ChemComm

Detection of Alcohol-Associated Cancer Marker by Single-Molecule Quantum Sequencing

Journal:	*ChemComm*
Manuscript ID	CC-COM-09-2020-005914.R1
Article Type:	Communication

COMMUNICATION

Detection of Alcohol-Associated Cancer Marker by Single-Molecule Quantum Sequencing

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

Yuki Komoto,[a,b] Takahito Ohshiro [a] and Masateru Taniguchi [*a]

Alcoholic beverages is a well-known risk factor of cancers. N^2-ethyl-2'-deoxyguanosine (N^2-Et-dG) is a promising biomarker of alcohol-associated cancers. However, the lack of a convenient detection method for N2-Et-dG hinders the development of practical DNA damage markers. Herein, we develop a detection method for N^2-Et-dG using the single-molecule quantum sequencing (SMQS) method and machine learning analysis. Our method succeeded in discriminating between N^2-Et-dG and dG with an accuracy of 99 %, using 20 signals. Our developped method quantified the mixing rathio of N^2-Et-dG from a mixed solution of N^2-Et-dG and dG. It is shown that our method has the potential for facilitating the development of DNA damage markers and the early detection and prevention of cancers.

Ethanol, a component of widely consumed alcoholic beverages, is a well-known cause of cancers such as oral cavity, liver, and oesophageal cancer and is accountable for 5.3% of the deaths in 2016 in the world.[1–3] It is necessary to develop biomarkers for alcohol-associated cancers for the early detection and treatment of cancers. Acetaldehyde, an ethanol metabolite, damages the DNA and disrupts the DNA replication process by adding itself to the DNA.[4,5] Fig. 1a shows the scheme of formation of the most abundant acetaldehyde aduct.[4–9] Acetaldehyde adds to the deoxyguanosine (dG) and dehydrates to form N^2-ethylidine-dG. To detect unstable N^2-ethylidine-dG *in vitro*,[9] the modified nucleotide is reduced to stable N^2-ethyl-2'-deoxyguanosine (N^2-Et-dG) by the addition of a reductant. In conventional methods, N^2-Et-dG is detected using liquid chromatography tandem mass spectrometry after enzymatic degradation.[4–9] Previous studies showed that N^2-Et-dG has the potential of alcohol-associated cancer markers. [4,7] However, the lack of a convenient and simple detection method for N^2-Et-dG hinders the development of practical DNA damage markers. Herein, we focused on single-molecule quantum sequencing (SMQS), theoretically proposed by Di Ventra's group.[10,11] SMQS is based on the mechanically controllable break junction (MCBJ) method, which is one of the most commonly used methods for single-molecule measurements; a narrow metal wire drawn is broken by bending the substrate to form a nano gap as shown in Fig. 1b-d.[12–16] The conductance of a single molecule passing through the nanogap is measured using this method as shown in Fig. 1b. The DNA and RNA sequencing using the SMQS method has attracted significant attention.[17–21] In the SMQS method, nucleobases are detectable and countable without complicated pre-treatment in principle. The purpose of this study is to establish a convenient method to detect N^2-Et-dG and differentiate between N^2-Et-dG and dG using the SMQS method.

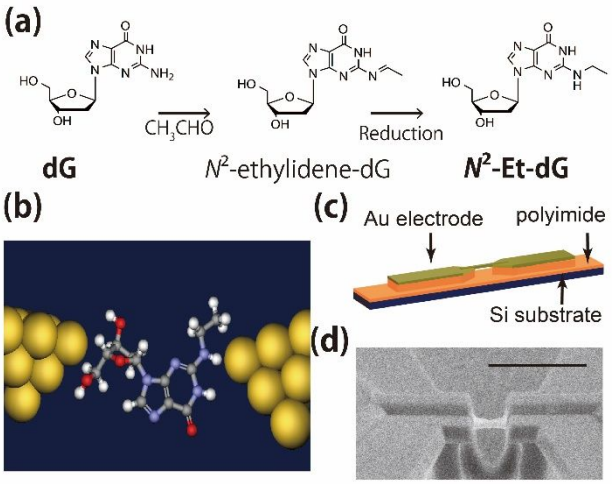


FIG.1 (a) Scheme of acetaldehyde addition to deoxyguanosine (dG). dG reacts with N^2-ethyl-2'-deoxyguanosine (N^2-Et-dG) via acetaldehyde addition, dehydration, and reduction. (b) Schematic of N^2-Et-dG single-molecule measurement. (c) Structures of SMQS substrate. Polyimide is coated onto a silicon substrate for electric insulation. A Au nanowire is fabricated with electron beam

[a] *The Institute of Scientific and Industrial Research, Osaka University, 8-1 Mihogaoka, Ibaraki, Osaka 567-0047, Japan E-mail: taniguti@sanken.osaka-u.ac.jp*
[b] *Artificial Intelligence Research Center, The Institute of Scientific and Industrial Research, Osaka University, 8-1 Mihogaoka, Ibaraki, Osaka 567-0047, Japan*
†Electronic Supplementary Information (ESI) available: Experimental and analysis details and estimation of classification accuracy. See DOI: 10.1039/x0xx00000x

lithography on the substrate. A Nanogap is formed by bending the substrate and breaking the Au nanowire. (d) SEM image of Au nanowire on the substrate. The scale bar is 1 μm.

Single-molecule measurements of N^2-Et-dG and dG solutions were carried out using the SMQS method.[18,20] The details of this method are described in the supplementary information (SI1). Figure 2a shows a part of the current measurement results of N^2-Et-dG. Single N^2-Et-dG signals were successfully detected. The histogram of the average current of the obtained signal is shown in Fig. 2b. A comparison between current histograms of N^2-Et-dG of dG shows the N^2-Et-dG structure to be in the higher current region compared to dG. The average current of the signals (indicated by arrows in Fig. 2(b)) at 0.1 V bias voltage is 47 pA (0.61 μG_0, G_0 is conductance quantum: $2e^2/h$, e,h denotes the elementary charge and Plank constant, respectively) for dG and 61 pA (0.79 μG_0) for N^2-Et-dG. N^2-Et-dG shows higher single-molecule conductance than dG. The histogram of N^2-Et-dG does not show well-defined conductance. In the conductance measurement of a single molecule, it has been reported that even a single molecule exhibits multiple conductance peaks or broad conductance peaks.[22–24] N^2-Et-dG was assumed to form multiple bridging structures with the electrodes. Both N^2-Et-dG and dG show similar shapes in the low-conductance region of the histograms. The molecular orbitals localized on the deoxyribose caused the peaks in the low conductance region.[25] The difference between the bases of N^2-Et-dG and dG appears in the high conductance region.[25] The difference in the conductance histograms of N^2-Et-dG and dG indicates the detection of differences between the bases.

DFT calculations using Gaussian 09[26] were performed to reveal the high conductance of N^2-Et-dG. The energy diagram based on molecular orbital calculations is shown in Fig. 2c. The HOMOs of N^2-Et-dG and dG are shown in Fig. 2d and (e). The HOMO levels are at 5.3 eV for N^2-Et-dG and at 5.4 eV for dG. HOMO of N^2-Et-dG has a smaller energy difference with respect to the Fermi level of gold at 5.1 eV.[27] In the Landauer picture,[28] the conductance of a single-molecule junction is proportional to the transmission. The transmission τ near the zero bias is expressed by Breit-Wigner equation (Eq. 1).[22,25,29,30]

$$\tau = \frac{4\Gamma_L\Gamma_R}{\varepsilon^2 + (\Gamma_L + \Gamma_R)^2} \qquad (1)$$

where $\Gamma_{L,R}$ and ε represent the coupling between the molecule-left and right electrodes and the energy difference between the Fermi energy of Au electrodes and the conduction orbital, respectively. The coupling indicates the delocalization between the molecular orbital and electrodes. From Figs. 2d and e, the orbital shapes of HOMO for both N^2-Et-dG and dG are similar. Therefore, the coupling of N^2-Et-dG and dG is assumed to be less significant than the effect of the energy difference between the Fermi level of the Au electrodes and the HOMO levels of the molecules. It can be interpreted that the smaller HOMO-Au Fermi energy difference of N^2-Et-dG causes higher conductance. The results of DFT calculations show the validity of single-molecule measurements.

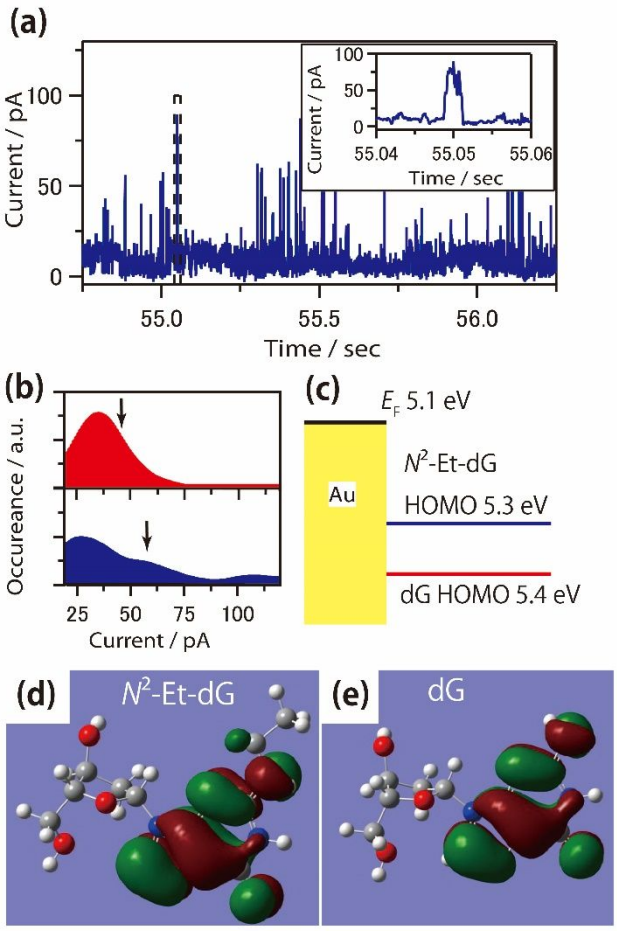


FIG.2 (a)Current profile of N^2-Et-dG single-molecule measurement using the SMQS method. Inset shows the expansion of a single pulse signal in the dotted line. (b) Current histogram of obtained pulse signal by single-molecule measurements of dG (red) and N^2-Et-dG (blue). The arrows indicate the average current. (c) Energy diagram of HOMO of the nucleobases and Au Fermi level. (d,e) HOMO of N^2-Et-dG (d) and dG (e).

The current histogram Figs. 2 confirms the statistical difference between N^2-Et-dG and dG. However, distinguishing the nucleobases from a single signal is difficult because of the large overlap between the histograms. Herein, the two nucleobases were distinguished using machine learning to quantify N^2-Et-dG (refer to SI.2).[31] First, we built a classifier between the dG and N^2-Et-dG signals. The classification scheme using machine learning is shown in Fig. 3a. The noise-removed signals by PUC[32] are classified by machine learning with the XGBoost classifier.[33] Figure 3b shows the classification result. It is observed that 81% of the N^2-Et-dG signals and 74% of the dG signals were correctly classified. *F*-measure, which is an index for measuring discrimination performance, was 0.77

(SI.2). Complete discrimination at a single-molecule level was not achieved. However, the two nucleobases are distinguishable using SMQS and ML-based analysis. Quantification among many molecules, instead of discriminating at a single-molecule level, is important for the application to detect cancers. The accumulated analysis improves the accuracy of statistical quantification. For example, when the accuracy for distinguishing only one signal is 0.77, the accuracy based on statistical analysis with 20 signals is over 99% (refer to SI3 in supplementary information). An *F*-measure of 0.77 is sufficiently high for a single signal. Machine learning can learn the average current as well as the current profile of statistical training data. High-precision discrimination of two nucleobases was achieved using a single signal. We succeeded in building a classifier that can distinguish between N^2-Et-dG and dG.

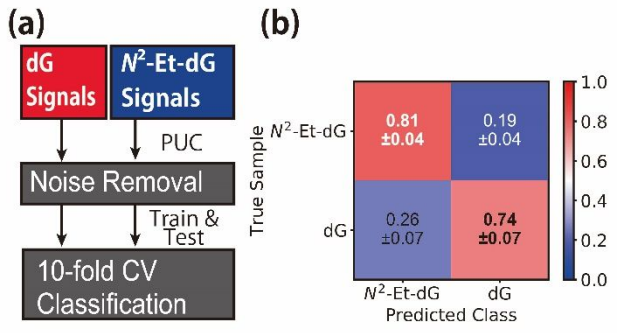


FIG.3 (a) Schematic flow of machine learning classification of N^2-Et-dG and dG single-pulse signals. (b) Prediction results between N^2-Et-dG and dG single-molecule signals. The matrix represents the average ratio and standard deviation of the 10-times classification.

To verify the robustness of the learned classifier and examine the applicability of the quantification of the mixture of N^2-Et-dG and dG, we carried out single-molecule measurements of the mixtures of N^2-Et-dG and dG. The concentration ratios of the mixed solutions were 1:3 and 3:1. As shown in Fig. 4a, the signals of the pure solutions were learned to build the classifier. Then, the signals of the mixture were classified as test data by the learned classifier. The concentration ratio of the mixed solution was determined from the ratio of the number of signals of the predicted class through machine learning. The obtained results are shown in Fig. 4b. Note that 70% of the signals from N^2-Et-dG:dG = 3:1 mixture were classified as N^2-Et-dG, and 80% of the signals from N^2-Et-dG:dG = 1:3 mixture were classified as dG. The concentration ratios predicted by machine learning were N^2-Et-dG:dG = 2.7:1 and 1:4.0 for N^2-Et-dG:dG = 3:1 and 1:3 solutions, respectively. The quantification of N^2-Et-dG in the presence of dG was confirmed by our SMQS method. After the DNA is extracted and monomerized, the ratio of N^2-Et-dG to dG in the DNA is examined by this method. This will allow the diagnosis of alcohol-associated cancers.

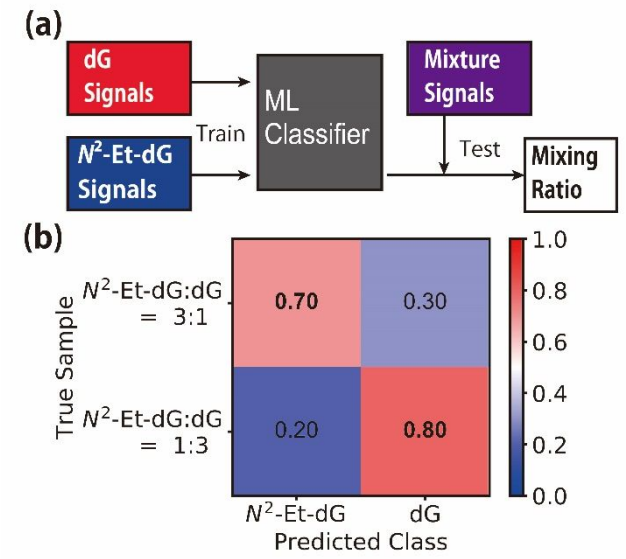


FIG.4 (a) Schematic flow of determination of mixing ratio using machine learning classification. The ML classifier was trained with pulse signals of N^2-Et-dG and dG. Signals obtained from mixture measurement were classified with the learned classifier. The concentration ratio of mixtures of N^2-Et-dG and dG were 1 μM:3 μM and 3 μM:1 μM, respectively. The signals denote the signals after noise removal with the PUC method. (b) The ML classification results of N^2-Et-dG and dG mixtures. The number of test signals for the signals obtained from the N^2-Et-dG :dG = 3:1 mixture and N^2-Et-dG: dG = 1:3 mixture were 123 and 178, respectively.

In conclusion, single-molecule measurements successfully detected the signals for dG and N^2-Et-dG. The current histogram of N^2-Et-dG showed a structure with higher conductance than dG. The DFT calculations revealed that the conductance of N^2-Et-dG increases as the gap between the HOMO level and Au Fermi level decreases. Machine learning classified the signals obtained from each solution with an F-measure of 0.77. The concentration ratios of dG to N^2-Et-dG were also determined using the machine learning-based method. The quantification of N^2-Et-dG under the coexistence of dG was confirmed using the SMQS and machine learning methods.

Our N^2-Et-dG detection method is promising for cancer diagnosis. Moreover, it can be used to investigate sequences damaged due to acetaldehyde. This study shows that N^2-Et-dG can be distinguished from dG by using SMQS. The development of direct analysis using machine learning for the DNA sequence provides sequence information of the DNA damage. This work will contribute to the development of practical DNA damage markers and the investigation of carcinogenic mechanisms affected by aldehyde-induced DNA damage.

This research was supported by KAKENHI Grant No. 19H00852, No.18K14091, and JST-CREST Grant Number JPMJCR1666. We would like to thank Editage (www.editage.com) for English language editing.

Conflicts of interest

There are no conflicts to declare.

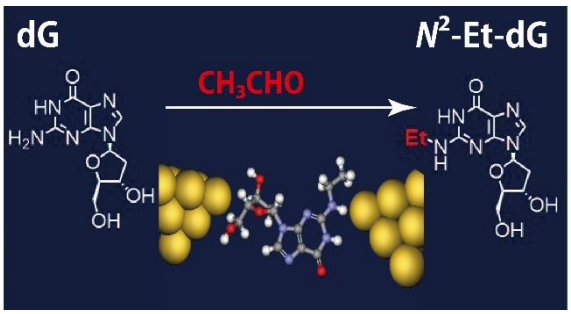
Notes and references

1 World. Health. Organization, *Global status report on alcohol and health 2018*, *World Health Organization*, 2019.
2 P. Boffetta and M. Hashibe, *Lancet Oncol.*, 2006, **7**, 149–156.
3 H. K. Seitz, C. Pelucchi, V. Bagnardi and C. La Vecchia, *Alcohol Alcohol.*, 2012, **47**, 204–212.
4 P. J. Brooks and J. A. Theruvathu, *Alcohol*, 2005, **35**, 187–193.
5 P. J. Brooks and S. Zakhari, *Environ. Mol. Mutagen.*, 2014, **55**, 77–91.
6 Y. Yukawa, S. Ohashi, Y. Amanuma, Y. Nakai, M. Tsurumaki, O. Kikuchi, S. Miyamoto, T. Oyama, T. Kawamoto and T. Chiba, *Am. J. Cancer Res.*, 2014, **4**, 279.
7 Y. Yukawa, M. Muto, K. Hori, H. Nagayoshi, A. Yokoyama, T. Chiba and T. Matsuda, *Cancer Sci.*, 2012, **103**, 1651–1655.
8 M. Wang, N. Yu, L. Chen, P. W. Villalta, J. B. Hochalter and S. S. Hecht, *Chem. Res. Toxicol.*, 2006, **19**, 319–324.
9 K. Hori, S. Miyamoto, Y. Yukawa, M. Muto, T. Chiba and T. Matsuda, *Biochem. Biophys. Res. Commun.*, 2012, **423**, 642–646.
10 J. Lagerqvist, M. Zwolak and M. Di Ventra, *Nano Lett.*, 2006, **6**, 779–782.
11 M. Zwolak and M. Di Ventra, *Nano Lett.*, 2005, **5**, 421–424.
12 C. A. Martin, D. Ding, H. S. J. Van Der Zant and J. M. Van Ruitenbeek, *New J. Phys.*, 2008, **10**, 65008.
13 Y. Komoto, S. Fujii and M. Kiguchi, *Mater. Chem. Front.*, 2018, **2**, 214–218.
14 Y. Komoto, S. Fujii, M. Iwane and M. Kiguchi, *J. Mater. Chem. C*, 2016, **4**, 8842–8858.
15 M. A. Reed, C. Zhou, C. J. Muller, T. P. Burgin and J. M. Tour, *Science (80-.).*, 1997, **278**, 252–254.
16 J. M. Van Ruitenbeek, A. Alvarez, I. Pineyro, C. Grahmann, P. Joyez, M. H. Devoret, D. Esteve and C. Urbina, *Rev. Sci. Instrum.*, 1996, **67**, 108–111.
17 M. Di Ventra and M. Taniguchi, *Nat. Nanotechnol.*, 2016, **11**, 117.
18 T. Ohshiro, Y. Komoto, M. Konno, J. Koseki, A. Asai, H. Ishii and M. Taniguchi, *Sci. Rep.*, 2019, **9**, 1–7.
19 T. Ohshiro, K. Matsubara, M. Tsutsui, M. Furuhashi, M. Taniguchi and T. Kawai, *Sci. Rep.*, 2012, **2**, 501.
20 T. Ohshiro, M. Tsutsui, K. Yokota and M. Taniguchi, *Sci. Rep.*, 2018, **8**, 1–8.
21 M. Tsutsui, M. Taniguchi, K. Yokota and T. Kawai, *Nat. Nanotechnol.*, 2010, **5**, 286–290.
22 Y. Komoto, S. Fujii, H. Nakamura, T. Tada, T. Nishino and M. Kiguchi, *Sci. Rep.*, 2016, **6**, 1–9.
23 S. Y. Quek, M. Kamenetska, M. L. Steigerwald, H. J. Choi, S. G. Louie, M. S. Hybertsen, J. B. Neaton and L. Venkataraman, *Nat. Nanotechnol.*, 2009, **4**, 230.
24 L. Venkataraman, J. E. Klare, I. W. Tam, C. Nuckolls, M. S. Hybertsen and M. L. Steigerwald, *Nano Lett.*, 2006, **6**, 458–462.
25 T. Furuhata, T. Ohshiro, G. Akimoto, R. Ueki, M. Taniguchi and S. Sando, *ACS Nano*, 2019, **13**, 5028–5035.
26 F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari and A. Rendell, *Inc. Wallingford, CT*.
27 H. B. Michaelson, *J. Appl. Phys.*, 1977, **48**, 4729–4733.
28 R. Landauer, *Philos. Mag.*, 1970, **21**, 863–867.
29 R. Frisenda, M. L. Perrin, H. Valkenier, J. C. Hummelen and H. S. J. van der Zant, *Phys. status solidi*, 2013, **250**, 2431–2436.
30 Y. Kim, T. Pietsch, A. Erbe, W. Belzig and E. Scheer, *Nano Lett.*, 2011, **11**, 3734–3738.
31 M. Taniguchi, T. Ohshiro, Y. Komoto, T. Takaai, T. Yoshida and T. Washio, *J. Phys. Chem. C*, 2019, **123**, 15867–15873.
32 C. Elkan and K. Noto, in *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2008, pp. 213–220.
33 T. Chen and C. Guestrin, in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.

A table of contents entry

Detection of Alcohol-Associated Cancer Marker by Single-Molecule Quantum Sequencing

Yuki Komoto, Takahito Ohshiro and Masateru Taniguchi *



N^2-ethyl-2'-deoxyguanosine, a promising biomarker of ethanol-derived cancers was detected with single-molecule quantum sequencer and discriminated from deoxyguanosine.