do begin to appear. These two isoacceptors do not separate well with any two parameters, but by projecting the data into 6-dimensional space the SVM analysis is able to achieve superior separation. After calibrating on a subset of the data from the two pure isoacceptors, the remainder of the pulses from the two samples are randomized to test the accuracy of the SVM. When the calibrated SVM is given a single translocation pulse it is able to identify the corresponding isoacceptor with an accuracy of 93% (see Table S13 for list of SVM parameters used for mixture analysis). After the system is calibrated with the pure samples the SVM is tested for its ability to sense differences in analyte ratios from several different mixtures. The results of the mixture analysis can be seen in Figure 5c. Once again, the called ratio does not exhibit a 1:1 correspondence with the actual ratio. This difference is likely due to differences in capture rates for the two analytes. For  $\sim 1$  ng/ $\mu$ L analyte concentrations, capture rates of 166 events/s and 112 events/s were observed for CAC and UAC, respectively. The capture rate for the UAC isoacceptor in the pure tRNA experiments was found to be  $\sim 0.67\times$  the capture rate of the CAC isoacceptor. This is comparable to the slope of 0.61 displayed by the SVM predicted ratio vs actual ratio plot in Figure 5c, though the capture rates may also be skewed by pipetting errors and solution heterogeneity.



**Figure 5.** Comparison of the UAC and CAC tRNA<sup>Ile</sup> isoacceptors. (a) Contour plots showing both isoacceptors, >3000 events were recorded for each at an applied voltage of 500 mV. (b) Two-dimensional projections using four of the SVM features. All units are normalized and shown from 0 to 1. UAC (green) and CAC (red). SVM analysis yields calling accuracy of 93% for individual pulses. Signal feature definitions: peak FFT 2 (power spectrum value at 18 kHz), log(power)/noise (the log of the sum of the power spectrum, divided by the noise spectrum of the nanopore system), top average (sum of the ionic current values within a translocation pulse), total power (log of the sum of the power spectrum) (c) Measured ratio of UAC to CAC signals vs the known molar ratio in solution. The ratios do not show 1:1 correspondence due to differences in the capture efficiency for the two molecules, but the linear correlation convincingly shows that the identity of single molecule events can be assigned). The error in the ratios was determined by segmenting the peaks by time ( $\sim 5$  s intervals). The ratios between the various analytes were then determined for each segment, normalized and the percent error was determined.

We have shown that each tRNA species tested here displays a characteristic peak in its dwell time value and that the relative position of these peaks remains consistent across all devices. In addition, SVM studies show that each specific tRNA produces a characteristic noise spectrum upon passage through the nanopore, possibly due to nanosecond to microsecond thermal fluctuations that are characteristic of specific tRNAs.<sup>7</sup> The results demonstrate that solid-state nanopores may be used as an effective tool for studying differences in tRNA flexibility.

Computer simulations may reveal a relationship between tRNA dynamics and dwell times. The results also indicate that the ribosome might monitor tRNA flexibility in achieving translational fidelity. In this connection, it would be of interest to determine whether there is a link between dwell time and the angle formed by the elbow region of ribosome-bound tRNA. With further studies on a complete set of tRNA species, it may be possible to identify the sources the different types of noises we have observed (asymmetric, Gaussian, non-Gaussian) with specific tRNA structural and dynamic features (vibrational modes, post-transcriptional modifications, etc.).

Finally, we have shown by using machine-learning algorithms and probing many different parameters of the ionic current that we can identify individual tRNA molecules as they translocate pores. These algorithms greatly increase the ability of the device to classify even single blockade events, allowing classification accuracies up to 90%. The time correction allows classification to occur even with changes in the local reading environment of the pore. This system is effective for differentiating between tRNA isoacceptors that are charged by the same amino acid and could possibly be extended to resolve tRNA isodecoders, as well as identical tRNAs that are posttranscriptionally modified to different extents. Our technique can be generalized to many other sensing modalities where signals exhibit time-dependent drift. Further optimization and accuracy could be achieved by utilizing a number of features that very specifically target data generated by tRNA translocation events. Additionally, adding Mg<sup>2+</sup>, which has been shown to increase tRNA stability and prolong the nanopore translocation process,<sup>28</sup> would increase the amount of data that could be acquired for each translocation.

**Materials and Methods.** Nanopores were fabricated on  $\sim 20$  nm thick silicon nitride freestanding membranes supported by a Si chip using a JEOL 2010FEG transmission electron microscope, as previously described.<sup>39</sup> Fabricated nanopores were cleaned using hot piranha solution and, following copious rinsing with deionized water, were assembled in a custom PTFE cell equipped with homemade Ag/AgCl electrodes.

To provide a seal for the device and reduce device capacitance, we applied an elastomeric gasket between the chip and cell.<sup>40</sup> We introduced solution containing 0.40 M KCl, 1 mM EDTA, and 10 mM Tris-buffered to pH 7.9 to both sides of the chip such that the nanopore forms a fluid channel joining the two chambers. In order to ensure a steady baseline current, we measured current traces for each device before the addition of tRNA molecules. Next, we added a purified sample of an individual tRNA species to the buffer on the *cis* side and applied a positive bias to the *trans* side while continuously monitoring the ion current until we detected a substantial number of molecular deformation events ( $n > 719$ ). We then thoroughly flushed the *cis* chamber to remove all tRNA as indicated by absence of events. We then repeated the process for all tRNA samples. Data was acquired at 4.17 MHz using a Chimera Instruments VC100 and low-pass filtered at 500 kHz to provide a suitable signal-to-noise ratio for event detection.

Current blockage and dwell time analysis of the traces were performed using Python, custom software written in the Wanunu lab (<https://github.com/rhenley/Pyth-Ion/>).

SVM data analysis was performed using custom code written in the Lindsay lab. The code is available online at [https://github.com/ochensati/SVM\\_DNA\\_TunnelVision](https://github.com/ochensati/SVM_DNA_TunnelVision)

In the experiments that contain mixtures of different tRNAs, we first obtained characteristic traces for tRNA<sup>Arg</sup> (mixture of isoacceptors) and tRNA<sup>Phe</sup><sub>GAA</sub>, both isolated from *S. cerevisiae*. We then collected data from different molar mixtures of the two samples and implemented our SVM algorithms to detect changes in the relative concentrations based on single-pulse tRNA calling.

Samples of tRNA<sup>Arg</sup> (a mixture of isoacceptors) from *S. cerevisiae*, and tRNA<sup>Tyr</sup><sub>QUA</sub> from *E. coli* were purchased from Chemical Block. tRNA<sup>Phe</sup><sub>AAG</sub> from *S. cerevisiae* was purchased from Sigma-Aldrich. tRNA<sup>Ile</sup><sub>UAU</sub>, tRNA<sup>Thr</sup><sub>CGU</sub>, tRNA<sup>Ser</sup><sub>GCU</sub>, tRNA<sup>Val</sup><sub>UAC</sub>, and tRNA<sup>Val</sup><sub>CAC</sub> were isolated from bulk *S. cerevisiae* tRNA (Roche Diagnostics) following procedures described in Barhoom et al., 2013. Analytical scale aminoacylations with radioactive amino acids resulted in the following aminoacylation efficiencies: 0.54 Arg/tRNA<sup>Arg</sup>, 0.45 Tyr/tRNA<sup>Tyr</sup>, 0.28 Phe/tRNA<sup>Phe</sup>, 0.68 Ile/tRNA<sup>Ile</sup>, 0.62 Thr/tRNA<sup>Thr</sup>, 0.72 Ser/tRNA<sup>Ser</sup><sub>GCU</sub>, 0.69 Val/tRNA<sup>Val</sup><sub>UAC</sub>, and 0.63 Val/tRNA<sup>Val</sup><sub>CAC</sub>. Purity and intactness of the tRNAs were verified by denaturing gel electrophoresis (Figure S19).

## ■ ASSOCIATED CONTENT

### ■ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.nanolett.5b03331.

Supporting Information contains experimental schematic, details on voltage dependence of tRNA molecules, sample events, additional data sets, time dependence of various parameters, raw current traces, full list of SVM parameters for each investigation, and PAGE gel of five tRNAs. (PDF)

## ■ AUTHOR INFORMATION

### ■ Corresponding Author

\*E-mail: wanunu@neu.edu. Fax: (617) 373 2943.

### ■ Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

We acknowledge support from the National Institutes of Health for this work (HG006873, MW; HG 006323, SML; GM080376, BSC).

## ■ REFERENCES

- (1) Crick, F. H. *Symp. Soc. Exp. Biol.* **1958**, *12*, 138–63.
- (2) Loftfield, R. B. *Biochem. J.* **1963**, *89* (1), 82–92.
- (3) Hirsh, D. *Nature* **1970**, *228* (5266), 57.
- (4) Ortiz-Meoz, R. F.; Green, R. *RNA* **2010**, *16* (10), 2002–13.
- (5) Schultz, D. W.; Yarus, M. *J. Mol. Biol.* **1994**, *235* (5), 1381–94.
- (6) Smith, D.; Yarus, M. *J. Mol. Biol.* **1989**, *206* (3), 503–511.
- (7) Lee, T. H.; Blanchard, S. C.; Kim, H. D.; Puglisi, J. D.; Chu, S. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (34), 13661–5.
- (8) Cochella, L.; Green, R. *Science* **2005**, *308* (5725), 1178–1180.
- (9) Dineshkumar, T. K.; Thanedar, S.; Subbulakshmi, C.; Varshney, U. *Microbiology* **2002**, *148* (12), 3779–3787.
- (10) Schrader, J. M.; Uhlenbeck, O. C. *Nucleic Acids Res.* **2011**, *39* (22), 9746–58.
- (11) Yikilmaz, E.; Chapman, S. J.; Schrader, J. M.; Uhlenbeck, O. C. *Biochemistry* **2014**, *53* (35), 5710–20.
- (12) Schrader, J. M.; Chapman, S. J.; Uhlenbeck, O. C. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* (13), 5215–20.
- (13) Novoa, E. M.; Pavon-Eternod, M.; Pan, T.; Ribas de Pouplana, L. *Cell* **2012**, *149* (1), 202–13.
- (14) Schmeing, T. M.; Voorhees, R. M.; Kelley, A. C.; Ramakrishnan, V. *Nat. Struct. Mol. Biol.* **2011**, *18* (4), 432–6.
- (15) Cantara, W. A.; Bilbille, Y.; Kim, J.; Kaiser, R.; Leszczynska, G.; Malkiewicz, A.; Agris, P. F. *J. Mol. Biol.* **2012**, *416* (4), 579–97.
- (16) Shi, H.; Moore, P. B. *RNA* **2000**, *6* (08), 1091–1105.
- (17) Basavappa, R.; Sigler, P. B. *EMBO J.* **1991**, *10* (10), 3105–11.
- (18) Rich, A.; RajBhandary, U. L. *Annu. Rev. Biochem.* **1976**, *45* (1), 805–60.
- (19) Cracknell, J. A.; Japrun, D.; Bayley, H. *Nano Lett.* **2013**, *13* (6), 2500–5.
- (20) van den Hout, M.; Skinner, G. M.; Klijnhout, S.; Krudde, V.; Dekker, N. H. *Small* **2011**, *7* (15), 2217–24.
- (21) Butler, T. Z.; Gundlach, J. H.; Troll, M. A. *Biophys. J.* **2006**, *90* (1), 190–9.
- (22) Kasianowicz, J. J.; Brandin, E.; Branton, D.; Deamer, D. W. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93* (24), 13770–13773.
- (23) van den Hout, M.; Vilfan, I. D.; Hage, S.; Dekker, N. H. *Nano Lett.* **2010**, *10* (2), 701–7.
- (24) Wanunu, M.; Dadosh, T.; Ray, V.; Jin, J.; McReynolds, L.; Drndic, M. *Nat. Nanotechnol.* **2010**, *5* (11), 807–14.
- (25) Shasha, C.; Henley, R. Y.; Stoloff, D. H.; Rynearson, K. D.; Hermann, T.; Wanunu, M. *ACS Nano* **2014**, *8* (6), 6425–30.
- (26) Wanunu, M.; Bhattacharya, S.; Xie, Y.; Tor, Y.; Aksimentiev, A.; Drndic, M. *ACS Nano* **2011**, *5* (12), 9345–53.
- (27) Gerland, U.; Bundschuh, R.; Hwa, T. *Phys. Biol.* **2004**, *1* (1–2), 19–26.
- (28) Smith, A. M.; Abu-Shumays, R.; Akeson, M.; Bernick, D. L. *Front. Bioeng. Biotechnol.* **2015**, *3*, 91.
- (29) Balan, A.; Machielse, B.; Niedzwiecki, D.; Lin, J.; Ong, P.; Engelke, R.; Shepard, K. L.; Drndic, M. *Nano Lett.* **2014**, *14* (12), 7215–20.
- (30) Venta, K.; Shemer, G.; Puster, M.; Rodriguez-Manzo, J. A.; Balan, A.; Rosenstein, J. K.; Shepard, K.; Drndic, M. *ACS Nano* **2013**, *7* (5), 4629–36.
- (31) Zhao, Y.; Ashcroft, B.; Zhang, P.; Liu, H.; Sen, S.; Song, W.; Im, J.; Gyrfas, B.; Manna, S.; Biswas, S.; Borges, C.; Lindsay, S. *Nat. Nanotechnol.* **2014**, *9* (6), 466–73.
- (32) Chang, S.; Huang, S.; Liu, H.; Zhang, P.; Liang, F.; Akahori, R.; Li, S.; Gyrfas, B.; Shumway, J.; Ashcroft, B.; He, J.; Lindsay, S. *Nanotechnology* **2012**, *23* (23), 235101.
- (33) Mathe, J.; Visram, H.; Viasnoff, V.; Rabin, Y.; Meller, A. *Biophys. J.* **2004**, *87* (5), 3205–12.
- (34) Heyman, T.; Agoutin, B.; Fix, C.; Dirheimer, G.; Keith, G. *FEBS Lett.* **1994**, *347* (2–3), 143–146.
- (35) Wanunu, M.; Sutin, J.; McNally, B.; Chow, A.; Meller, A. *Biophys. J.* **2008**, *95* (10), 4716–25.
- (36) Carson, S.; Wilson, J.; Aksimentiev, A.; Wanunu, M. *Biophys. J.* **2014**, *107* (10), 2381–93.
- (37) Grandvalet, Y.; Canu, S. Adaptive Scaling for Feature Selection in SVMs. *Adv. Neural Inf. Process. Syst.* **2002**, 553–560.
- (38) Vercoutere, W.; Winters-Hilt, S.; Olsen, H.; Deamer, D.; Haussler, D.; Akeson, M. *Nat. Biotechnol.* **2001**, *19* (3), 248–252.
- (39) Kim, M. J.; Wanunu, M.; Bell, D. C.; Meller, A. *Adv. Mater.* **2006**, *18* (23), 3149–3153.
- (40) Tabard-Cossa, V.; Trivedi, D.; Wiggin, M.; Jetha, N. N.; Marziali, A. *Nanotechnology* **2007**, *18* (30), 305505.