Supporting Information for:

Nanopores Discriminate among Five C5-Cytosine Variants in DNA

Zachary L. Wescoe, Jacob Schreiber, Mark Akeson*

Department of Biomolecular Engineering, Baskin School of Engineering, MS SOE2
University of California, Santa Cruz, California 95064, United States

*Correspondence: makeson@soe.ucsc.edu

Figure S1. **Sequence of DNA template strands used in this study.** The strands are grouped by XnCGY 4mers (red letters). C5-cytosine variants within the XnCGY 4mers are as follows: C = cytosine, E = 5-hydroxymethylcytosine, F = 5-formylcytosine, J = 5-carboxylcytosine, Z = 5-methylcytosine. Blue letters represent trimer labels used to identify the C5-cytosine variant. Within these labels, ‘T’, ‘C’, ‘G’, and ‘A’ denote canonical bases and ‘a’ denotes an abasic residue. A is the first templating base during strand synthesis.

5′-CATCATCATCATCATCATGGTCATCATCACGATCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCATCATCATCATCAZGATCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCATCaTCATCATCAEGATCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCATTTTCATCATCAFGATCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCAaaaTCATCATCAJGATCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′

5′-CATCATCATCATCATCATGGTCATCATCACGCTCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCATCATCATCATCAZGCTCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCATCaTCATCATCAEGCTCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCATTTTCATCATCAFGCTCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCAaaaTCATCATCAJGCTCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′

5′-CATCATCATCATCATCATGGTCATCATCACGGTCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCATCATCATCATCAZGGTCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCATCaTCATCATCAEGGTCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCATTTTCATCATCAFGGTCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCAaaaTCATCATCAJGGTCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′

5′-CATCATCATCATCATCATGGTCATCATCACGTTCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCATCATCATCATCAZGTTCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCATCaTCATCATCAEGTTCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCATTTTCATCATCAFGTTCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCAaaaTCATCATCAJGTTCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′

5′-CATCATCATCATCATCATGGTCATCATCCCGATCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCATCATCATCATCCZGATCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCATCaTCATCATCCEGATCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCATTTTCATCATCCFGATCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCAaaaTCATCATCCJGATCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′

5′-CATCATCATCATCATCATGGTCATCATCCCGCTCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCATCATCATCATCCZGCTCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCATCaTCATCATCCEGCTCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCATTTTCATCATCCFGCTCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCAaaaTCATCATCCJGCTCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′

*
5′-CATCATCATCATCATCATGGTCATCATCCCGGTCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCATCATCATCATCCZGGTCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCATCaTCATCATCCEGGTCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCATTTTCATCATCCFGGTCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCAaaaTCATCATCCJGGTCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′

5′-CATCATCATCATCATCATGGTCATCATCCCGTTCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCATCATCATCATCCZGTTCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCATCaTCATCATCCEGTTCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCATTTTCATCATCCFGTTCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCAaaaTCATCATCCJGTTCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′

5′-CATCATCATCATCATCATGGTCATCATCGCGATCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCATCATCATCATCGZGATCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCATCaTCATCATCGEGATCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCATTTTCATCATCGFGATCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCAaaaTCATCATCGJGATCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′

5′-CATCATCATCATCATCATGGTCATCATCGCGCTCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCATCATCATCATCGZGCTCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCATCaTCATCATCGEGCTCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCATTTTCATCATCGFGCTCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCAaaaTCATCATCGJGCTCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′

5′-CATCATCATCATCATCATGGTCATCATCGCGGTCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCATCATCATCATCGZGGTCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCATCaTCATCATCGEGGTCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCATTTTCATCATCGFGGTCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCAaaaTCATCATCGJGGTCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′

5′-CATCATCATCATCATCATGGTCATCATCGCGTTCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCATCATCATCATCGZGTTCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCATCaTCATCATCGEGTTCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCATTTTCATCATCGFGTTCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCAaaaTCATCATCGJGTTCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′

5′-CATCATCATCATCATCATGGTCATCATCTCGATCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCATCATCATCATCTZGATCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCATCaTCATCATCTEGATCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCATTTTCATCATCTFGATCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCAaaaTCATCATCTJGATCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′

5′-CATCATCATCATCATCATGGTCATCATCTCGCTCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCATCATCATCATCTZGCTCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCATCaTCATCATCTEGCTCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCATTTTCATCATCTFGCTCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCAaaaTCATCATCTJGCTCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′

5′-CATCATCATCATCATCATGGTCATCATCTCGGTCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCATCATCATCATCTZGGTCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCATCaTCATCATCTEGGTCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCATTTTCATCATCTFGGTCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCAaaaTCATCATCTJGGTCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′

5′-CATCATCATCATCATCATGGTCATCATCTCGTTCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCATCATCATCATCTZGTTCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCATCaTCATCATCTEGTTCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCATTTTCATCATCTFGTTCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′
5′-CATCATCATCATCATCAaaaTCATCATCTJGTTCATCATCATCATCATCATTATTATTAGCATTCTCATGCAGGTCGTAGCC-3′

Figure S1 (continued)

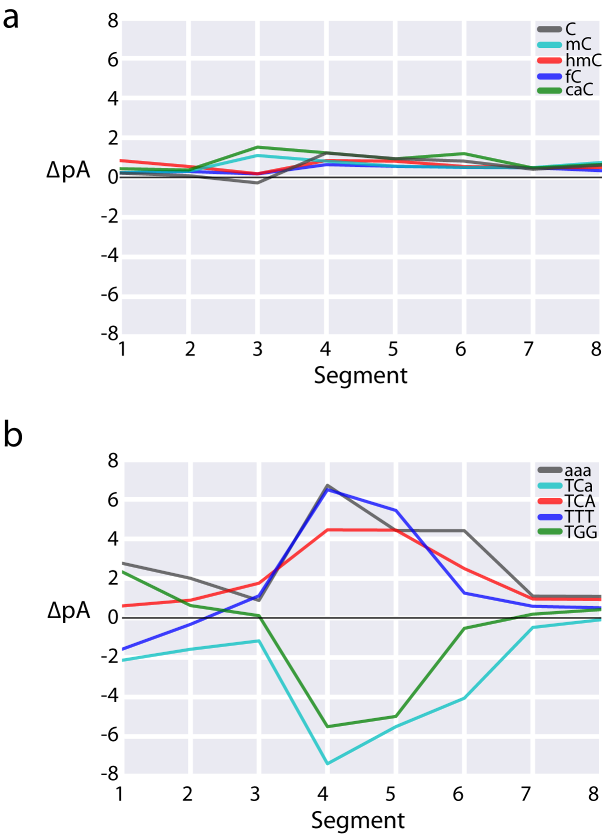

Figure S2. **Label swap experiment.** Our objective in this experiment was to reconfirm that the trimer labels built into the template strands did not affect the ionic current amplitude of segments used to identify the C5-cytosine variants. To do this we swapped labels among cytosine variants in the CnCGG context and compared the ionic current readouts. These label swapped templates were identical to those listed for CnCGG (see *) in Figure S1 except that the labels were changed as follows: C, TGG to aaa; mC, TCA to TTT; hmC, TCa to TCA; fC, TTT to TGG; caC, aaa to TCa, where 'T', 'C', 'G', and 'A' denote canonical bases and 'a' denotes an abasic residue. We collected 250 events for each of the two sets of template strands. Using the data from the eight key ionic current segments, we calculated ten consensus events, one for each cytosine variant/label pair, by averaging the ionic current amplitudes observed for each segment. To make comparisons we subtracted the consensus ionic current means of one set from the consensus ionic current means of the other set. (a) Comparison of strand pairs that contain the same cytosine variant, but different labels. For example, the blue line represents the difference in consensus events from the two cytosine containing strands where one contained the TGG label and the other the aaa label. The difference in the consensus events for all pairs with the same C5-cytosine variant is minimal (<2pA). (b) Comparison of strand pairs that contain the same label, but different C5-cytosine variants. For example, the blue line represents the difference in consensus events from the two strands bearing the TTT label where one contained mC and the other fC. The difference in the consensus events for all pairs with the same label but different C5-cytosine variants is large for segments 4, 5, and 6.

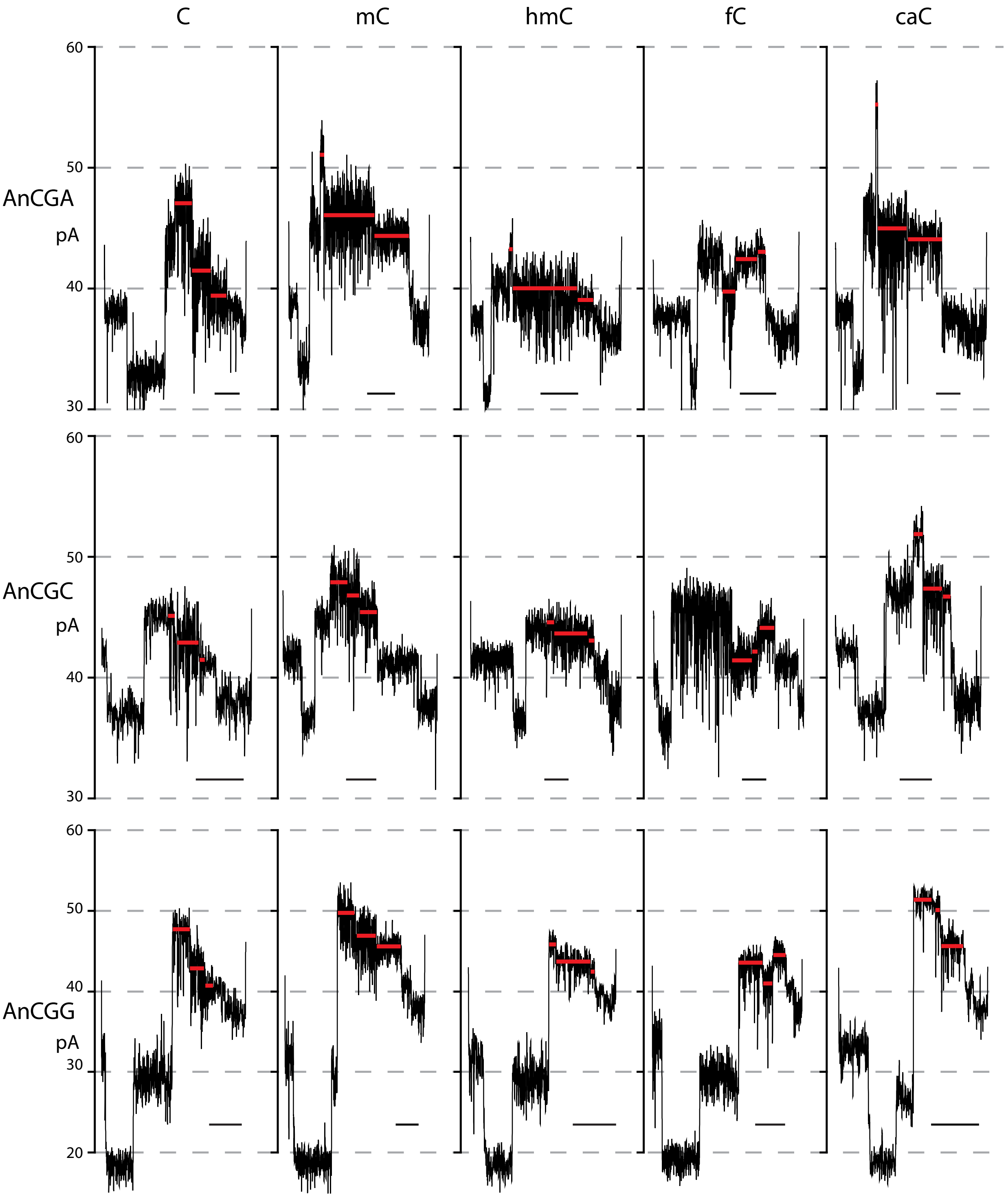

Figure S3. **Representative ionic current traces for C5-cytosine variants in fifteen XnCGY 4mer contexts.** Each panel shows eight consecutive ionic current segments that we quantified. The cytosine variant present on the strand is denoted above the trace. Segments 4, 5, and 6, which differed the most between C5-cytosine variants, are highlighted with horizontal red lines. Time scale bars are 30ms. Comparable traces for the GnCGC context are shown in Figure 3.

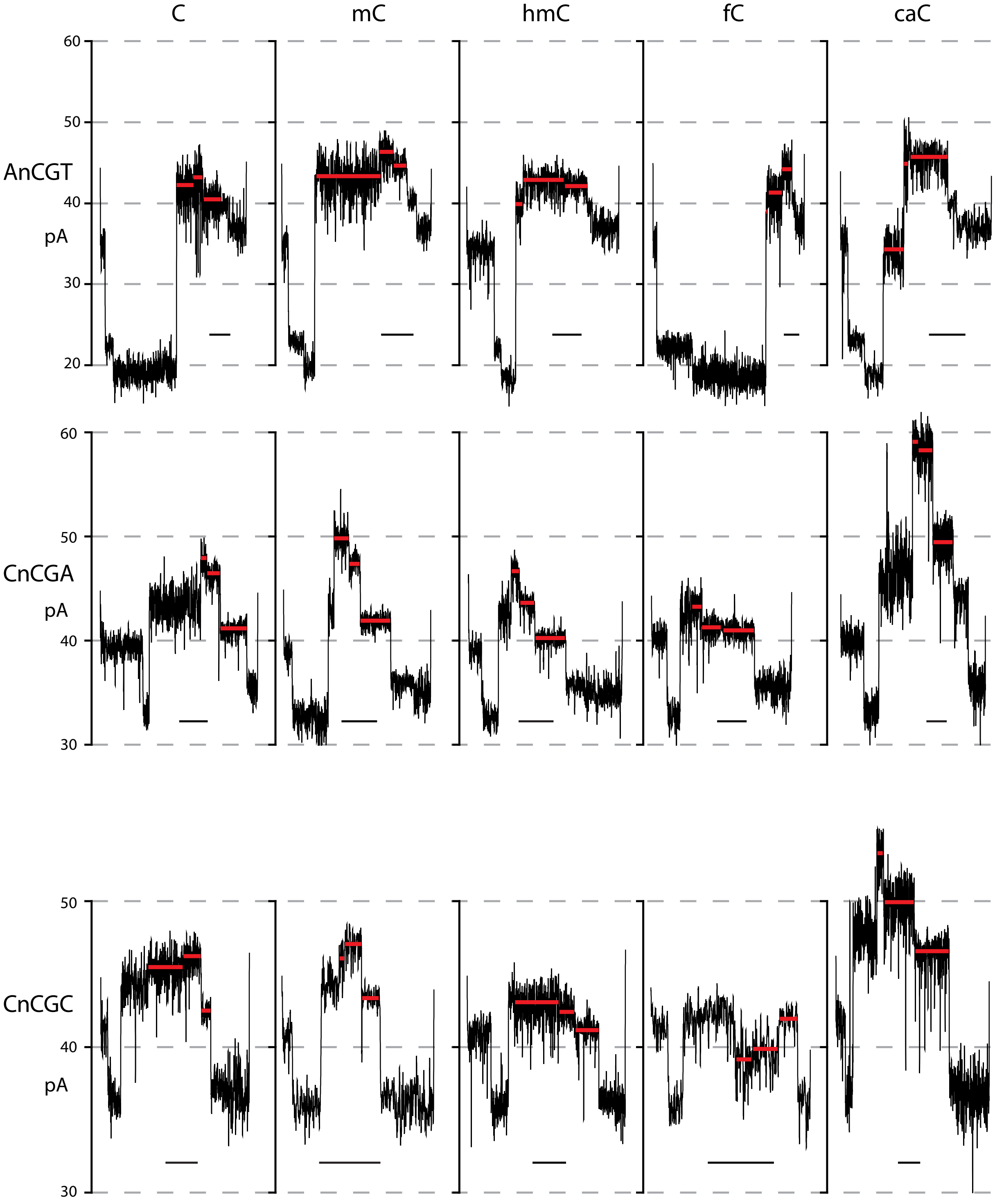

Figure S3 (continued)

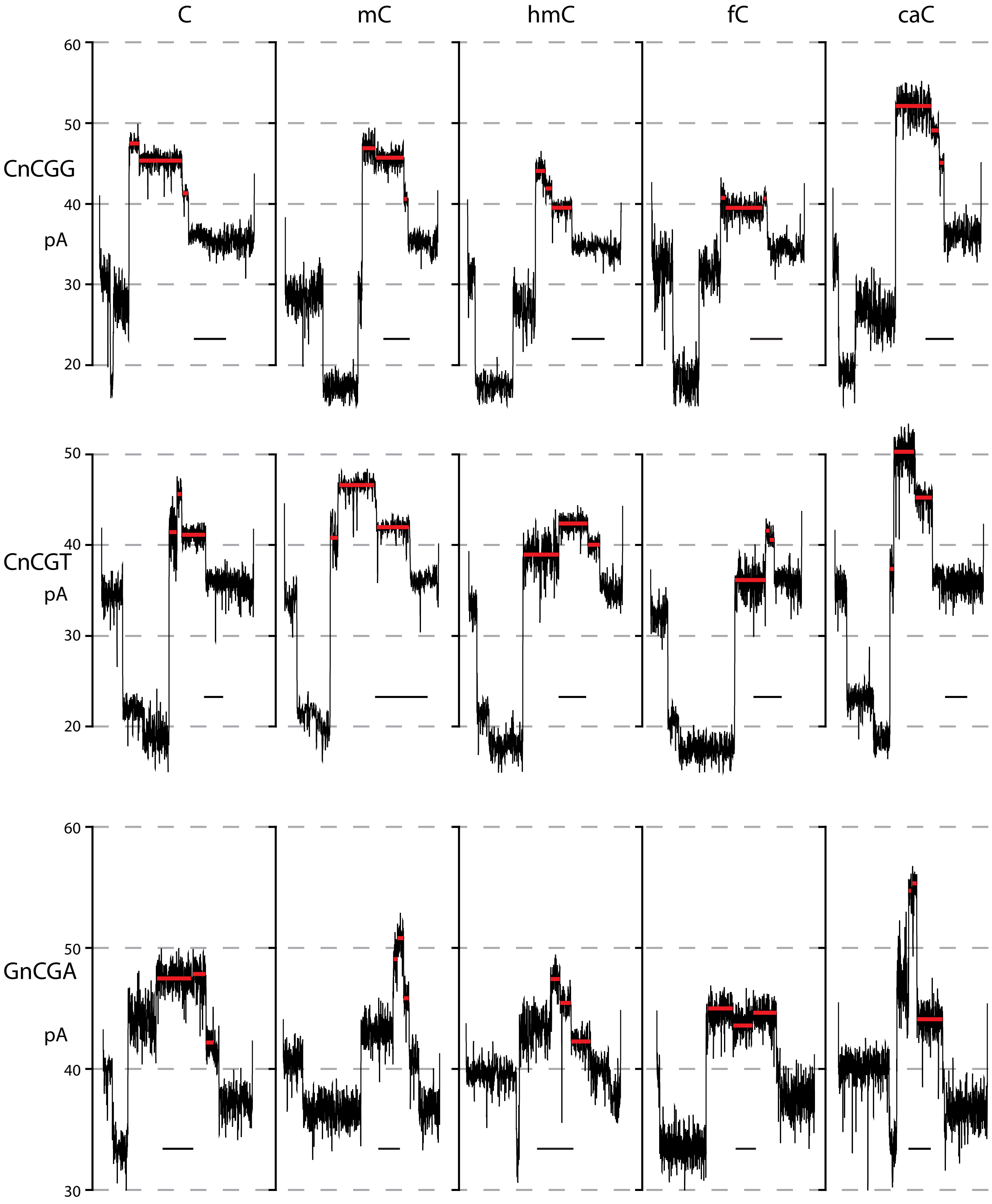

Figure S3 (continued)

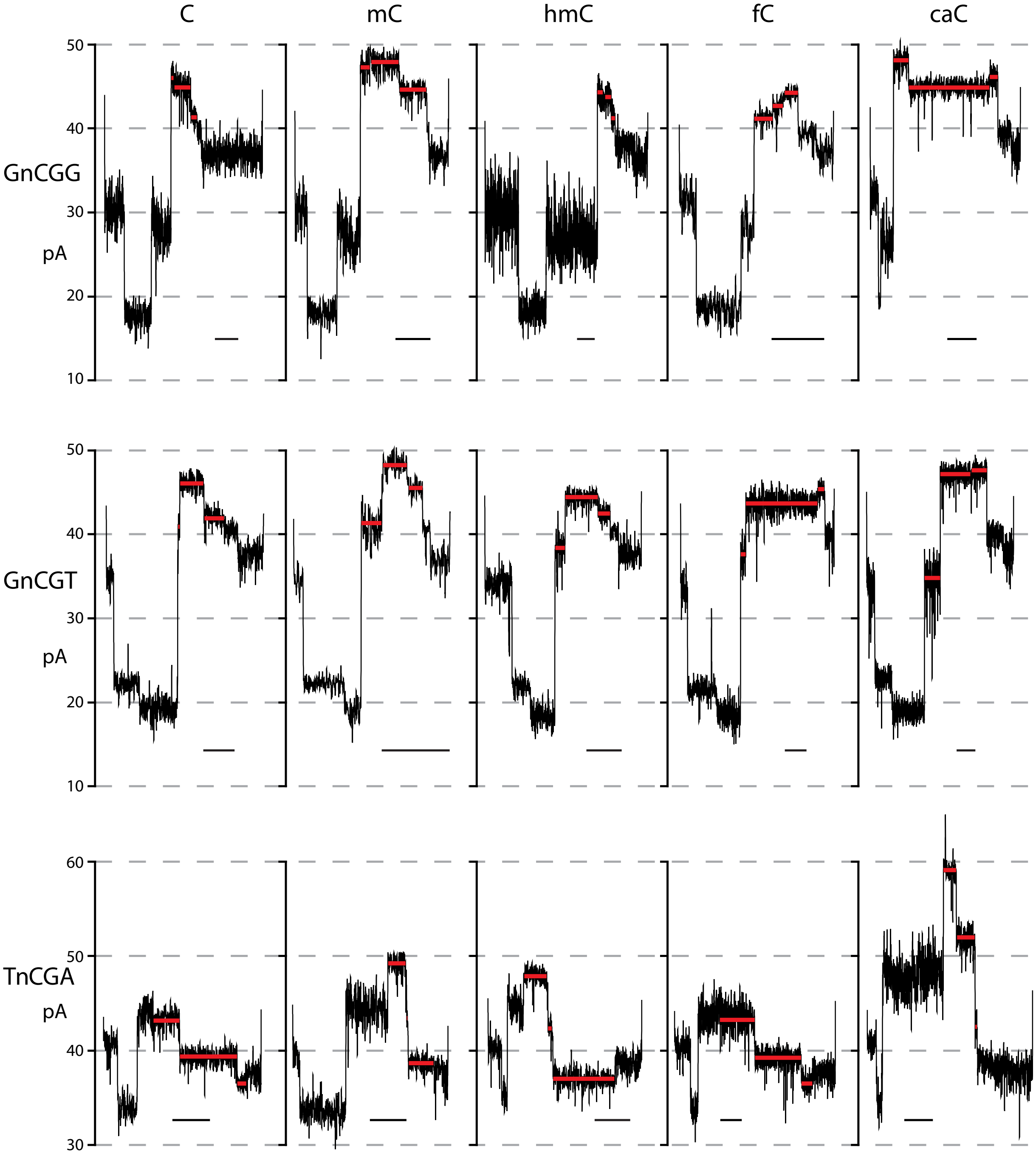

Figure S3 (continued)

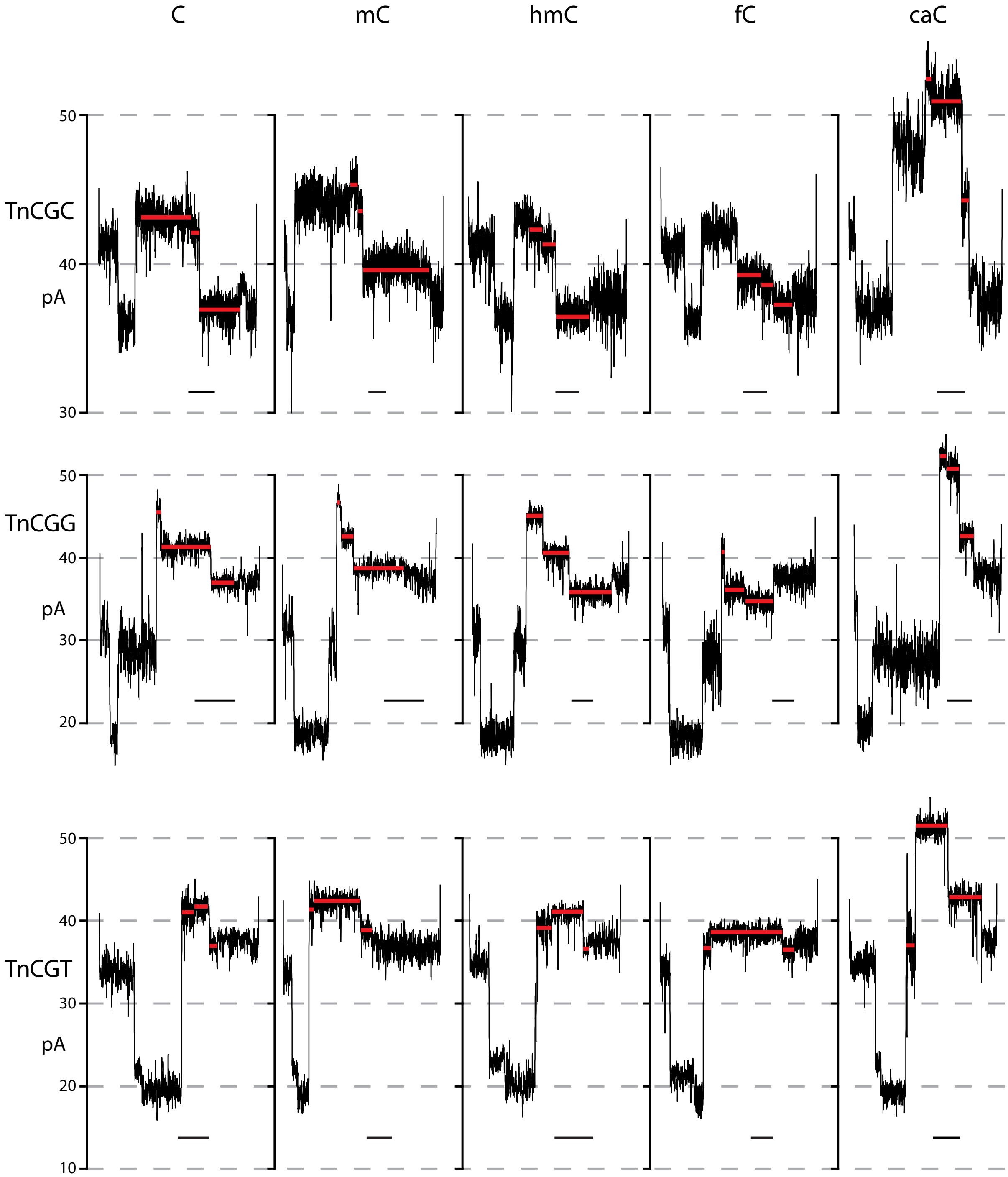

Figure S3 (continued)

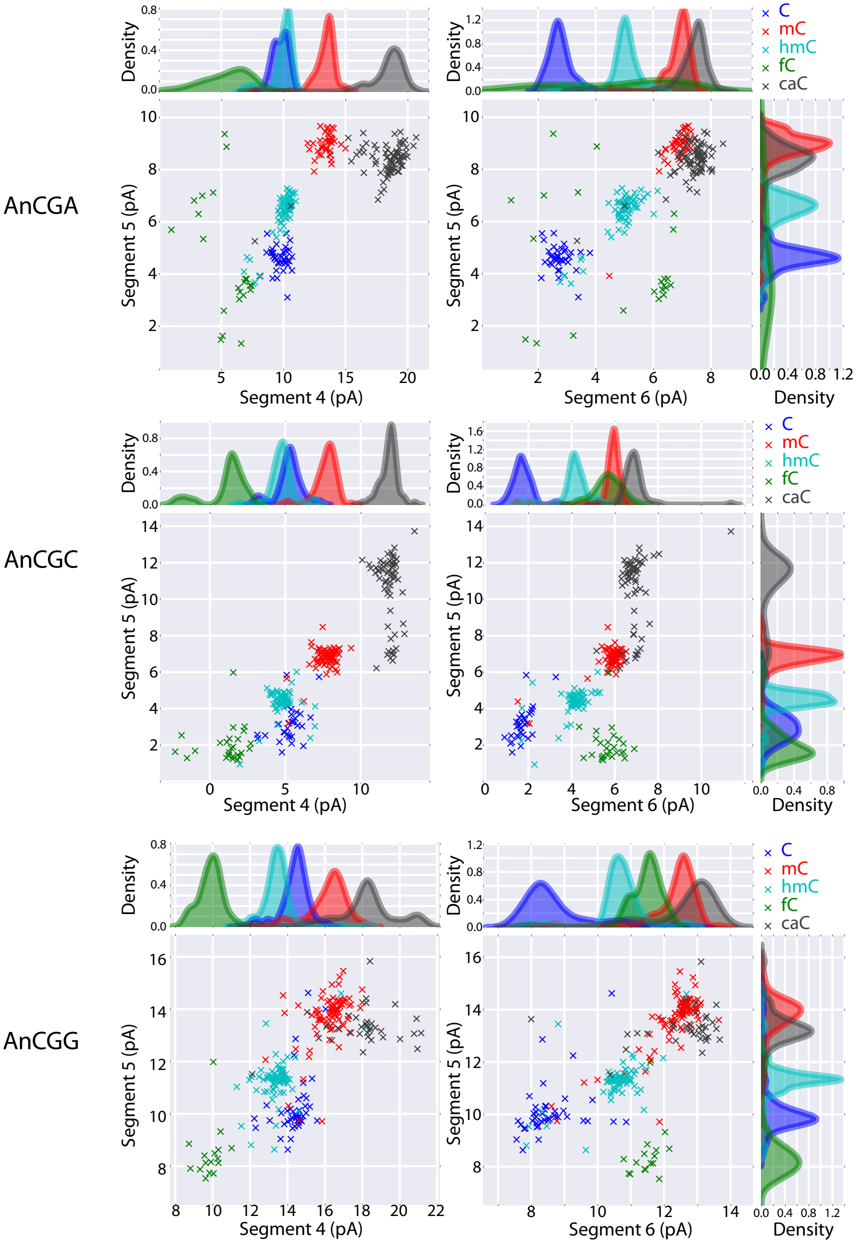

Figure S4. **Discrimination among C5-cytosine variants based on important ionic current segments.** The five letters at the left of each graph denote the XnCGY context for that graph. Because each experiment contained all five cytosine variants for a given context in the *cis* well bathing a single M2MspA nanopore for each experiment, the downstream label on each template was used to establish the C5-cytosine variant identity at nC. In all panels, Xs represent normalized ionic current values for an individual DNA strand read one time (continued next page).

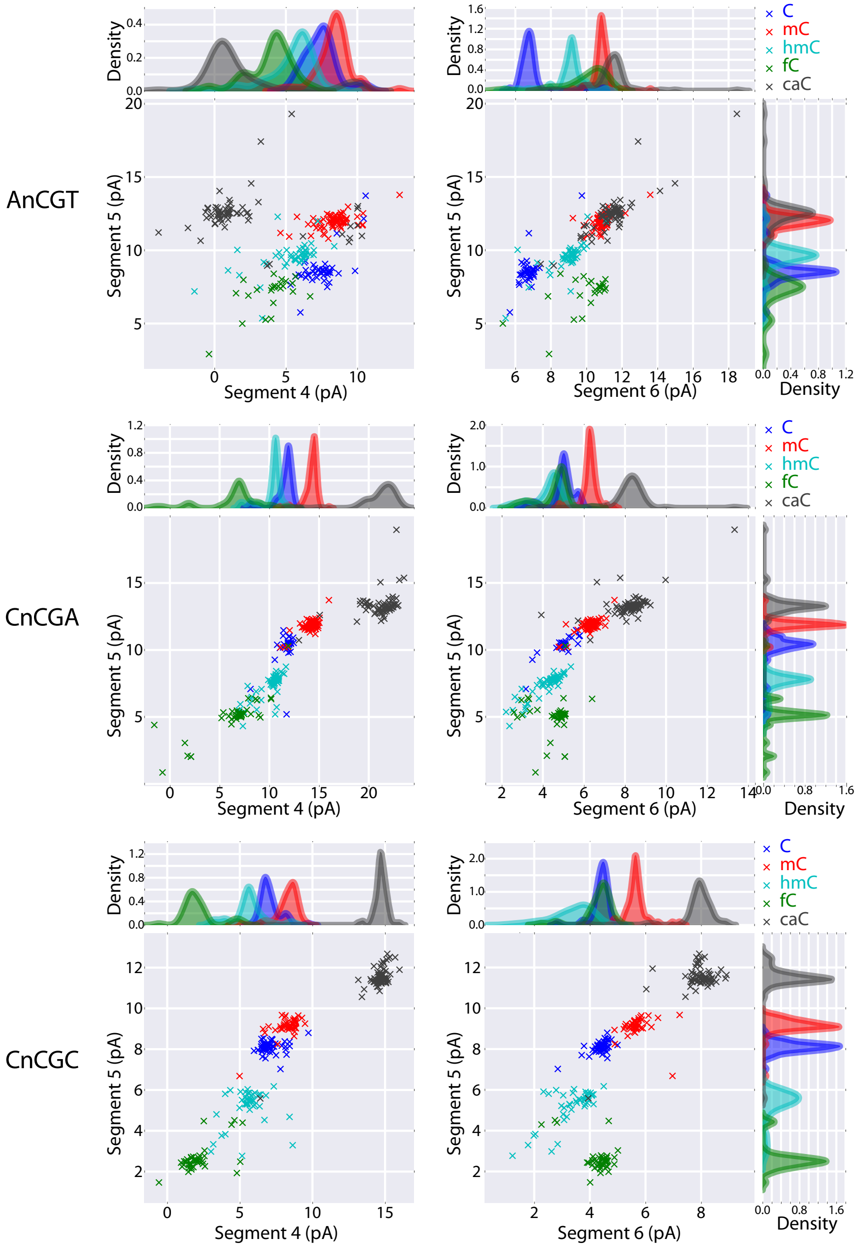

Figure S4 (continued). The color key for the C5-cytosine variant and its paired label is as follows: blue, cytosine/TGG; red, mC/TCA; cyan, hmC/TCa; green, fC/TTT; and grey, caC/aaa. For all panels, the Y-axis is normalized ionic current for segment 5. In the panels at left for each XnCGY context, the X-axis is normalized ionic current for segment 4; in the panels at right for each XnCGY context, the X-axis is normalized ionic current for segment 6. Marginal histograms using kernel densities are shown for each of the three segments used in classification. Histogram colors correspond to colors assigned to Xs.

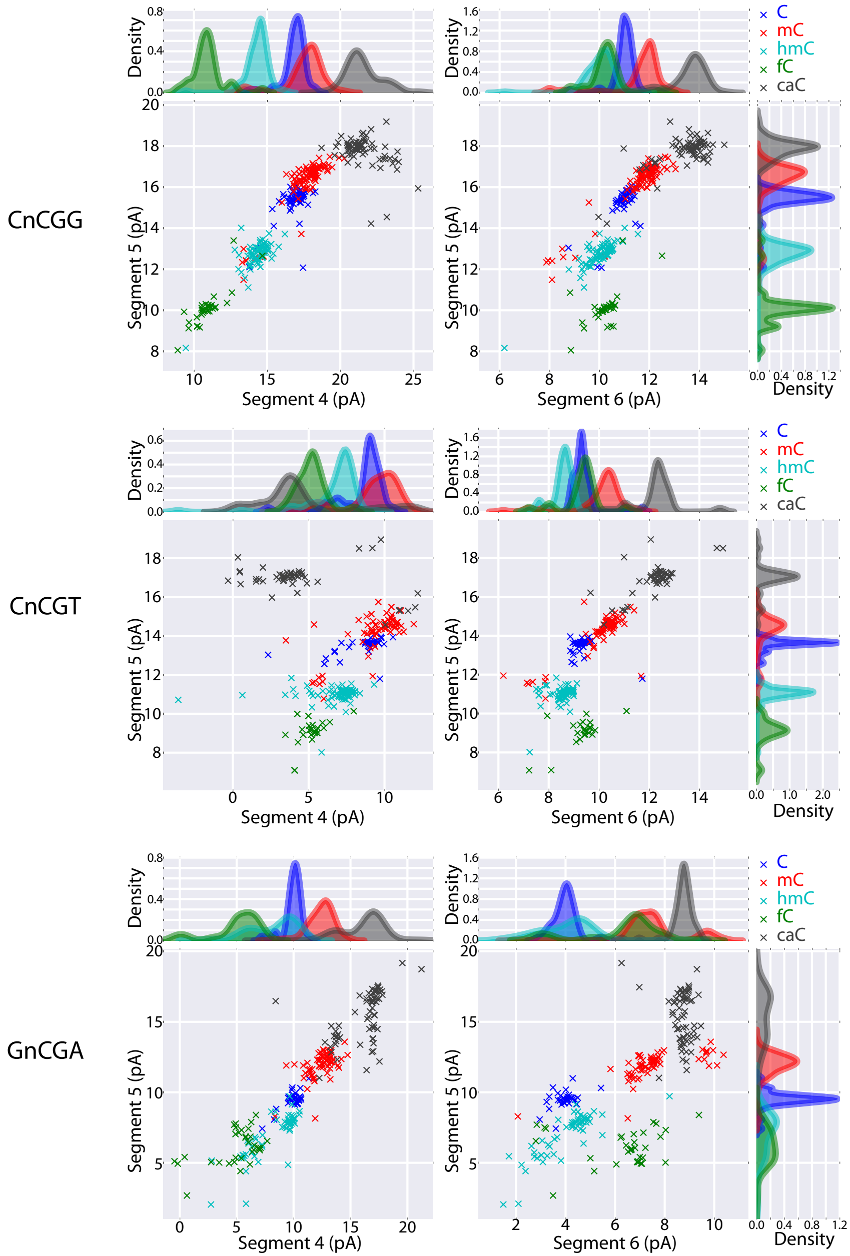

Figure S4 (continued)

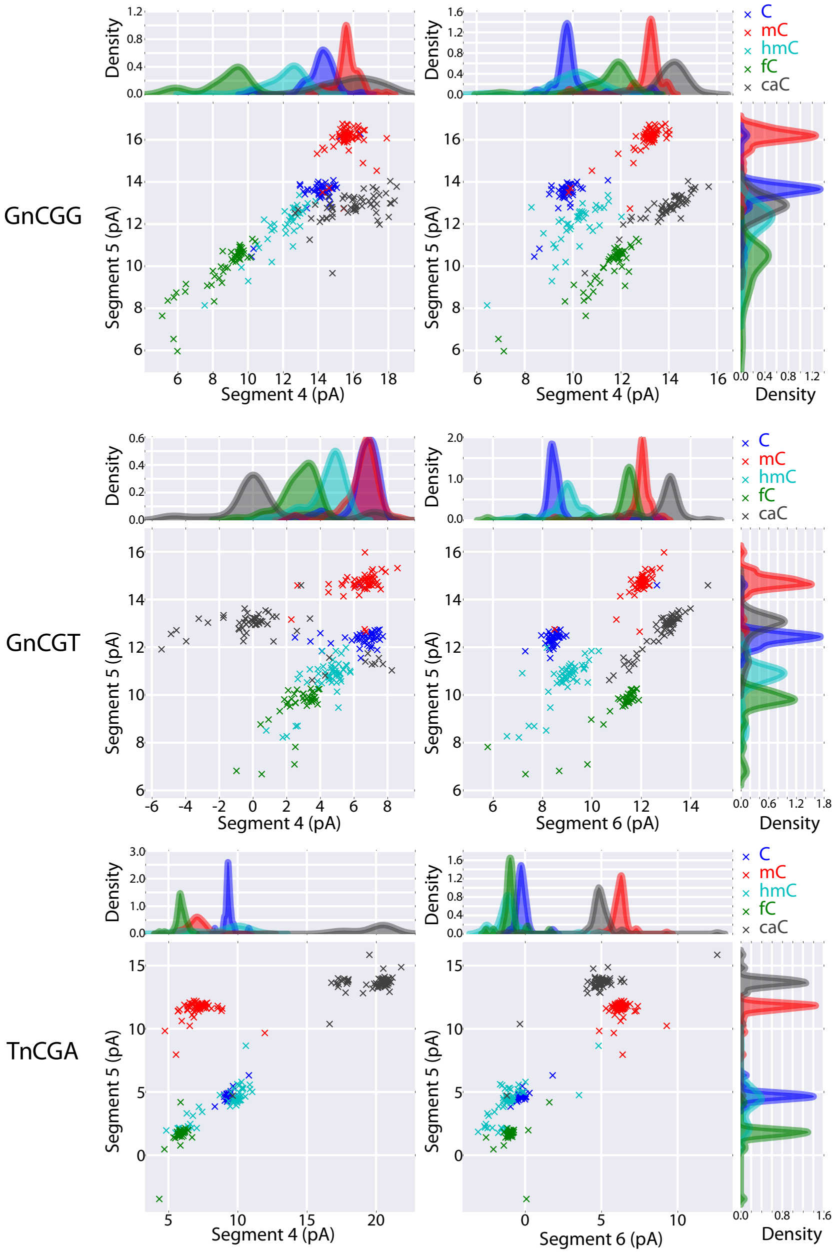

Figure S4 (Continued)

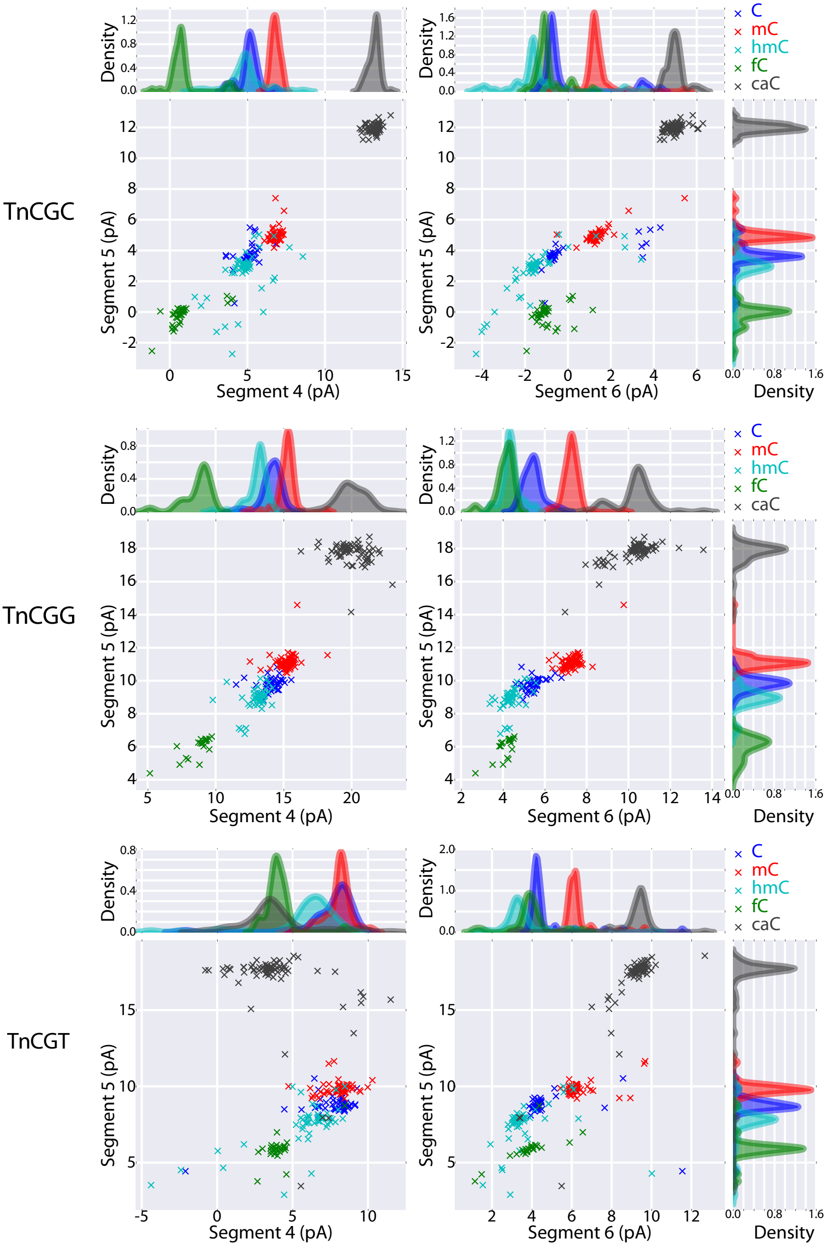

Figure S4 (continued)

```
AnCGA
----
       C     mC    hmC   fC    caC
TGG    0.95  0.00  0.04  0.00  0.00
TCA    0.01  0.97  0.02  0.00  0.01
TCa    0.01  0.00  0.99  0.01  0.00
TTT    0.03  0.02  0.02  0.93  0.00
aaa    0.01  0.01  0.01  0.00  0.96
```

```
AnCGC
----
       C     mC    hmC   fC    caC
TGG    0.94  0.03  0.02  0.00  0.00
TCA    0.03  0.95  0.01  0.00  0.00
TCa    0.01  0.01  0.96  0.02  0.00
TTT    0.04  0.01  0.00  0.94  0.00
aaa    0.00  0.00  0.00  0.00  1.00
```

```
AnCGG
----
       C     mC    hmC   fC    caC
TGG    0.92  0.05  0.02  0.00  0.00
TCA    0.03  0.92  0.04  0.00  0.01
TCa    0.02  0.03  0.94  0.00  0.01
TTT    0.00  0.00  0.06  0.93  0.00
aaa    0.02  0.09  0.03  0.00  0.86
```

```
AnCGT
----
       C     mC    hmC   fC    caC
TGG    0.92  0.04  0.00  0.01  0.02
TCA    0.00  0.97  0.01  0.00  0.02
TCa    0.02  0.05  0.87  0.04  0.02
TTT    0.06  0.00  0.08  0.86  0.00
aaa    0.01  0.03  0.01  0.00  0.95
```

Figure S5. **Confusion matrices for C5-cytosine variant calls.** The five letters above each matrix represent the XnCGY 4mer for that matrix. The rows are labels that identify the nC content of the template strand. For those labels, G, A, T, and C have their usual meanings and ‘a’ means an abasic residue. The columns are the C5-cytosine variant calls (See Figure 2 for details). The diagonal values (bold) represent the fraction of times the C5-cytosine variants calls agreed with the labels. Values in standard font are the fraction of times the predicted C5-cytosine variant was miscalled as an incorrect C5-cytosine variant. In some instances the rows do not sum to 1.00 because of rounding. The number of translocation events for each XnCGY context is in Table 1 of the main text.

```
CnCGA
----
      C     mC    hmC   fC    caC
TGG   0.81  0.08  0.10  0.00  0.00
TCA   0.04  0.95  0.00  0.00  0.02
TCa   0.00  0.00  0.98  0.02  0.00
TTT   0.03  0.00  0.04  0.93  0.00
aaa   0.01  0.03  0.00  0.00  0.96


CnCGC
----
      C     mC    hmC   fC    caC
TGG   0.97  0.01  0.02  0.00  0.00
TCA   0.02  0.97  0.00  0.00  0.00
TCa   0.02  0.00  0.95  0.03  0.00
TTT   0.00  0.00  0.08  0.91  0.00
aaa   0.00  0.00  0.01  0.00  0.98


CnCGG
----
      C     mC    hmC   fC    caC
TGG   0.87  0.08  0.06  0.00  0.00
TCA   0.06  0.91  0.03  0.00  0.01
TCa   0.01  0.01  0.95  0.03  0.00
TTT   0.00  0.00  0.09  0.90  0.00
aaa   0.00  0.00  0.00  0.00  1.00


CnCGT
----
      C     mC    hmC   fC    caC
TGG   0.91  0.07  0.01  0.00  0.00
TCA   0.09  0.90  0.00  0.00  0.01
TCa   0.00  0.00  0.99  0.00  0.00
TTT   0.00  0.02  0.11  0.86  0.00
aaa   0.00  0.06  0.00  0.00  0.94


GnCGA
----
      C     mC    hmC   fC    caC
TGG   0.87  0.03  0.09  0.00  0.00
TCA   0.00  0.96  0.02  0.00  0.02
TCa   0.01  0.02  0.93  0.04  0.00
TTT   0.00  0.00  0.12  0.88  0.00
aaa   0.00  0.03  0.00  0.00  0.97
```

Figure S5 (continued)

```
GnCGC
----
      C     mC    hmC   fC    caC
TGG   0.97  0.00  0.03  0.00  0.00
TCA   0.00  0.99  0.00  0.00  0.01
TCa   0.02  0.00  0.97  0.00  0.00
TTT   0.00  0.00  0.00  0.99  0.00
aaa   0.01  0.00  0.01  0.00  0.98


GnCGG
----
      C     mC    hmC   fC    caC
TGG   0.89  0.04  0.05  0.02  0.00
TCA   0.03  0.95  0.00  0.00  0.02
TCa   0.06  0.00  0.86  0.08  0.00
TTT   0.00  0.00  0.01  0.99  0.00
aaa   0.00  0.00  0.02  0.00  0.97


GnCGT
----
      C     mC    hmC   fC    caC
TGG   0.96  0.02  0.02  0.00  0.00
TCA   0.02  0.96  0.00  0.00  0.02
TCa   0.02  0.00  0.97  0.01  0.00
TTT   0.00  0.00  0.06  0.94  0.00
aaa   0.00  0.01  0.00  0.01  0.98


TnCGA
----
      C     mC    hmC   fC    caC
TGG   0.85  0.00  0.15  0.00  0.00
TCA   0.00  1.00  0.00  0.00  0.00
TCa   0.05  0.02  0.93  0.00  0.00
TTT   0.00  0.00  0.06  0.94  0.00
aaa   0.00  0.00  0.02  0.00  0.98


TnCGC
----
      C     mC    hmC   fC    caC
TGG   0.83  0.03  0.11  0.03  0.00
TCA   0.00  0.99  0.00  0.00  0.00
TCa   0.05  0.05  0.88  0.02  0.00
TTT   0.00  0.00  0.04  0.95  0.00
aaa   0.00  0.00  0.00  0.00  1.00
```

Figure S5 (continued)

```
TnCGG
----
      C     mC    hmC   fC    caC
TGG   0.86  0.06  0.08  0.00  0.00
TCA   0.00  0.98  0.00  0.00  0.01
TCa   0.09  0.00  0.91  0.00  0.00
TTT   0.00  0.00  0.05  0.94  0.00
aaa   0.00  0.00  0.00  0.00  0.99
```

```
TnCGT
----
      C     mC    hmC   fC    caC
TGG   0.86  0.08  0.04  0.02  0.01
TCA   0.01  0.99  0.00  0.00  0.00
TCa   0.04  0.04  0.84  0.06  0.02
TTT   0.04  0.01  0.07  0.89  0.00
aaa   0.00  0.00  0.04  0.00  0.96
```

Figure S5 (continued)