

1 分子 DNA シークエンサーが生み出すビッグデータ

Big Data Produced by Single-Molecule Electrical Sequencing of DNA

谷口 正輝 川合 知二 筒井 真楠 横田 一道 鷺尾 隆
Masateru Taniguchi Tomoji Kawai Makusu Tsutsui Kazumichi Yokota Takashi Washio

大阪大学 産業科学研究所

The Institute of Scientific and Industrial Research, Osaka University

Single-molecule electrical sequencing is a potential innovative technology to realize drug discovery, diagnostics, and therapeutics on the basis of genomic information and a potential destructive technology to realize the rapid growth of the big DNA sequencing market. By analyzing big data obtained from the time traces of single-molecule conductance, this single-molecule technology allows us to read genomic information that existing technologies cannot read. Here, we introduce the current status of R&D on single-molecule electrical sequencing technology and the issues that are expected to be overcome using machine learning.

1. はじめに

2003 年, ヒトのゲノムを全て解読したヒトゲノム計画の終了は, 遺伝子に基づく個別化医療の幕開けであり, 画期的な診断法, 治療法, 薬が次々に生み出されると期待されていた. ところが, ヒトの遺伝子を解読する費用と時間が個別化医療の大きな障壁となり, ヒトゲノム計画を主導した米国では, 低コスト・ハイスループットな DNA シークエンサーの開発が, 全米体制で 2004 年から開始された. その究極のターゲットが, 1 分子 DNA シークエンサーである[Di Ventra 16].

1 分子 DNA シークエンサーは, DNA を構成する 4 つの塩基分子を, 1 塩基分子を流れる極微小電流により識別する原理を持つ(図 1). この電流は, 1 分子の電子状態を反映したトンネル電流であり, 4 つの塩基分子の電子状態が異なるため, 電流による分子種の判別が可能となり, DNA の塩基配列が決定される. 1 塩基分子を流れるトンネル電流を計測するためには, 塩基分子の大きさと同程度となる 1nm の電極間距離を持つナノギャップ電極を作製することが必須となる. ところが, 先端の半導体技術を持ってしても, 1nm のナノギャップ電極を作製することは困難であり, 夢の技術と考えられていた.

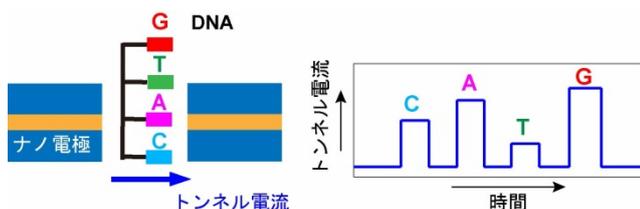


図 1: 1 分子 DNA シークエンシングの原理図

我々は, 3 点曲げの要領で機械的に金属細線を破断させ, 1nm 以下のナノギャップ電極を作製する方法(機械的破断接合)を開発した[Tsutsui 10]. この方法を用いて, DNA と RNA を構成するそれぞれ 4 つの塩基分子をトンネル電流により識別できることを示し, 短い DNA と RNA の塩基配列決定を実証してきた[Ohshiro 12]. さらに, ガンのバイオマーカーになることが知られているものの, 既存技術では直接識別できない化学修飾された塩基分子も, トンネル電流により 1 分子で識別できることを示

してきた[Tsutsui 11]. 最近, 既存技術では直接解析不可能なペプチドのアミノ酸配列や, 疾病マーカーとなる修飾アミノ酸分子の識別に成功した[Ohshiro 14]. このように, 1 分子 DNA シークエンサーは, DNA の塩基配列のみならず, RNA の塩基配列やペプチドのアミノ酸配列の決定にまで応用可能と期待されるが, その塩基配列決定法には 1 分子計測に特有な課題が見えてきた. 本講演では, 1 分子 DNA シークエンサーの開発状況と機械学習に期待する課題を紹介する.

2. トンネル電流による 1 分子識別

1 分子を流れるトンネル電流の計測は, 機械的破断接合を用いて作製した 1nm 程度のナノギャップ電極を用いて行われた. 1 塩基分子を水に溶解し, その水溶液の電流を計測すると, 図 2a のような最大電流値(I_p)と電流持続時間(t_d)の 2 つのパラメータで特徴付けられるシグナルが得られた. DNA の 4 つの塩基分子に関して電流-時間プロファイルを計測し, I_p/V のヒストグラムを作成すると, 図 2b のように, 全ての塩基分子のヒストグラムが 1 つのピーク電流値を持つことが明らかとなった. 塩基分子の電流値の大きさは, グアニン(G) > アデニン(A) > シトシン(C) > チミン(T)となり, この順列は塩基分子の電状態を反映した酸化電位の順列と同じであることが分かった[Ohshiro 2012]. RNA の 4 つの塩基分子, 化学修飾塩基分子, アミノ酸分子, および化学修飾アミノ酸分子のトンネル電流計測でも, DNA の塩基分子と同様なシグナルが得られた.

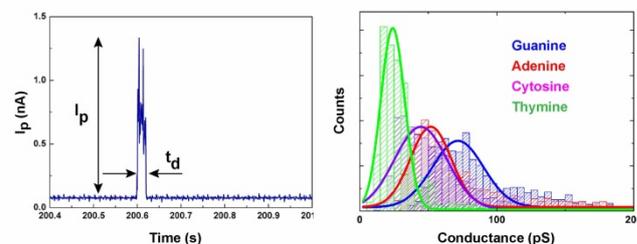


図 2: (a) グアニンの電流-時間プロファイルと, (b) 4 塩基分子の電流伝導度(I_p/V)のヒストグラム

トンネル電流は, 電極に対する分子の距離と配向に大きく依存する性質を持つ. 1 塩基分子のトンネル電流の計測では, 塩

基分子は、ブラウン運動とナノ電極間の電場の2つの力で、電極に対する距離と配向が決まると理論的に説明されている [Lagerqvist 2007]。1塩基分子は、1 価の負電荷と電気双極子を持つため、電界が存在するナノギャップ電極では、1 つの安定な距離と配向を持つと期待される。この結果、各塩基分子のトンネル電流のヒストグラムに 1 つのピーク電流が観察されたと考えられる。一方、1分子にかかるブラウン運動の影響が大きいと、観察されるトンネル電流に分散が観察されたと考えられる。現在の電流-時間プロファイルの解析法は、全ての電流シグナルを塩基分子由来とし、得られる電流ヒストグラムはガウス関数に従うと仮定している。この解析法は、1分子計測のルールであり、明確な根拠無しで電流シグナルの取舍選択を行わないことになっている。しかし、1nm のナノギャップ電極を用いる電流計測では、電極先端の原子の動きや不純物の通過も電流変化として観察されるため、解析に用いるシグナルには塩基分子由来ではないものが含まれている可能性がある。もしも、機械学習で 1 つ 1 つのシグナルを高い精度で判別でき、かつ判別に用いた特徴量を物理・化学の現象として説明することができれば、ヒストグラムは小さな分散を持つと予測される。小さい分散は、1分子の識別精度を向上させる大きなメリットを持つ。

3. トンネル電流による塩基配列決定

DNA, RNA, およびペプチドの水溶液の電流-時間プロファイルを計測すると、同様なシグナルが観察された。ここでは、グアニン(G)とチミン(T)の配列を持つ GTG の電流-時間プロファイルを示す(図 3)。ある時間領域を切り出して電流ヒストグラムを作成すると、2 つの電流ピークが観察されるため、2 つの塩基分子が溶液中に含まれていることが分かる。

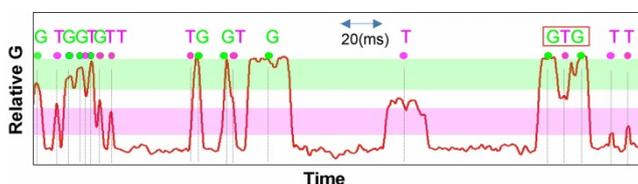


図 3: GTG の電流-時間プロファイル

次に、1 塩基分子の計測で得られた電流ヒストグラムを参照すると、ピーク電流値の比から、ヒストグラムで得られたピークがグアニンとチミンであると推定できる。実際の解析では、切り出したシグナルデータから電流ヒストグラムを作成し、各塩基分子のヒストグラムに分解する。各塩基分子のヒストグラムを用いて、計測で得られた電流-時間プロファイルの 1 つ 1 つのシグナルに含まれる G と T の比率を求め、高い比率の塩基分子をシグナルに帰属している。

現在の解析法には用いていないが、多くの DNA シークエンサーでは、1 つのシグナルにおける塩基配列決定に隠れマルコフモデルが使われている。既存の DNA シークエンサーでは、塩基分子そのものを識別している訳では無く、1 つの DNA の末端に化学的に結合した 1 つの色素分子の発光を用いて塩基分子を識別している。この原理では、4 つの塩基分子に 4 つの色(色素分子)を対応させており、色素分子は DNA の末端に結合しているので、観察される 4 色の光は独立事象である。ところが、1分子 DNA シークエンサーでは、1分子 DNA 上で隣接塩基分子と化学結合でつながった 1 塩基分子が、トンネル電流で読み出される。ある 1 塩基分子は、隣接塩基分子の電子状態や運動状態に影響を受けるため、1 塩基分子から得られるトンネル電流は隣接塩基分子の影響を受けることが予測され、独立

事象とは言い難い。つまり、GTG と ATA の T から得られるトンネル電流は異なると予測される。また、1 分子 DNA シークエンサーは、重要なガンマーカマーであり、既存技術では直接検出不可な修飾塩基分子もトンネル電流により識別する。この結果、2 種類の化学修飾塩基分子がトンネル電流で識別される場合、これらは第 5, 第 6 の塩基分子として取り扱われるため、既存の DNA シークエンサーに使われている隠れマルコフモデルをそのまま適用することが出来ない。このように、1 分子計測が 1 分子の受ける摂動を電流変化として観測するため、得られる 1 分子のシグナルと隣接塩基分子のシグナルが互いに干渉する特有の課題があり、この課題を解くことは、高い塩基配列決定精度に結び付くと期待される。

4. まとめ

1分子 DNA シークエンサーが、遺伝子情報に基づく個別化医療を実現するために必須の技術であることは広く認識されている。一方、RNA の塩基配列、ペプチドのアミノ酸配列、さらに化学修飾塩基分子と化学修飾アミノ酸分子は、1分子シークエンサーのみが直接読取可能である。特に、化学修飾塩基分子はガンマーカマーになることが知られているが、既存技術ではマーカー探索が困難であるため、1 分子 DNA シークエンサーの早期開発が強く求められている。1 分子 DNA シークエンサーは、1 分子計測に基づく大量のデータを用いる解析を基本としているため、その解析には新たな機械学習のモデル開発が必要と考えられる。また、1分子計測で得られるトンネル電流は、正確な制御、あるいは理解しきれない物理・化学現象を必然的に含むため、機械学習で得られる特徴量が、理解を深めるヒントを与えてくれると期待している。

参考文献

- [Di Ventra 16] Di Ventra, M., Taniguchi, M.: Decoding DNA, RNA and Peptides with Quantum Tunnelling, *Nature Nanotech.*, Vol. 11, pp. 117-126 (2016).
- [Lagerqvist 07] Lagerqvist, J., Zwolak, M., Di Ventra, M.: Influence of the Environmental and Probes on Rapid DNA Sequencing via Transverse Electronic Transport, *Biophys. J.*, Vol. 93, pp. 2384-2390 (2007).
- [Ohshiro 12] Ohshiro T., Matsubara K., Tsutsui M., Furuhashi M., Taniguchi M., Kawai T.: Single-Molecule Electrical Random Resequencing of DNA and RNA, *Sci. Rep.*, Vol. 2, No. 501 (2012).
- [Ohshiro 14] Ohshiro T., Tsutsui M., Yokota K., Furuhashi M., Taniguchi M., Kawai T.: Detection of Post-translational Modifications in Single Peptides Using Electron Tunnelling Currents, *Nature Nanotech.*, Vol. 9, pp. 835-840 (2014).
- [Tsutsui 10] Tsutsui M., Taniguchi M., Yokota K., Kawai T.: Identifying Single Nucleotides by Tunneling Current, *Nature Nanotech.*, Vol. 5, pp. 286-290 (2010).
- [Tsutsui 11] Tsutsui M., Matsubara K., Ohshiro T., Furuhashi M., Taniguchi M., Kawai T.: Electrical Detection of Single-Methylcytosines in a DNA Oligomer, *J. Am. Chem. Soc.*, Vol. 133, pp. 9124-9128 (2011).