Electronic Supplementary Material (ESI) for Lab on a Chip

Supplementary Information for

Enhanced discrimination of DNA molecules in nanofluidic channels through multiple measurements

Buffer used for the experiment

The buffer used in the experiment was diluted from 1X Phosphate Buffer Saline (10g NaCl, 0.25g KCl, 1.8 g Na_2HPO_4, 0.3 g KH_2PO_4, 1000 ml H_2O) at pH=7.4 by a factor of 1/15. The resultant solution had an estimated conductivity of $\sigma = 0.1441$ S/m. The conductivity of the buffer was calculated from first principles using the electrophoretic mobilities of the ions (5.4×10^{-8} $m^2/V.s$ for sodium ions and 8.1×10^{-8} $m^2/V.s$ for potassium ions). The concentration of NaCl was roughly 11 mM, which is equivalent to 7.7 mM KCl based on the ratios of their electrophoretic mobilities. By choosing a relatively low ionic concentration, the device yields significant current increase while having negligible current decrease due to volume occupancy effect[1].

Choice of voltage bias

Since the translocation amplitude is proportional to the applied voltage, an applied voltage of 1 Volt was chosen to maximize the translocation amplitude. The resultant translocation duration (roughly 4ms, sampling rate of 20 kHz) and the current change of 20~50 pA were large enough for triggering voltage reversal in LabVIEW. The large voltage (1V) also produced a high electric field, which was necessary for a high recapture probability. Calculation of the recapture radius is given in the next section.

Device design rationale

The device geometry was designed to yield (1) detectable DNA translocation current signals, and (2) a *sufficiently strong* electric field outside the nanochannel to enable a high recapture probability.

For the first criterion, the resistance of the nanochannel had to be much larger than the resistance of the microchannel so that the majority of the voltage drop would be dissipated across the nanochannel (Equation 1 in the main text). The dimensions of microchannel (0.8 cm × 1 mm × 10 μm, length by width by height) led to an electrical resistance roughly 50 times smaller than that of nanochannel (4μm × 200 nm × 500 nm).

For the second criterion, the geometry was designed to have an electric field that dominates over diffusion at relevant distances away from the nanochannel entry. Since the cross-sectional area of nanochannel is an order of magnitude smaller than that of connecting microchannel, the distribution of electric field around the nanochannel

entrance was assumed to be spherically symmetric. Under this assumption, we used the calculation from Golovchenko et al., and define the location where the combined electrical velocity equals to diffusion velocity as "recapture radius"[2]. Since the device geometry is ¼ of a sphere (instead of ½ because the nanochannel is at the bottom of the microchannel), the radius of recapture, $R = \mu I/\sigma\pi D = 88\ \mu m$, with I = 3.7 nA (with all other variables the same as used in the main text). This radius is much larger than the microchannel height. When the distance away from the nanochannel exceeds that of the height of microchannel, the 3-D spherical environment becomes 2D, which confines the electric field for recapturing the molecule. Therefore, the large recapture radius, and the 2D confinement when the distance away from the channel exceeds the channel height, are indicators of a high recapture probability.

Statistical analysis of translocation events

The goal of this section is to provide statistical verification of the hypothesis that each series contains measurements on a single DNA molecule. The verification procedure involved performing paired t-tests to test for (1) differences in mean translocation between series, and (2) the homogeneity of measurements within each series.

Homogeneity within a series means that the series is a set of measurements on **one and only one** type of molecule. Thus, testing homogeneity requires characterizing the probability that the series switches from sampling the translocation current distribution of one molecule to sampling the translocation current distribution of another molecule. In particular, we take advantage of the fact that there are two types of molecules, λ-DNA and T7 DNA, which have distinct translocation current distributions.

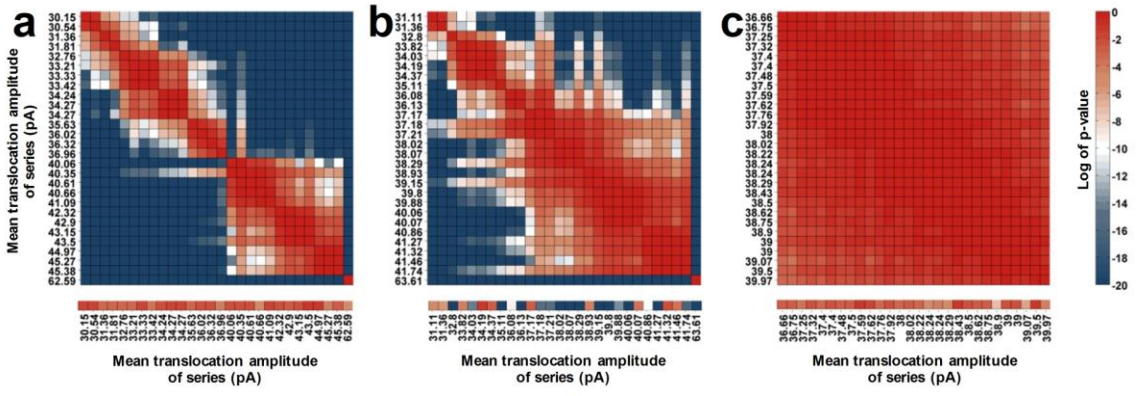
To test the homogeneity of a given series of N measurements, it was divided into pairs of sub-series {1,2,…10},{11,12,…N}; {1,2,…11},{12,13,…N}; … ; {1,2,…N-10},{N-9,N-8,…N}. For each pair, we tested for statistical differences between the mean of the sub-series (i.e. compare the means of the subseries {1,2,…10} and the subseries {11,12,…N}). If the molecule switched with another molecule of a different type after *N-n* measurements, we expect there to be a statistical difference in the means of the two subseries{1,2,...N-n}, {N-n+1,N-n+2,…N}. For each series, there are *N -1* pairs of t-tests that can be performed. In practice, we required that each sub-series have a minimum of 10 data points so that the distribution of the sub-series was well defined. Thus, N-21 t-tests were performed.

The statistical test used was a paired t-test. The null hypothesis for each of the paired t-tests was that there is no significant difference in the mean value of the translocation amplitude between the two sub-series was tested. The minimum p-value among all sub-series comparisons within a series was taken as an indication of whether or not a switch occurred within the series.

For a single paired t-test, a p-value of 0.01 implies that there is a 1% chance of making a Type I error (where the t-test would indicate a significant difference in the mean, when in fact there is none). Since a rejection of the null hypothesis at *any* of the

breakpoints constitutes the rejection of the hypothesis that the same molecule is being interrogated during the series, a 1% probability of making a Type I error per t-test corresponds to a larger probability of making a Type I error on the null hypothesis that the same molecule is being interrogated over the whole series. Therefore, to determine the appropriate criterion for rejection of the null hypothesis based on the minimum p-value recorded during the (N-21) t-tests on a single series, the analysis procedure was reproduced on a set of data artificially generated from a single Gaussian distribution. The probability of making at least one Type I error (incorrect rejection of the null hypothesis that we are measuring the same molecule) out of the complete ensemble of 32 series was computed as a function of the minimum rejection p-value. The probability of making a single Type I error in any one of 32 series decreased significantly for p-values less than 10^{-5}, and was 0.1 at a p-value of 10^{-10}. Therefore, 10^{-10} was determined to be an indicator of the rejection p-value value for series homogeneity, with a p-value below 10^{-10} corresponding to the detection of a switching event within the series.

Statistical differences between sub-series were detected in 4 of the 32 series, suggesting that the molecule being measured was switched during the series; these series were excluded from further analysis. Comparisons between the remaining series are presented as a heat map (Figure 5a) revealing regions of statistical difference of the mean signal amplitudes between several series (blue color). The map is divided into blocks, with each block presumably corresponding to a different type of molecule. The color bar at the bottom (Figure 5a) indicates homogeneity within the corresponding series. To verify the ability of the analysis to detect switching of molecules within a series, we preserved the chronology of translocation events, but randomly chose the locations of the series terminations. As expected, the number of switching events detected within the series increased significantly with this treatment (Figure 5b, bar at bottom), while the heat map failed to resolve into distinct blocks as seen in Figure 5a. Finally, to model the effect of averaging over consecutive translocations in the absence of multiple measurements, the chronological order of translocation events was randomized while maintaining the series termination locations. In this case, none of the series exhibited significant differences in their mean translocation amplitudes, and all heterogeneity was lost (Figure 5c).



Supplementary Figure 1. Analysis of DNA translocation events. a) Heat map of p-values of t-tests comparing pairs of series arranged in order of current amplitude reveal significant differences between the

mean translocation amplitudes of different series. P-values of t-tests within each series are represented by a bar (bottom). b) Random assignment of series while maintaining the chronological order and number of series exhibits broadening of the heat map. T-tests within each series reveal a mixture of homogeneous and heterogeneous series (bottom bar). c) Heat map with chronological order of translocation events randomized fails to reveal significant differences between series. Similarly, no heterogeneity is detected within any series (bottom). All p-values are represented on a log scale (right).

The enhanced discrimination facilitated by averaging over multiple measurements, the ability to detect switching of molecules and significant differences in the mean translocation amplitudes between different series, and the complete loss of heterogeneity after randomization of the chronological order strongly suggest that a majority of the series represent multiple measurements on the same molecule. The data also indicate a very low probability of escape or switching per translocation event: only 7 DNA escape events and 4 switching events were observed in the 32 series comprising approximately 4200 translocation events, with an average of ~130 translocation events per series. Since switching of molecules of different types (λ-DNA to T7 DNA and T7 DNA to λ-DNA) represents half of the possible types of switching events that can occur, roughly equal number of switching events between molecules of the same kind is expected. This corresponds to approximately 8 switching events in 4200 translocations, or a switching probability of ~0.19%. The 7 escape events in approximately 4200 translocation events correspond to a recapture probability of 99.83% (see Equation (3) in main text).

Determination of Series Termination

The goal of this section is to describe the method by which termination of a series was determined. A series of measurements is strictly defined by the following criteria:

1. One and only one translocation occur between voltage reversals
2. No greater than 500 ms between translocation events

Consider a current trace in which a series of translocations, $\{\delta I_k\}$ is detected, where each translocation occurs at time t_k. The current trace also contains large discontinuities in the current corresponding to a change in the polarity of the voltage. These voltage reversals occur *only when* the real time feedback algorithm has triggered a voltage reversal. The last translocation in a series is denoted as I_N, occurring at time t_N. The closest voltage reversal after the last translocation is denoted as V^N, occurring at time T_N.

The series failures are:

1) Second Molecule Translocating before voltage reversal occurs
    a. Criterion: δI_{N+1} occurs before V_N, and $t_{N+1} - t_N < 50$ ms
    b. Explanation: This indicates a second molecule translocating through the nanochannel, causing a break in the series
2) Failure of Real Time algorithm to detect DNA molecule

a. Criterion: δI_{N+1} occurs before V_N, and $t_{N+1} - t_N > 50$ ms

b. Explanation: The mean pre-reversal time is 30 ms, with nearly no pre-reversal times greater than 50 ms. Had the program detected the molecule and reversed the voltage, then the second translocation would occur after V^N.

3) DNA molecule escape

   a. Criterion: δI_{N+1} occurs after V_N, and $t_{N+1} - t_N > 500$ ms

4) Real Time algorithm incorrectly detects a DNA molecule

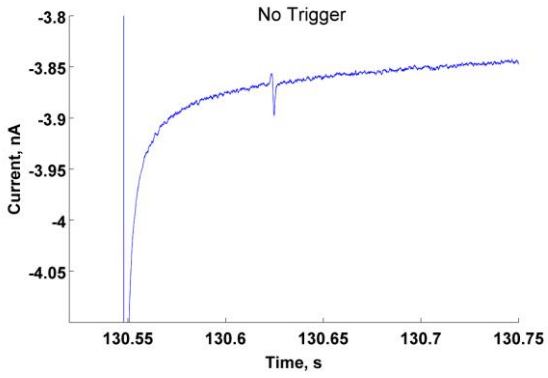
   a. Criterion: δI_{N+1} occurs on the voltage reversal after V^N (corresponding to no translocations between two voltage reversals)

The only criterion that seems to be missing is when: δI_{N+1} occurs after V^N, and $t_{N+1} - t_N <$ 500 ms. If this is the case, and (4) is not satisfied, then it is easy to see that the translocation δI_{N+1} satisfies the criterion for being part of the series, and would have been included as such. Thus, these four criteria map out all possible causes of series failure.

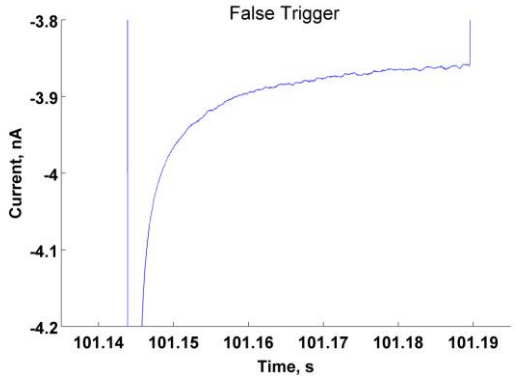
Supplementary Figure 2 provides examples for each of the four causes of failure. Preceding each current trace are translocations (not shown), each of which satisfy the criterion for belonging to a series with greater than 32 measurements in the series.

a)

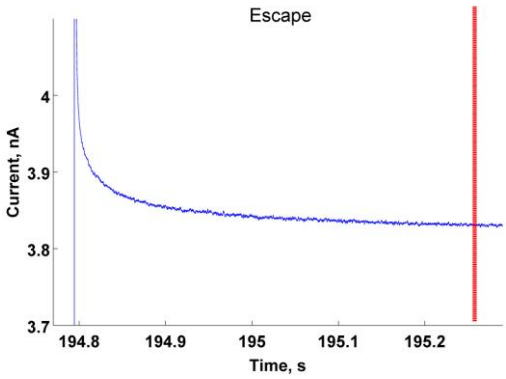
b)



c)



d)



Supplementary Figure 2. Causes of Series Failure. (a) Double translocation: a second molecule translocating within 50 ms of the first translocation. (b) No Trigger: the real-time algorithm fails to detect the translocation event. Consequently, there is no voltage reversal soon after the translocation. (c) False Trigger: LabVIEW falsely detects a translocation event, and initiates a voltage reversal prior to the translocation of the molecule. (d) Escape: no translocation is detected within 500 ms after the last translocation in the series. The red line indicates the 500 ms cutoff wait time.

1. R. M. M. Smeets, U. F. Keyser, D. Krapf, M. Y. Wu, N. H. Dekker and C. Dekker, *Nano Letters*, 2006, **6**, 89-95.
2. M. Gershow and J. A. Golovchenko, *Nature Nanotechnology*, 2007, **2**, 775-779.