DETECTION OF NUCLEIC ACID DISEASE BIOMARKERS USING SOLID-STATE NANOPORES

BY

OSAMA KAZI ZAHID

A Dissertation Submitted to the Graduate Faculty of

WAKE FOREST UNIVERSITY GRADUATE SCHOOL OF ARTS AND SCIENCES

in Partial Fulfillment of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

Molecular Genetics and Genomics

December, 2017

Winston-Salem, North Carolina

Approved By:

Adam R. Hall, Ph.D., Advisor

Timothy D. Howard, Ph.D., Chair

Dorothy A. Erie, Ph.D.

Douglas S. Lyles, Ph.D.

Ethan W. Taylor, Ph.D.

Martin Guthold, Ph.D.

DEDICATION AND ACKNOWLEDGEMENT

All Praise belongs to Allah (term in arabic for God or Creator), the Lord of the Universe.

'Recite in the name of your Lord who created. Created man from a clinging substance. Recite, and your Lord is the most Generous. Who taught by the pen. Taught man that which he knew not.' Chapter 96, Verses 1-5, The Holy Quran. These five verses of the Holy Quran was the first revelation sent to the Prophet Muhammad (peace be upon him). A reminder, we know so little and should continue throughout our lives to read and learn.

'We will show them Our signs in the horizons and within themselves until it becomes clear to them that it is the truth. But is it not sufficient concerning your Lord that He is, over all things, a Witness?' Chapter 41, Verse 53, The Holy Quran.

First and foremost I would like to thank the Almighty God for blessing me with a wonderful life. Blessing me with family, health and friends and giving me the opportunities that so many people all over the world unfortunately lack. I am truly and sincerely humbled when I reflect upon the things (which happens rarely) I've been lucky enough to be blessed with. First off I have been blessed with a wonderful, supportive and patient family. I would like to dedicate this work to my lovely parents, my father, Kazi Zahid Ali and my mother, Rana Kirmani. No words can express the gratitude, honor and respect they deserve. They sacrificed their entire lives for their children and continue to do so to this day just so we can have a comfortable and successful life. I hope I can follow in their footsteps and be as honest, kind, and generous as they are. I want to thank my two extremely supportive and loving brothers, the twins, Hamza and Omair Kazi Zahid. Throughout the past 6 years they have gone above and beyond to help me in any

way possible, it saddens me to think about the time apart we've spent from each other. I love my entire family; they have all been so patient and continue to pray for my success. I know I can never make it up to them but I want them to know I am extremely thankful to them and will always remember them in my prayers.

Next I would like to thank my advisors and mentors who have influenced my life, I would not be here today if it weren't for their kind words and support. I would like to acknowledge all of the professors in the Chemistry Department of UAE University, if not for them I would not have thought it'd be possible for me to pursue a doctorate degree in the US. Just being around them made me believe I could achieve that goal. I would especially like to thank Dr. Mohammed Meetani for giving me the opportunity to work as a research assistant in his lab, that was my first gig that kick-started my scientific career. Another person who deserves credit for this opportunity is Dr. Salman Ashraf, if it weren't for him I wouldn't have been introduced to Dr. Meetani and both of these wonderful human beings are always available whenever I need help or advice. Once I moved to NC, another person who was extremely helpful and influential was Shanaz Qadri and his wife, he was pursuing his doctorate degree as well and he helped me out a lot to settle down in Greensboro. Honestly, it was his advice that helped me make probably the most important decision in graduate school, choosing the right advisor. Which led me join Adam's lab and I can honestly say that Shahnaz was right it was the correct choice. I sincerely would like to thank Adam for everything that he has done for me throughout graduate school (and outside of it). I sincerely hope I can be half as dedicated, driven, ambitious and creative as he is. I will always be in awe of his belief and sheer will to succeed. He has taught me so much, to be sincere and conduct quality,

reproducible research, to ask the right questions and continue to be curious. He has been extremely patient with me and I am truly grateful to him for giving me the opportunity to join him when he moved his lab to Wake Forest University. For some reason, I don't know why I was the only graduate student that was lucky enough to get that chance and move to another great research institution but most importantly this allowed me to continue working on a research project that I was really interested in. Furthermore I would like to thank him for his generosity and patience during this past year. Its been a rough road, first I tore my ACL and had to get surgery to recover from that injury, and Adam was very supportive and patient until I completely recovered from that. He was extremely patient once again when I was stuck outside the country for two months trying to get my visa renewed, it was a tough and depressing time for me but Adam continued to be supportive. He is an extremely dedicated and creative scientist but most of all he is a kind person and for that I am grateful. Two other mentors deserve a shout out, Jan Ruzicka and Autumn Carlsen, these two are wonderful human beings and extremely hard working scientists, both helped me during initial graduate school years. Finally, I am grateful to all members of my thesis committee, their insight and suggestions helped navigate me in the right direction, without their guidance, this work wouldn't have been possible. I am sincerely and truly thankful to all of these people and will always remember them and their families in my prayers.

This has been a long and arduous road and I could not have gone through these ups and downs without a supportive group of friends, other family members and colleagues. It would take too long to go into precise details of why I am thankful to each and every individual on this list, but suffice is to say, they directly or indirectly helped me

achieve this goal. My first roommate Omar Lowe and his wonderful sister Awamary and her family, they really treated me like family and that helped me settle in North Carolina. My second roommate Michael Blankenship, he's just a lovely human being with a big heart, without him I wouldn't have survived the transition to Wake Forest. My North Carolina friends, Diana Garcia, Lakmini Premadasa, Max Burnett, Daniel Partridge, and Dave Herman we have shared a lot of good laughs together. Some friends are always there for you through a lot of tough times and they become more than just friends, they are family, Robby Harris is my American brother. My dear friends Naiara Sanchez and Caroline Xavier, my beautiful sisters, without their help and support I would not have survived my time here. They literally helped me get back on my two feet after my ACL reconstruction surgery. While we have fallen out of touch, I am grateful to Pablo Diaz, it was through him I met these wonderful people. I have to thank my supportive Muslim brothers and sisters at the Winston-Salem Community Mosque, especially Sajid and Shahid and their families. It's a tough time trying to be a Muslim in the West and I am grateful for having a strong supportive group to always keep me grounded in my faith. My lovely and beautiful extended family in India and across the globe, especially Nida, Farzana, Abbas, and Uncle Zia. I would like to dedicate this work to Sabiha Kirmani as well, my aunt who suffered from cancer and was taken from us too early, I and everyone else miss her dearly.

Finally, I would like to thank the rest of my colleagues Fanny, Felipe, and Shiny for maintaining a fun, collegiate and supportive lab culture. It's always nice to have to work with people who we get along with. I would especially like to thank Felipe for his

comments and suggestions, without his help I wouldn’t have been able to finish this document.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

Main text: List of figures and tables.

APPENDIX: LIST OF FIGURES AND TABLES

LIST OF ABBREVIATIONS

SS - Solid-state

MS – Monovalent Streptavidin

Bio – Biotinylated

ssBioDNA – single-stranded biotinylated

dsBioDNA – double-stranded biotinnylated

cf – cell free

OxBS-Seq – Oxidative Bisulfite Sequencing

TAB-Seq – Tet-Assisted Bisulfite Sequencin

5mC – 5-methylcytosine

5hmC – 5-hydroxymethylcytosine

5fC – 5-formylcytosine

5caC – 5carboxylcytosine

BS – Bisulfite treatment

βGT – β-glycosyltransferase

FOV – Field of view

HIM – Helium Ion Microscope

ABSTRACT

Nanopore sensing technology was initiated with the ultimate goal of developing a rapid, single-molecule sequencer. With the release of Oxford Nanopore Technology's Minion sensor in 2014, that goal has been achieved. While this device utilizes biological nanopores, Solid-State Nanopore sensors were also initially investigated to develop sequencing devices. Significant efforts are being made to this day towards this end, with advances and improvements being made to amplifiers and using various 2-D materials for devices such Graphene, and MoS_2. Other embodiments include, plasmonic nanopore devices, or integration of tunneling electrodes with SS-Nanopores. However, sequencing is not the only application that can be addressed by SS-nanopores, these devices have tremendous potential to detect various types of disease biomarkers such as nucleic acids, epigenetic modifications and proteins. Several studies have made significant efforts to develop novel assays to detect biological markers of disease such as cancer, and infectious disease agents. However, most of these involve significant interpretation of the electronic current signature, which can be difficult due to minor structural differences between analytes.

This work specifically focuses on development of a novel binary nanopore detection assay for detection of cancer-specific nucleic acid biomarkers. First, we report on the assay and its binary nature of exclusively detecting dsDNA molecules with biotin modifications due to selective binding of a protein ligand. We demonstrate the event rate exclusively correlates to the concentration of the protein-bound dsDNA complex, thereby allowing quantification of biotinylated dsDNA in a target sample.

We adapt the assay to detect specific single-stranded nucleic acid sequences by designing biotinylated ss-DNA binding probes. We demonstrate the assay generates events when protein is bound to biotinylated ds- as opposed to ssDNA. These detection differences arise due to the overall structure of the two nucleic acid strands at the investigated length scales. We demonstrate detection of a particular sequence from a background of non-specific ss-DNA sequences. Finally we detect a microRNA sequence, miR155, an important biomarker of lung-cancer.

We employ the assay to detect another important cancer biomarker, the epigenetic modification 5-hydroxymethylcytosine (5hmC). This modification is significantly depleted in all forms of cancer and the ability to quantify using a highly sensitive but affordable technique would help in early cancer screening. Since our nanopore assay works on biotinyalted dsDNA, we perform an enzymatic labeling technique to attach a biotin linker to 5hmC. We demonstrate selective detection of target 5hmC containing DNA by the nanopore assay once it is labeled with biotin. As a proof-of-concept we quantify 5hmC content in murine genomic DNA, and find our assays results to be within error of mass spectrometry.

Finally, we adapt a novel labeling assay developed in our lab that replaces damaged DNA bases with biotinylated nucleotides to address cytosine DNA epigenetic modifications. Other than 5hmC, there are three other cytosine epigenetic modifications, 5mC, 5fC and 5caC. We demonstrate successful labeling of the demethylation pathway modifications with biotin.

1. Chapter 1 - Introduction

All prokaryotes and eukaryotes are cellular organisms, while prokaryotes are single-cell entities eukaryotes are multi-cellular beings, with different types of cells and several smaller molecular machines and compartments called organelles. Common aspect among these different types of organisms is that they are cellular organisms, i.e. they have a boundary referred to as the membrane that encloses organelles and the code of life, the genome or genetic material. All information that leads to the formation of cells and cellular components, to tissues etc., is stored in the genome, and is passed on to the offspring, each individual unit is referred to as a gene. While in the case of eukaryotes the genetic material is enclosed within another protective structure called the nuclear membrane and the structure itself is called the nucleus, most prokaryotic organism's genome is not. Within the nucleus there are compact structures called chromosomes, observable under a microscope when stained with dyes, they are made up of DNA and proteins referred to as histones (Figure 1-4). With the advent of molecular biology, several experiments established that DNA is the carrier of genetic information in cells. In order to pass on the genetic information the DNA should be accurately copied and passed on or transferred.

1.1. Central Dogma of Life

Shortly after the discovery of the DNA double helix, a clearer picture of the flow of information was formulated. DNA is first *transcribed* or copied into another form of nucleic acid molecule, the ribonucleic acid (RNA) where the sugar contains a hydroxyl group at the 2' C as opposed to DNA. RNA is generally present in a single-stranded

form, it acts as short-term transfer of information and due to its flexibility can fold upon itself to form complex structures, some with catalytic activity such as ribosomes. The RNA transcript called the messenger RNA (mRNA), is then *translated* into a polymer of amino acids called proteins. The whole process of DNA transcription to RNA, and subsequent translation of RNA to protein is referred to as the Central Dogma of life (Figure 1-1). Proteins form specific well-defined structures depending on their amino acid sequence, they form complex structural components in animals such as muscles and are responsible for most of the enzymatic activity.

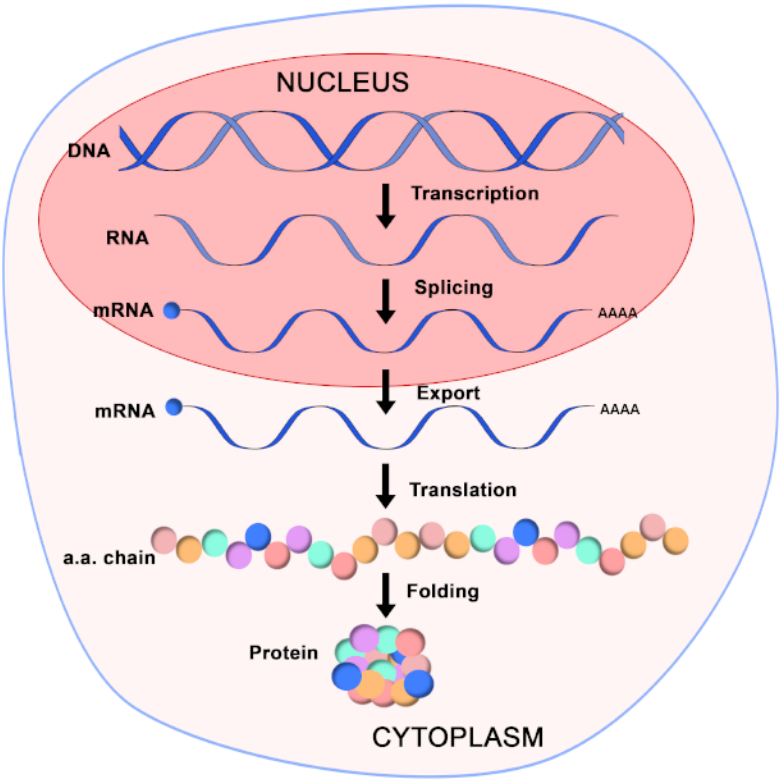

Figure 1-1 Central Dogma of Life. A cartoon representation of a cell illustrating the concept, Central Dogma of Life. DNA is first transcribed into an RNA molecule, the primary transcript. The primary transcript is then spliced to form a mature mRNA (other RNA machinery are also generated – tRNA, rRNA etc. not represented here). Mature

mRNAs are exported into the cytoplasm where the information is translated into an amino acid polypeptide that folds into a functional protein molecule.

Since there are four DNA bases and 20 amino acids, a sequence of three bases called codons, code for each amino acid. Codon sequences within the mRNA dictate the initiation of translation, determine the amino acid sequence and finally terminate the reaction resulting in a particular amino acid polypeptide, that folds into functional proteins. Several different proteins are generated that conduct enzymatic reactions or form important structural elements, such as the DNA packaging units, histones. Other proteins span across the cell membrane, these transmembrane proteins allow for movement of ions and other molecules in and out of the cell, allowing them to communicate with their environment along with many other functions.

1.2. Dexoyribonucleic Acid

Genes, or genetic information is stored in DNA. The first indication that DNA is the 'transforming principle'[1] i.e. it carries genes was demonstrated by an experiment conducted by Oswald and colleagues. DNA is composed of four nitrogen bases either purines, Adenine (A) or Guanine (G), or pyrimidines like Cytosine (C) and Thymine (T), a ribose sugar group and a phosphate backbone. The monomers, referred to as nucleotides constitute a nitrogen base, a pentose sugar and a phosphate head, individual nucleotides connect via a phosphodiester bond formed between the 3'OH group on the sugar to the 5' phosphate head group of the neighboring nucleotide forming a polymeric strand (Figure 1-2a). In 1953[2], Watson and Crick delineated the structure of DNA from the X-ray diffraction images taken by Rosalind Franklin and Raymond Gosling and a key discovery made by Erwin Chargaff[3]; that all organisms have a 1:1 ratio of adenine to

thymine and cytosine to guanine bases. Using these key findings they demonstrated that DNA is composed of two intertwined antiparallel strands that formed a helical structure, the DNA double helix, with the bases at the center forming hydrogen bonds between them (See Figure 1-3). The bases on one strand are complementary to the opposing strand, this allows for each strand to serve as a template during the replication process. Adenine pairs with thymine and guanine pairs with cytosine forming 2 and 3 hydrogen bonds between them, respectively, so the four bases (A, T, G and C) form the genetic coding language. Furthermore, the strands are stabilized by the interactions of the twisted base pair stacks[4]. There are several different configurations of dsDNA (B, A and Z forms)[5], in most organism, at physiological conditions the most common is the B-form of DNA[5]. This form, has a diameter of ~2.2nm, is right-handed and a full twist occurs at every 10.4 nucleotides with a rise of 3.4 nm[6]. The distance between each base pair stack is 0.34nm[6]. Since the strands run anti-parallel to each other the structure twists asymmetrically and forms a major and minor groove, each 2.2 and 1.2nm wide respectively[7] (Figure 1-3 b). These grooves allow specific interactions, or binding sites for proteins, such as transcriptions factors, or DNA/RNA polymerases[8].

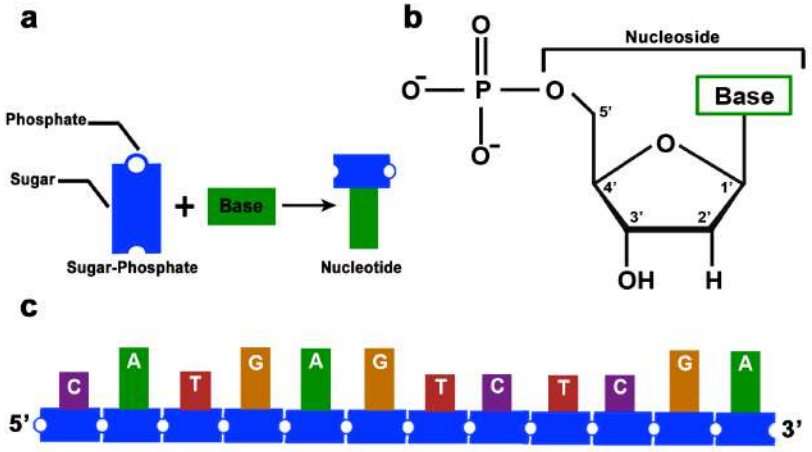

Figure 1-2 Schematic and chemical composition of a nucleotide. (a) Schematic representation of a nucleotide monomer, representing the sugar phosphate head and a nitrogen base. (b) Chemical composition of the 5 carbon deoxyribose sugar with the base bonded to the carbon at 1' and the phosphate head to the 5' C. (c) A cartoon representation of an oligonucleotide, demonstrating the 5' phosphate covalently linked to the 3' OH group of the neighboring nucleotide via a phosphodiester bond.

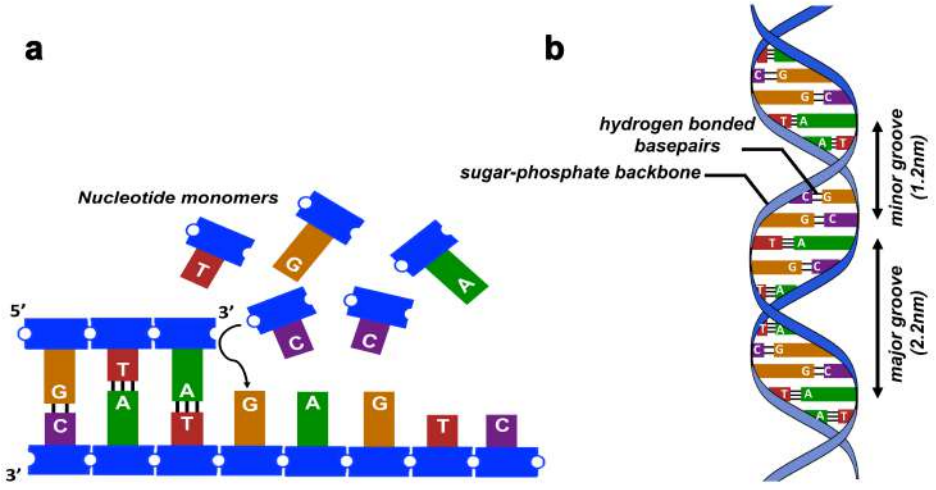

Figure 1-3 Schematic of DNA double Helix. (a) The two strands run anti-parallel to each other, nucleotides are added to the 3' end while the other strand serves as a template. (b) Cartoon representation of the B-form of DNA double helix.

So how is this stored information converted into physical cells, tissues and enzymes? As mentioned earlier (See Chapter 1.1) essentially a three base sequence encodes for an amino acid[9,10]. These three DNA base sequences are called codons, when found in a long series in a stretch of DNA are referred to as a gene. Each gene represents and codes for a sequence of amino acids, which form polypeptides that fold into specific structures, proteins. Genes in DNA are distinguished by another set of codons, the start and stop codons[11], which initiate transcription of gene and subsequently terminate it. In certain cases, codons within genes can get mutated to code for a different amino acid or a stop codon resulting in early termination of the gene, which can result in certain disease states such as cancer.

1.3. Cancer

Cancer is a major public health problem worldwide and is the second leading cause of death in the United States, its estimated that approximately 1,688,780 new cases of cancer and 600,920 patient deaths will occur in 2017 alone[12]. Along with high mortality rates, diagnostic and treatment costs are estimated to increase significantly, ~$158 billion by 2020[13]. Cancer is a disease that arises due to uncontrolled growth of cells, in general the cells in our body have a very defined and structured lifetime. Healthy cells divide and form new cells when necessary, as the cells grow older and their genetic make-up is damaged, they are destroyed in a controlled manner by a process called apoptosis[14]. However, in the case of cancer cells this ordered process breaks down and the cells continue to proliferate uncontrollably generating a large mass referred to as tumors[15]. Cells once they switch and become cancerous, can bypass markers and signals of apoptosis and continue to survive, tumors that do not spread to different regions of the

body are benign. Malignancy is an advanced stage of cancer, where the cancer cells from the primary site of tumor are released, travel through the lymphatic or circulatory system and invade other tissues. The tumor microenviorment is significantly different from healthy tissues, since tumors divide uncontrollable they require excessive amounts of nutrients, cancer cells can signal the generation of increased vascularization to keep up with their energy demands[16]. Cancers generally arise due to mutations in gene bodies that can be hereditary or spontaneous[15,17], environmental stresses[18] such as UV radiation damage, toxins from cigarette smoke and other sources can lead to mutations. Mutations arising specifically in proto-oncogenes, tumor suppressor genes and those of DNA Damage Repair pathway (DDR)[19] along with chromosomal rearrangements[20] are considered to be 'drivers' of cancer. These could be the primary cause of switching to the cancerous state or be the consequence of a primary event. Genetic makeup of cells within the tumor can also be extremely heterogenous and a single cancer type can be a result of mutations that arise at different levels of the signaling pathway. An example of this can be found in colorectal cancer, where the Epidermal Growth Factor (EGFR), a transmembrane receptor is overexpressed in a large percentage of colorectal cancers cases[21,22]. Anti-EGFR monoclonal antibodies can be used to treat colorectal cancer[23,24], except in another form of colorectal cancer where the k-ras gene is mutated, this type is resistant to monoclonal anti-body therapy[25,26]. This is indicative of the extremely complicated nature of cancer, and the necessity to develop affordable early-stage cancer diagnostic methods and personalized therapies. Significant improvement of patient survival and recovery can be achieved with early stage cancer detection and personalized

treatments. Furthermore, genetic mutations aren't the sole contributor to tumorigenesis, it's now been established that epigenetics plays a significant role as well[27,28].

1.4. Epigenetics

Epigenetics was derived from the Greek word *epigenesist*, which literally means extra growth. The prefix *'Epi'* generally means 'on top of' or 'in addition to'. With reference to describing a scientific phenomenon or affect, epigenetics was first coined by C.H. Waddington in 1942[29,30]. He used it to describe the process of cells differentiating into various types from their initial totipotent state. Waddington wasn't too far off, because epigenetic changes as we understand them today, vary as a cell differentiates; therefore they play an important role in development[31,32]. Once the concept of genes and their hereditary nature was established later, Holliday revised it, "the study of the mechanisms of temporal and spatial control of gene activity during the development of complex organisms."[33]. Therefore, epigenetics refers to any heritable (and now encompasses non-hereditary) changes that occur to the genome which do not change or affect the DNA sequence but affect the development of the organism, or expression etc. These are essentially chemical modifications that arise on either histones, or nucleic acids. Several Epigenetic changes have been identified in chromatin[34], as well as within DNA sequences[35], therefore the field is extremely broad. With this in consideration the NIH defines epigenetics as 'epigenetics refers to both heritable changes in gene activity and expression (in the progeny of cells or of individuals) and also stable, long-term alterations in the transcriptional potential of a cell that are not necessarily heritable."[36]

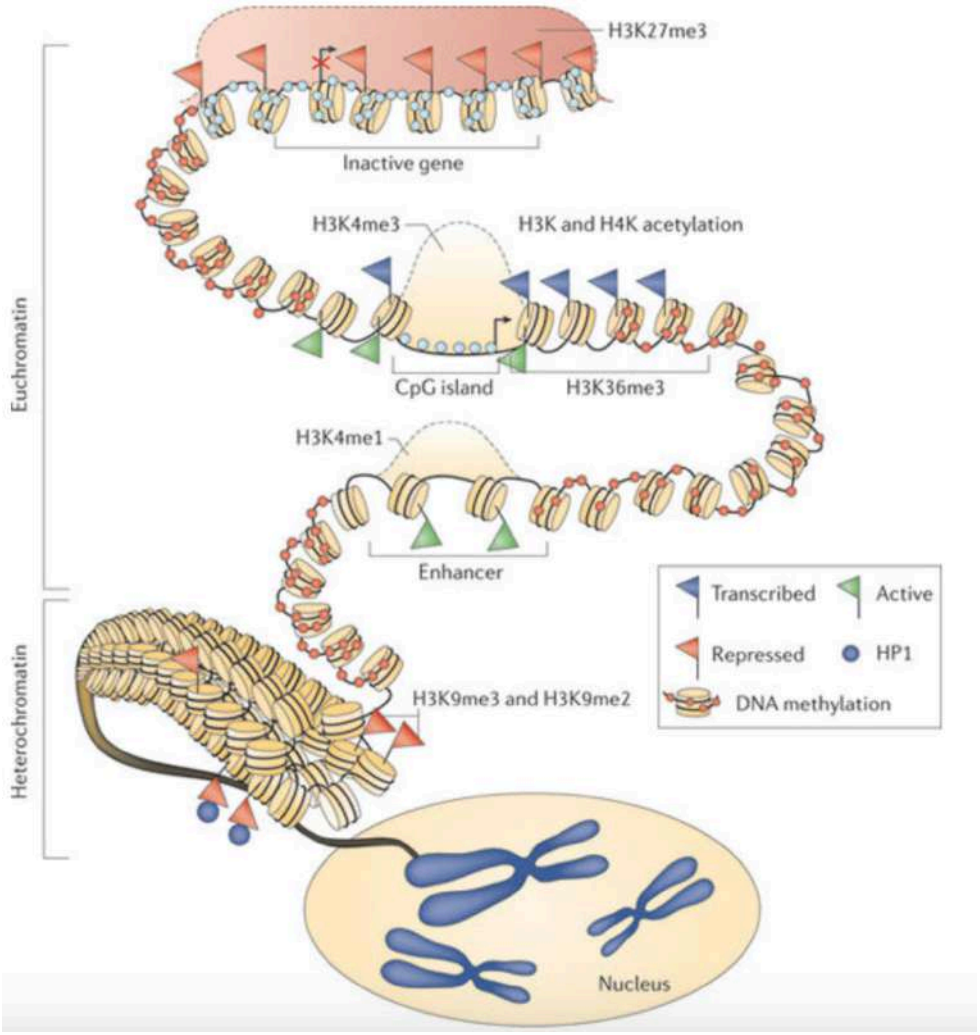

Figure 1-4 The Epigenetic Landscape. Image demonstrates the dynamic epigenetic state of chromatin, representing the repressed or inactived gene region (top) and actively transcribed gene regions (center). DNA is wrapped around nucleosomes, the transcribed gene regions that are 'loosely packed' are referred to as Euchromatin. Nucleosomes are further wrapped into 30nm thick fibers, referred to as the heterochromatin. Both the nucleosomes and DNA contain epigenetic marks, the red flags represent histones containing gene silencing or repressive modifications whereas the blue flags indicate histones with transcriptional permissive modifications. Green flags represent an actively transcribed gene body and its corresponding enhancer. Active gene regions and enhancer

are not heavily methylated (blue circles) whereas silent regions (red circles) are marked by 5-methylcytosine modifications. The red shaded region refers to an inactive gene state occupied by histone modifying proteins that add repressive modifications. These marks can be reversed by another set of activating histone modifying proteins and initiate gene transcription. Reprinted by permission from Macmillan Publisher Ltd. Nat. Can. Rev., Ref [**27**], copyright (2011).

Epigenetic modifications arise within the DNA sequence and the Chromatin. Nucleosome is the fundamental unit of chromatin that is made up 2 copies each of 4 separate histone protein (H2A, H2B, H3 and H4), about 147 bps of double-stranded (ds-)DNA wraps around it. The N-terminal tails of these core histone proteins carry amino acid residues that are ordained with various chemical modification groups[34]. These modifications allow for the recruitment of several non-histone proteins to the chromatin that carry out specific types of functions, for example initiate DNA repair, replication, transcription and several others[34] (See Figure 1-4). Similarly, at the DNA level several different types of chemical modifications have been identified, the most common is methylation (5-methylcytosine, 5mC) of the cytosine residue[35] (See Figure 1-5). Within the human genome, the 5mC mark almost exclusively occurs in the context of cytosine-phosphate-guanine dinucleotides (CpG). The methyl group is added to a cytosine nucleotide by a class of enzymes called methyltransferases that transfer a methyl group from S-adenosyl-L-methionine (SAM) to the carbon 5 atom of the cytosine ring. CpG dinucleotides are dispersed across the genome with significant repeating units of CpG dinucleotides referred to as CpG islands[37]. CpG islands are often but not always present in gene promoter regions, the degree of methylation represents the transcriptional state of the

gene[38] (Figure 1-4). 5mC methylation marks in CpG dinucleotides occur globally in highly repetitive regions referred to as satellites, these methylation marks are important for maintaining the genome stability[39]. Epigenetic modifications are dynamic and changes occur in 5mC marks, localized gain of methylation (hypermethylation) and loss of methylation (hypomethylation), are often associated with a diseased state[27]. 5mC is the most commonly studied epigenetic mark, decades of research has given insight into the location and functional roles of this modification in healthy and tumor cells. It is now widely accepted that cancer development arises due to a combination of genetic and epigenetic anomalies[28]. Significant focus on studying epigenetics and its role in cancer has given rise to the concept and importance of the cancer epigenome[27]. Although whether epigenetic changes are a cause or consequence of cancer development is yet to be determined, a clear link between the two has been established.

For several decades, 5mC was considered to be the predominant DNA epigenetic modification, until less than a decade ago an oxidized form of 5mC, 5-hydroxymethylcytosine (5hmC) was unequivocally identified by the Rao and Heintz group independently[40,41]. Upon further investigation it was established that 5hmC can be oxidized by the same Tet family of Fe(II)- and α-ketoglutarate (α-KG) dependent dioxygenases to 5-formylcytosine (5fC), and ultimately to 5-carboxylcytosine (5caC)[42]. The latter two modifications are excised by thymine DNA glycosylase[43,44], therefore the action of TET dioxygenases and subsequently the Base Excision Pathway enzymes, formulate the active demethylation pathway[43]. Recent reports have led to establishing 5hmC role's in cellular differentiation[45], neurodegenerative disorders[46] and cancer[47]. Intense research into epigenetic changes occurring in cancer has lead to the understanding

of the cancer epigenome. Genetic mutations are not the only changes that occur in cancer but significant restructuring of epigenetic modifications of the chromatin and DNA has been observed and its been established that epigenetic abnormalities also contribute to carcinogenesis[27]. Understanding and determining when and where these aberrant epigenetic marks arise can give us insight into the development of tumor. For instance, abnormal gain of the 5-methylcytosine mark (hypermethylation) has been observed in gene bodies of several types of cancer[48]. Recognizing regular epigenetic patterns and their evolution in particular cancers can allow us to establish certain modifications as important biomarkers of cancer development and progression.

1.4.1. 5-Hydroxymethylcytosine

Since its discovery in 2009[40,41], research has been conducted to determine the biological role of 5hmC in the mammalian genome. Evidence suggests that 5hmC is involved in cellular differentiation[45], aging[49] and neurological disorders such as Alzheimer's[50]. With respect to cancer, a promising discovery was made, the 5hmC mark is significantly depleted in all forms of cancer[51]. As opposed to 5mC, which has varying levels of methylation patterns in cancer cells, 5hmC modification is drastically reduced. This is a ubiquitous phenomenon observed in all cancer types and has been observed by several groups independently[52–54]. These results have established quantification of 5hmC is an ideal biomarker target to determine an individual's cancer susceptibility.

1.4.2. Detection of 5-Hydroxymethylcytosine

Two main techniques have been developed to probe 5hmC position and amounts in the genome, Oxidative Bisulfite Sequencing (OxBS-Seq)[55] and Tet-Assited Bisulfite

Sequencing (TAB-Seq)[56]. Both methods are performed in conjunction with classical bisulfite sequencing along with a second reaction pathway wherein the genomic sample is first treated with either an oxidizing chemical or with a set of enzymes, respectively.

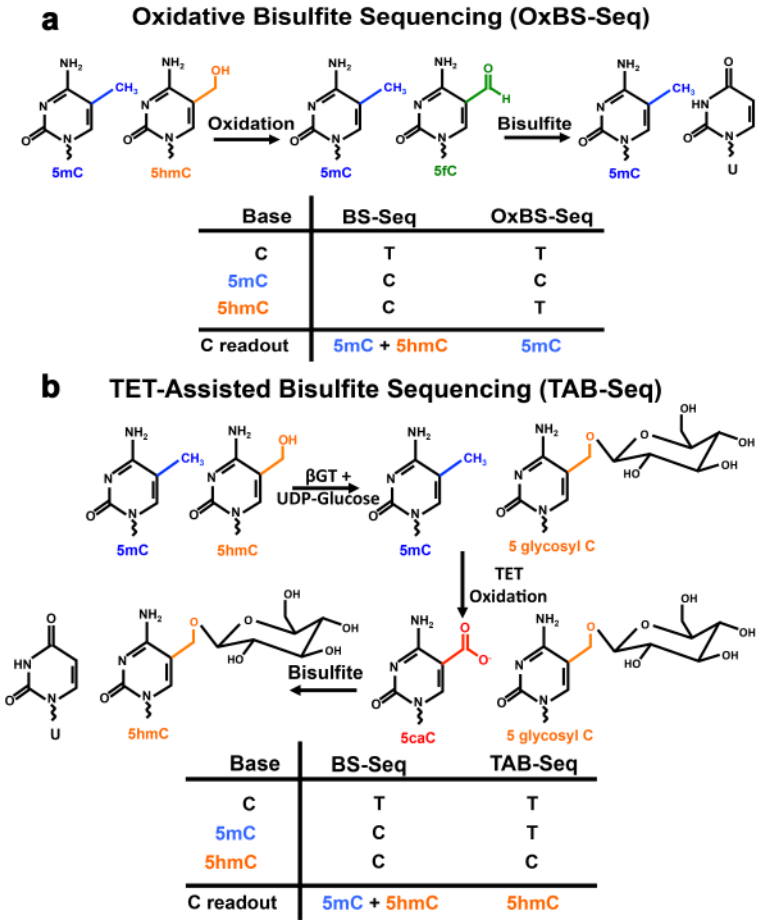

Figure 1-5 Methods to probe 5hmC. (a) Reaction scheme and sequencing output of oxidative bisulfite sequencing. The oxidation reaction converts 5hmC to 5fC, subsequent BS treatment results in only 5mC being read as C, 5hmC can be distinguished by comparing output with traditional BS sequencing. (b) In TAB-Seq, 5hmC is blocked by glycosylation, subsequent treatment with TET oxidizes 5mC to 5caC. BS treatment of the sample results in 5hmC as the only base that is read as C, comparison with regular BS-Seq allows for 5mC discrimination.

In the case of OxBS-Seq the oxidation step converts 5hmC to 5fC, subsequent treatment by BS, results in 5fC being converted to uracil (and appears to be read as thymine in sequencing reaction). Comparative analysis with classical BS reaction, which converts all cytosine to uracil but reads both 5mC and 5hmC as cytosine determines the 5hmC position (Figure 1-5 a). In TAB-Seq, a UDP-glucose molecule is added to the hydroxyl group of 5hmC by β-glucosyltransferase (β-GT), samples are then treated with excess amounts of TET enzyme that oxidizes 5mC to either 5fC or 5caC, while 5hmC remains unaltered due to the presence of UDP-glucose. Subsequent treatment with bisulfite results in 5mC being converted to uracil which is then read in the sequencing reaction as thymine as opposed to 5hmC which is continued to be read as cytosine.

Another method recently developed by the He group, nano-hmC-seal was used determine genomic regions where 5hmC was enriched with limited sample input[57]. Here, they adapt a method developed by Adey et al.[58] where a Tn5 transposase is used to fragment the genomic DNA down to ~300-500 bps length fragments and incorporate sequencing-compatible adaptors in a single step. Subsequently, they treat the genomic DNA with a synthetic form of UDP-glucose containing an azide group in the presence of β-GT, which incorporates the UDP-N_3-Glucose to 5hmC. Biotin linkers can then be attached to the azide group by copper free click chemistry, affinity purification by streptavidin-coated beads collects any sequence regions where 5hmC is present. These genomic regions are amplified using primers that selectively anneal to the linkers previously added to the ends of the genomic fragments, subsequent sequencing of the amplified library determines the genomic regions 5hmC is prevalent in. Analysis of these regions established that 5hmC is present in transcriptionally active gene regions,

associated with permissive histone modifications or 'open' chromatin marks[57]. A subsequent study by the group using nano-hmC-seal technique was conducted on liquid biopsy samples collected from over 250 patients suffering from 5 different types of cancer and compared those to ~100 healthy individuals[59]. They conducted a bioinformatic analysis and found unique set of genes being overexpressed in cancer patients and demonstrated that once again 5hmC was enriched in active gene bodies, associated with open chromatin marks. Furthermore, the gene expression profiles of the liquid biopsy samples were comparable to the cancer tissue biopsy profile. They confirmed this finding in patient derived xenograft mouse models, where they analyzed the tumor 5hmC gene profiles of the cancer tissue xenograft to the 5hmC gene profiles of the cfDNA collected from the mouse blood plasma. Li et al. finally concluded out of the five different cancer types they studied, 5hmC was an ideal cancer-specific biomarker for colorectal and gastric cancers, and showed strong evidence to suggest that it is possibly a superior method to conventional cancer biomarker determinants.

While all of these methods have shed significant insight into the role of 5hmC and established its importance as a cancer-specific biomarker, they are all based on next-generation sequencing analysis, which generates huge amounts of data and requires cumbersome analysis to determine the position and content of 5hmC. Similarly, High-Performace Liquid Chromatography (HPLC)[60,61], the gold standard method for quantification of total nucleic-acid content has been used to determine 5hmC (and other epigenetic modification content), however, it suffers from requiring sophisticated equipment, high sample concentration, (≥1μg) prolonged experimental run time and loss

of sequence information. Therefore, both methods can be difficult to implement in regular clinical workflow and screen individuals for cancer susceptibility.

To overcome these drawbacks, alternative highly sensitive, affordable and rapid methods are required for detection of 5hmC content. Thus, allowing them to be implemented in clinics for regular monitoring of epigenetic modifications such as 5hmC and others.

An optical detection technique developed recently by the Ebenstein lab[62], is identical to the labeling method pioneered by Song et al.[63], where a UDP-N_3-Glucose molecule is enzymatically added to 5hmC but instead of attaching a biotin label using click chemistry, the Ebenstein lab employs a fluorophore linker. They determine the relative change of 5hmC content between healthy and tumor tissues by measuring the fluorescent labels intensity, lower intensity is observed for the tumor samples thereby correlating cancer incidence by measuring the 5hmC content. An ELISA based immunoassay was developed [64], where the signal is enhanced by utilizing a biotinylated secondary antibody, which can then subsequently bind to at least three horseradish peroxidase (HRP) enzyme units instead of the traditional single (or two) HRP. The assay was used to screen 5mC, 5hmC and 5caC modification content from liquid biopsies of patients suffering from lung, pancreatic and bladder cancer and compared to healthy individuals. They demonstrated ability to measure reduced 5hmC content of patients suffering from lung cancer, while no significant change was observed for the other cancer types in comparison to healthy individuals. However, even in the case of lung cancer patient samples the error was significant and overlapped with epigenetic content measured in healthy individuals. While both studies demonstrated ability to detect and quantify 5hmC

content inexpensively, the former method while highly sensitive requires extensive DNA sample preparation, and the latter lacks specificity and sensitivity. Therefore, techniques that can rapidly quantify 5hmC content (as well as other epigenetic base modifications) with high-sensitivity and selectivity would be ideal for implementation in a clinical setting. A goal of this research has been to realize a novel, highly sensitive, single-molecule detection assay that can rapidly quantify 5hmC (See Chapter 5.1).

1.4.3. Detection of Formyl- & Carboxylcytosine

As mentioned earlier, 5mC and 5hmC are not the only cytosine epigenetic marks, 5fC and 5caC are other oxidized forms of cytosine base modifications. In order to investigate these modifications, a handful of methods have been developed that sequence 5fC and 5caC at single-base resolution. However, all of them involve chemical treatment performed in conjunction with traditional bisulfite sequencing to determine the position of the base modification. Two methods were recently developed to address 5fC, reduced bisulfite sequencing (redBS-Seq)[65] and 5fC-assisted bisulfite sequencing (fCAB-Seq)[66] by the Balasubramanian and He group respectively (**Figure 1-6**a). Both involve chemically blocking 5fC thereby preventing its deamination during bisulfite treatment. Therefore 5fC along with 5mC and 5hmC is read as C. The read out is compared to traditional bisulfite sequencing where untreated 5fC is deaminated and is therefore read as T, thus identifying the position of 5fC.

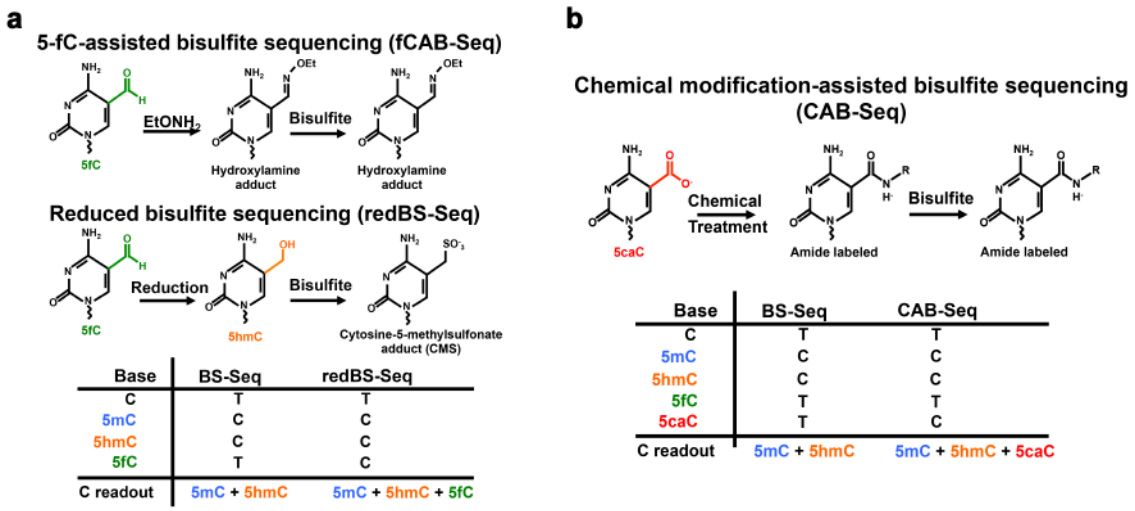

Figure 1-6 Schematic and read out of methods developed to probe 5fC and 5caC. (a) Two methods, fCAB-Seq and redBS-Seq distinguish 5fC and have identical outputs. In fCAB-Seq (top), sample is chemically treated with hydroxylamine to form an adduct that is resistant to BS deamination. In redBS-Seq, chemical treatment results in selective reduction of 5fC to 5hmC, which is also resistant to deamination by BS. In both cases, 5fC is read as C and therefore can be distinguished by comparing results to BS-Seq. (b) In CAB-Seq, chemical treatment results in the addition of an amide group to 5caC, thereby protecting it from deamination by BS. 5caC is therefore read as C and can be differentiated by comparing results to BS-Seq.

A similar method was developed to sequence 5caC, chemical modification assisted bisulfite sequencing (CAB-Seq)[67], wherein DNA is first treated with 1-ethyl-3-[3-(dimethylamino)propyl]-carbodiimide hydrochloride (EDC), which selectively reacts with 5caC and forms an amide (**Figure 1-6**b). The amide labeled 5caC is resistant to deamination by subsequent sodium bisulfite treatment, and is therefore read as C whereas 5fC and C are deaminated and read as T. Subtraction of regular BS-Seq data from CAB-Seq, where C, 5fC and 5caC are all deaminated and thus read as T (5mC and 5hmC are

both unaffected, read as C) gives quantitative information on 5caC. Methylase assisted bisulfite sequencing (MAB-Seq)[68,69]is a subtraction independent method that addresses both 5fC and 5caC but suffers from the inability to distinguish between the two. Here, the sample is first treated with a bacterial DNA CpG methyltransferase M.SssI, that converts all cytosines (within the CpG context) to 5mC. Subsequent treatment with BS results in the deamination of 5fC and 5caC that are read T as opposed to all cytosines (converted to 5mC), 5mC and 5hmC which are resistant to BS and therefore read as C. A slight modification to the latter method was made to develop 5caC methylase-assisted bisulfite sequencing (caMAB-Seq)[69], here 5caC can be resolved at single base resolution. caMAB-Seq employs an additional chemical treatment post methyltransferase reaction, NaBH4 is introduced that selectively reduces 5fC to 5hmC. Subsequent BS treatment deaminates 5caC exclusively and is read as T whereas all other cytosines are protected since they are either in the form of 5mC or 5hmC and therefore read as C. All of these methods require BS treatment and sequencing in order to get relevant information, thus suffer from damaging the target DNA significantly. Novel BS free methods are required that can generate important biologically relevant results without conducting whole genome sequencing, similar to 5hmC-nano-seal (see 1.4.2).

1.5. MicroRNA

MicroRNAs are single-stranded, non-coding RNA molecules that are approximately 18-24 nucleotides in length and play an important role in eukaryotic post-transcriptional gene silencing[70]. An important aspect is that they are readily found in the blood plasma as cell-free nucleic acids and are therefore an ideal target for minimally invasive liquid biopsies for cancer diagnosis and progression[71]. Over 2500 microRNAs have been

recognized[72] and there is strong evidence demonstrating at least two-thirds of human genes are targeted by microRNAs[73]. The origin of cell free microRNAs is still a hot topic of debate, with two main hypotheses. Passive release from apoptotic cells [74,75] or shuttled from cells in either vesicles[71] or bound to RNA-binding protein such as Argonaute[76]. While they might arrive in blood plasma through one or all of the aforementioned sources, the degree and extent to which each contributes is also contentious. However, what is important and established is that cell free microRNA can be utilized for cancer screening and specific microRNAs correlate to the type of cancer. MicroRNA expression profiles are unique to not only the cancer type but give insight to the stage of tumor development[77]. All of these characteristics make microRNAs an ideal nucleic acid biomarker for cancer screening and regular monitoring of patients undergoing treatment.

1.5.1. Detection of MicroRNAs

Several methods for microRNA detection have been developed, these include, quantitative reverse transcription PCR (qRT-PCR), microarray-based hybridization methods and next generation sequencing with qRT-PCR among them being them most popular[78]. However, due to their small size and minor sequence differences, detection of these short nucleic acids can be challenging, as conventional methods tend to introduce bias and lack specificity[78,79].

To overcome these short-comings, nanotechnology based methods are being developed. Recently, the Mirkin group employed spherical nucleic acid (SNA) gold nanoparticle conjugates to screen for prostate cancer specific mircoRNAs[80]. Label-free electrochemical-based assays have been utilized as well to monitor cfmicroRNAs. However, sophisticated fabrication methods are necessary to fashion these devices and a

diagnostic tool has yet to be realized. Therefore, novel highly sensitive techniques are required that can rapidly monitor these important disease biomarkers. Nanopore-based sensors[81] have demonstrated tremendous potential for molecular sensing of biomolecules and are an ideal tool for the task.

1.6. Nanopores

The detection principle of Nanopore sensing is based on an invention developed by Wallace H. Coulter in the late 1940's, commonly referred to as the Coulter Counter[82].

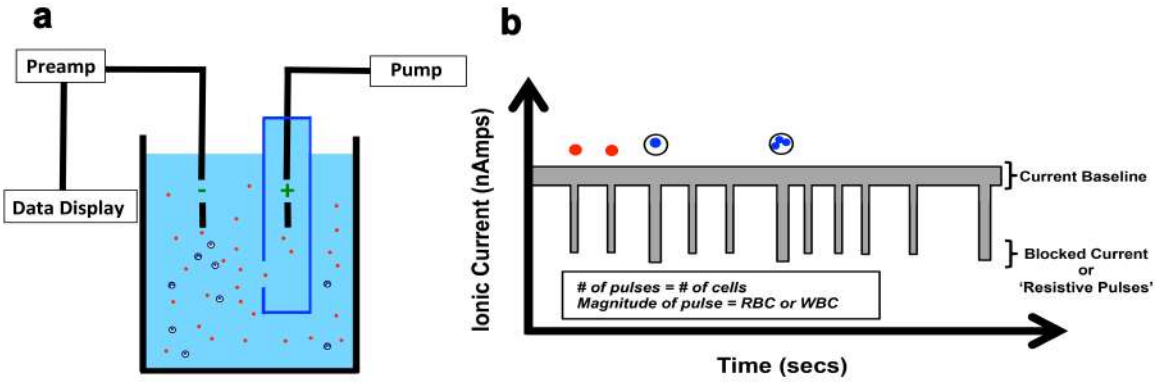

Figure 1-7 The Coulter Counter. (a) A schematic representation of the Coulter Counter demonstrating the sensing principle, an orifice separates two chambers, a voltage bias generates a measureable ionic current. Solution from one chamber (containing cells) is pumped into the other and presence of cells blocks the current. (b) Schematic of typical results from instrument, a baseline ionic current level is measured across the chambers, presence of cells generate resistive pulses. The frequency of the pulses is indicative of the number of cells in a particular sample and the magnitude of the resistive pulses represents either RBC or WBC.

The initial design of the invention would appear primitive; an orifice was generated in a cellophane cigarette wrapper using a hot needle. This aperture with a diameter of several

microns, is placed between two chambers filled with electrolyte solution, application of a voltage bias across the chambers resulted in a measureable ionic current (See Figure 1-7). Blood sample is introduced to one of the chambers (*cis*) and pressure applied to sample containing chamber such that solution flows through the orifice into the second chamber (t*rans*) (Figure 1-7a). For a pure electrolyte solution the ionic current remains unperturbed, however in the presence of cells, discrete *downward spikes* or *resistive pulses* are observed indicating the solution conductivity is temporarily reduced. These pulses are generated when cells are present within the orifice and occupy the sensing volume thus displacing ions and temporarily blocking the ionic current. The total number of pulses measured over time co-relates to the total amount of cells present in the sample and the magnitude of each ionic current drop is indicative of the cell size (Figure 1-7b). Therefore, not only can you rapidly determine the number of cells in a given sample you can also differentiate between the cell type (Red Blood Cells or White Blood Cells). The Coulter Counter revolutionized the field of hematology and is still widely used in hospitals to this day.

Nanopore sensing (Figure 1-8) is also a resistive pulse based technique that has tremendous potential to usher in the next revolution in diagnostic medicine. In this case, however the aperture dimensions are in the nanoscale and the target analytes aren't micron-sized cells but individual biomolecules such as DNA[83], RNA[84] or proteins[85,86]. In a typical setup, a nanopore is placed between two chambers containing electrolyte solution; application of a voltage bias across the chambers generates a strong electric field at the pore. Charged biomolecules are electrophoretically driven through the pore individually and their presence within the pore perturbs the measured ionic current

(Figure 1-8c). These brief electrical pulses are commonly referred to as 'events' and the electrical profile can be used to characterize biomolecules, or study biomolecular interactions[87].

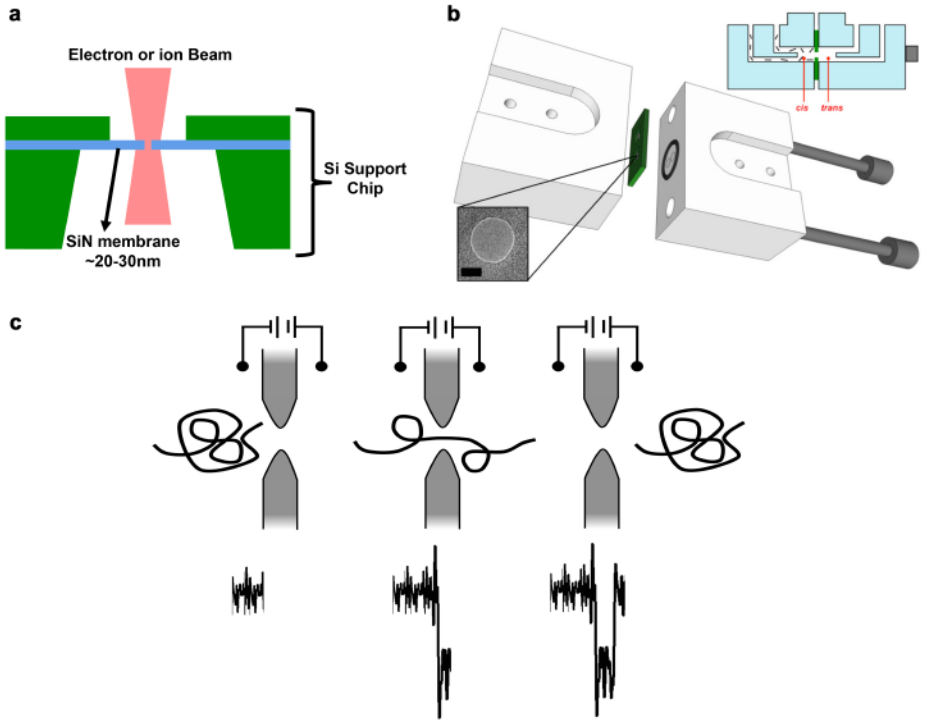

Figure 1-8 Nanopore device and detection. (a) Schematic of a thin SiN membrane (20-30nm thick) supported on a silicon chip forms the basis of the device. A high-energy beam is used to ablate the atoms to fabricate a pore. (b) Chip containing a single nanopore is mounted on a custom designed flow cell. A lateral view of the final device is represented (top-right), demonstrating the nanopore serving as the only connection between the cis and trans chambers of the device. Lower left inset is a TEM image of a typical nanopore fabricated in SiN membrane. Scale bar is 10nm. (c) Cartoon representation of the translocation process with corresponding ionic current signal below: baseline open-pore current (bottom, left); threading of the molecule decreases ionic

current (center); current returns to baseline upon molecular passage (bottom, right). The total electrical signal is referred to as an event.

1.6.1. Biological Nanopores

Nanopore based biomolecular sensing was initiated using a transmembrane protein channel, the α-hemolysin toxin isolated from the bacterium *staphylococcus aureus*. This heptameric porin, with an approximate diameter of 2nm, self-assembles and easily embeds into lipid bilayer membranes[88]. In 1999 Kasianowicz et al. [89] demonstrated the detection of RNA homopolymers through the pore, and sparked the possibility of developing a next-generation sequencing platform based on nanopore ionic current detection. They hypothesized that each base would block a certain amount of current based on its unique structure as it passed through the pore and the ionic current signatures would therefore contain the sequence information of the DNA or RNA molecule. The first confirmation of this hypothesis was provided by Akeson et al.,[90] they investigated the translocation of an RNA block copolymer that consisted of 70 Cytosine and 30 Adenine bases, respectively. The ionic current signatures of this polymer resulted in two distinct current blockade levels, corresponding to the levels previously observed for homopolymers of these bases. This provided compelling evidence that under the correct parameters, nanopores could be used for sequencing. Over a decade of experiments were conducted by various groups to delineate the ideal conditions for sequencing, initially it was hypothesized that an exonuclease tethered atop the α-hemolysin channel across which a voltage bias is applied can determine the sequence[91]. The exonuclease cleaves off nucleosides from a target polymer and each cleaved nucleoside would be detected as the porin electrophoretically captures it. However, this strategy results in loss of sequence

information, the process is stochastic and not every cleaved nucleoside is captured. The solution was provided by utilizing a phi29 polymerase that slows the translocation speed of the target polymer being sequenced[92,93]. The polymerase adds nucleotides to one of the strands with free 3' end in a step-wise single nucleotide addition while the complementary strand occupies the pore sensing region and gradually ratchets through the pore. As the polymerase continues the extension the template strand is gradually pulled out of the pore, and the ionic current profile corresponds to all the bases occupying the α-hemolysin porin. Similar experiments pioneered by the Akeson group were performed using the MspA porin isolated from mycobacterium smegmatis[94]. Both systems have successfully demonstrated the capability to sequence nucleic acid targets.

Several other porins such as ClyA[95] and OmpG[96] have also been used to detect a variety of proteins. However, biological nanopores while highly reproducible suffer from geometric constraints, most pore channel diameters are no larger than ~2nm and can therefore only allow biomolecules less than 2nm in diameter to translocate through them. The lipid bilayers that these channels are embedded in are highly unstable and tend to disrupt easily rendering the system ineffective. Biological nanopores, especially with larger (>2nm) diameters are not ideal for molecular detection, as most of porins exhibit gating, a phenomenon where the ionic current measured through the pore varies significantly due to stochastic structural changes of the channel walls. Over 25 years of research has resulted in only a few biological nanopores that can be utilized for molecular analysis.

1.6.2. Solid-State Nanopores

To overcome these stability and geometric constraints nanopores were fabricated in thin free-standing solid state membranes[81]. These nanopore are highly robust, mechanically stable and can withstand high temperature as well as a variety of solvent conditions[97–99].

The pore diameter and membrane thickness[100,101] can be fine-tuned depending on the desired application. The first demonstration[102] of a nanopore device was conducted using a high energy Ar+ ion beam, however this method lacked sensitive control of the desired diameter. Subsequently, a fabrication method utilizing Transmission Electron Microscope, wherein an electron beam is used to form pores was developed[103]. This technique has fine-control over the nanopore dimensions and is currently the most popular method for fabrication of SS-nanopore devices. Other methods such as dielectric breakdown have also been developed[104], here a high bias is applied across a thin solid-state membrane, over time this leads to the formation of a nanopore. A novel method developed in the Hall lab takes advantage of focus ion beam milling, using the Helium Ion Microscope (HIM), this has significant advantages over other methods with respect to fabrication time and throughput[105]. By tuning total ion dose we can precisely control the nanopore device dimensions as can be seen in Figure 1-9 a. Furthermore, the method allows for rapid fabrication of nanopore array devices (see Figure 1-9 b) that can be employed for high-throughput screening techniques and simultaneous optical and electrical detection[106] setups. Another distinct advantage of the HIM is the ability to conduct localized reduction of membrane thickness, and subsequently fabricating pores within the thinned region[101]. Standard silicon fabrication techniques generate $\geq$ 10nm thick free-standing SiN membranes, it is difficult to achieve sub 10nm membrane

thicknesses, although certain protocols have been developed to achieve these dimensions[107]. Reduction of membrane thickness enhances signal-to-noise ratio, which is useful and in some cases necessary for detection of certain types of molecules. A variety of solid-state materials have also been employed for nanopore sensing, such as HfO_2[108], including two-dimensional materials such as graphene[109–111], and MoS_2[99] that are promising materials for sequencing applications.

A detailed account of signal enhancement and interpretation of translocation dynamics using nanopores fabricated in sub-10nm membranes is discussed in Chapter 2. *All experiments in this study were conducted with devices fabricated by the Helium Ion Microscope, See* Appendix *for nanopore fabrication protocol.*

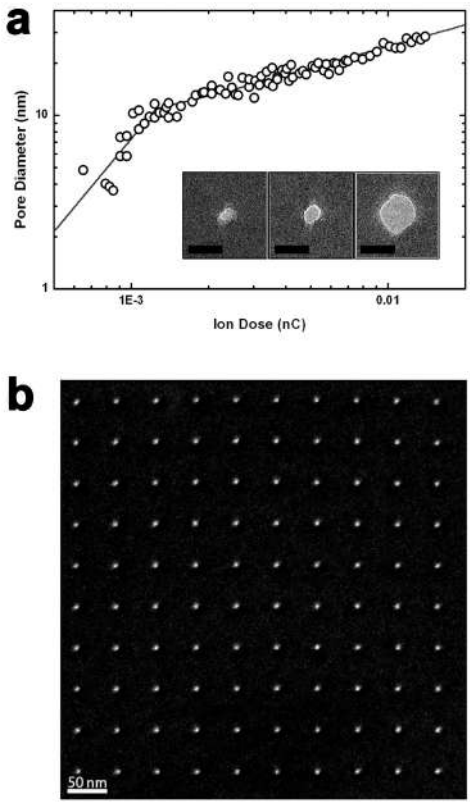

Figure 1-9 Nanopore devices fabricated by HIM. (a) Log–log plot of He+ dose versus resultant nanopore diameter, showing a rapid fabrication regime <10nm and a slow one

above that. Solid lines are power law fits to the respective regions. Inset: TEM images of individual nanopores with diameters (L-R) of 5, 10 and 20 nm. Scale bars are 20 nm. (b) STIM micrograph of a 10 × 10 array of 5 nm diameter nanopores.

1.6.1. Applications of Nanopores: Sequencing

The inception of Nanopore sensing was based initially to develop a rapid single-molecule sequencing device. In 2014, Oxford Nanopore Technologies a nanopore based sequencing company founded in 2005, released a portable handheld sequencer, the MinION[112]. Over 25 years of intense research finally culminated in the development of a nanopore device that can sequence 'ultra-long' fragments (~80 kbp) and has been used to sequence the entire human genome[113]. The ability to sequence long-length fragments has significant advantage over conventional sequencing technologies, wherein the target nucleic acid molecule is fragmented to short lengths (<500 bp) and post sequencing has to be aligned back together like a jigsaw puzzle. However, nanopore sequencing does not suffer from this constraint, and this ability to sequence long length fragments on a portable device has allowed identification of novel antibiotic resistance islands[114], monitor the Ebola epidemic on-site[115] and determine the transmission of the Zika virus[116].

1.6.2. Applications of Nanopores: Detecttion of Nucleic Acid Biomarkers

Applications are not limited to sequencing, SS-nanopores specifically have been used to probe DNA[102,117], study RNA interactions with small molecules like antibiotics[118], differentiate various tRNA species[119], study DNA-protein interactions and utilized for genome mapping[120,121]. The first indication of genomic mapping was demonstrated by Kowalczyk et al.[120], where dsDNA was investigated with varying concentrations of the RecA protein. The authors observed individual events with multiple ionic current

blockade levels, corresponding to either just the DNA or DNA bound by protein, indicating that a single DNA molecule was decorated with discrete protein bound regions. Recently Chen et al.[121] demonstrated mapping a specific 4 base palindromic sequence TCGA, recognized by a methyltransferase. The authors used a S-adenosyl methionine analog that contains a biotin modification and therefore tricks the enzyme to add a biotin modification instead of a methyl group[122]. Labeled DNA was incubated with MS, regions bound by MS excluded more ions and therefore blocked more current. The protein bound level could easily be distinguished from the DNA level, and total number of spikes translates to the total number of that specific sequence present in the DNA target. The location of the protein bound spike with respect to the start of the event is indicative of the position of the sequence thereby allowing rapid genomic mapping of specific sequences.

Cancer specific biomarkers such as MicroRNAs have also been detected using both SS-[123] and biological nanopores[124]. A lung cancer specific biomarker, miR155 that is over-expressed in cancer patients was detected using a complimentary probe by an α-hemolysin pore[124]. The complementary probe was slightly longer than the microRNA sequence (22 nucleotides) and therefore once annealed generated a ds probe-microRNA hybrid with single-stranded overhangs. The hybrid generated a unique current signature, a prolonged multi-level event indicative of a multi-step process; unzipping of the probe, probe translocation and then subsequent translocation of the target microRNA. Wanunu et al. demonstrated detection of microRNAs using specialized SS-nanopore devices, wherein small <5nm diameter nanopores were fabricated in sub 10nm membranes referred as ‘ultra-thin’ nanopore devices. They observed significant improvement in

signal to noise ratio using ultra-thin nanopore devices and were able to detect microRNAs extracted from murine liver tissue. While both approaches demonstrate successful and sensitive detection of microRNAs, significant analysis of the current signal is necessary to conclusively determine presence of target analyte. Furthermore, in the case of Wanunu et al., enrichment of target microRNAs coupled with ultra-thin nanopore devices that are challenging and time consuming to fabricate were an essential requirement for detection. *We have utilized a novel assay for assessing microRNAs that does not require any sophisticated device modification or signal analysis, as demonstrated in Chapter 4.1.*

Nanopores have been used to study epigenetic modifications, specifically those occurring on the cytosine base and oxidative damage adducts like 8-oxoguanine. Biological pores have assessed the ability to sequence these modifications and determine their exact positions[125]. Presence of 5-methylcytosine in target nucleic acid sequences was determined by investigating binding of methylated CpG binding proteins (MBD1) to methylated CpG regions. The authors demonstrated they could distinguish and quantify DNA targets containing 5mC modifications. Those nucleic acids generated a significantly different electronic signature due to binding of MBD1 proteins from non-methylated DNA molecules[126]. Nucleic acid targets containing 5-hydroxymethylcytosine[127] and recently 5-hydroxyuracil[128] were studied by SS-nanopores, these reports were conducted on label-free molecules. However, ultra-thin nanopore devices were necessary for detection of molecules containing these epigenetic marks. Furthermore, molecules with >15-30% of 5hmC modification were successfully differentiated, these synthetic targets aren't an accurate representation of physiological genomes where the 5hmC content is

sparse[35,129]. *To date, no attempt has been made to detect or quantify genomic epigenetic content using SS-nanopores, chapter 5.1 discusses at length our successful demonstration of 5hmC content analysis of murine brain genome.*

References

1. Studies on the chemical nature of the substance inducing transformation of pneumococcal types | JEM. Available at: http://jem.rupress.org/content/79/2/137. (Accessed: 5th November 2017)

2. Watson, J. D. & Crick, F. H. C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **171,** 171737a0 (1953).

3. Chargaff, E., Lipshitz, R. & Green, C. Composition of the desoxypentose nucleic acids of four genera of sea-urchin. *J. Biol. Chem.* **195,** 155–160 (1952).

4. Yakovchuk, P., Protozanova, E. & Frank-Kamenetskii, M. D. Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic Acids Res.* **34,** 564–574 (2006).

5. Ghosh, A. & Bansal, M. A glossary of DNA structures from A to Z. *Acta Crystallogr. D Biol. Crystallogr.* **59,** 620–626 (2003).

6. The dimensions of DNA in solution - ScienceDirect. Available at: http://www.sciencedirect.com/science/article/pii/0022283681900991. (Accessed: 5th November 2017)

7. Wing, R. *et al.* Crystal structure analysis of a complete turn of B-DNA. *Nature* **287,** 287755a0 (1980).

8. Privalov, P. L. *et al.* What Drives Proteins into the Major or Minor Grooves of DNA? *J. Mol. Biol.* **365,** 1–9 (2007).

9. Nirenberg, M. W. & Matthaei, J. H. The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides. *Proc. Natl. Acad. Sci.* **47,** 1588–1602 (1961).

10. Crick, F. H., Barnett, L., Brenner, S. & Watts-Tobin, R. J. General nature of the genetic code for proteins. *Nature* **192,** 1227–1232 (1961).

11. Edgar, B. The Genome of Bacteriophage T4. *Genetics* **168,** 575–582 (2004).

12. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2017. *CA. Cancer J. Clin.* **67,** 7–30 (2017).

13. Cancer costs projected to reach at least $158 billion in 2020. *National Institutes of Health (NIH)* (2015). Available at: https://www.nih.gov/news-events/news-releases/cancer-costs-projected-reach-least-158-billion-2020. (Accessed: 1st October 2017)

14. Alberts, B. *et al.* Programmed Cell Death (Apoptosis). (2002).

15. What Is Cancer? *National Cancer Institute* Available at: https://www.cancer.gov/about-cancer/understanding/what-is-cancer. (Accessed: 1st October 2017)

16. Vaupel, P., Kallinowski, F. & Okunieff, P. Blood flow, oxygen and nutrient supply, and metabolic microenvironment of human tumors: a review. *Cancer Res.* **49,** 6449–6465 (1989).

17. Ames, B. N., Gold, L. S. & Willett, W. C. The causes and prevention of cancer. *Proc. Natl. Acad. Sci.* **92,** 5258–5265 (1995).

18. Ames, B. N. Identifying Environmental Chemicals Causing Mutations and Cancer. *Jurimetr. J.* **20,** 326 (1979).

19. Helleday, T., Petermann, E., Lundin, C., Hodgson, B. & Sharma, R. A. DNA repair pathways as targets for cancer therapy. *Nat. Rev. Cancer* **8,** 193–204 (2008).

20. Mitelman, F., Johansson, B. & Mertens, F. The impact of translocations and gene fusions on cancer causation. *Nat. Rev. Cancer* **7,** 233–245 (2007).

21. Spano, J. P. *et al.* Epidermal growth factor receptor signaling in colorectal cancer: preclinical data and therapeutic perspectives. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* **16,** 189–194 (2005).

22. Krasinskas, A. M. EGFR Signaling in Colorectal Carcinoma. *Pathology Research International* (2011). Available at: https://www.hindawi.com/journals/pri/2011/932932/. (Accessed: 6th November 2017)

23. Meyerhardt, J. A. & Mayer, R. J. Systemic therapy for colorectal cancer. *N. Engl. J. Med.* **352,** 476–487 (2005).

24. Van Cutsem, E. *et al.* Open-label phase III trial of panitumumab plus best supportive care compared with best supportive care alone in patients with chemotherapy-refractory metastatic colorectal cancer. *J. Clin. Oncol.* **25,** 1658–1664 (2007).

25. Lievre, A. *et al.* KRAS mutation status is predictive of response to cetuximab therapy in colorectal cancer. *Cancer Res.* **66,** 3992–3995 (2006).

26. Amado, R. G. *et al.* Wild-type KRAS is required for panitumumab efficacy in patients with metastatic colorectal cancer. *J. Clin. Oncol.* **26,** 1626–1634 (2008).

27. Baylin, S. B. & Jones, P. A. A decade of exploring the cancer epigenome — biological and translational implications. *Nat. Rev. Cancer* **11,** 726–734 (2011).

28. Vogelstein, B. *et al.* Cancer Genome Landscapes. *Science* **339,** 1546–1558 (2013).

29. Waddington, C. H. The epigenotype. 1942. *Int. J. Epidemiol.* **41,** 10–13 (2012).

30. Tronick, E. & Hunter, R. G. Waddington, Dynamic Systems, and Epigenetics. *Front. Behav. Neurosci.* **10,** (2016).

31. Reik, W., Dean, W. & Walter, J. Epigenetic reprogramming in mammalian development. *Science* **293,** 1089–1093 (2001).

32. Cantone, I. & Fisher, A. G. Epigenetic programming and reprogramming during development. *Nat. Struct. Mol. Biol.* **20,** 282–289 (2013).

33. Holliday, R. DNA methylation and gene regulation - DNA methylation and epigenetic inheritance. *Phil Trans R Soc Lond B* **326,** 329–338 (1990).

34. Kouzarides, T. Chromatin Modifications and Their Function. *Cell* **128,** 693–705 (2007).

35. Breiling, A. & Lyko, F. Epigenetic regulatory functions of DNA modifications: 5-methylcytosine and beyond. *Epigenetics Chromatin* **8,** 24 (2015).

36. Roadmap Epigenomics Project - Overview. Available at: http://www.roadmapepigenomics.org/overview. (Accessed: 6th November 2017)

37. Suzuki, M. M. & Bird, A. DNA methylation landscapes: provocative insights from epigenomics. *Nat. Rev. Genet.* **9,** 465–476 (2008).

38. Deaton, A. M. & Bird, A. CpG islands and the regulation of transcription. *Genes Dev.* **25,** 1010–1022 (2011).

39. Ehrlich, M. DNA Methylation and Cancer-associated Genetic Instability. in *Genome Instability in Cancer Development* 363–392 (Springer, Dordrecht, 2005).

40. Tahiliani, M. *et al.* Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. *Science* **324,** 930–935 (2009).

41. Kriaucionis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **324,** 929–930 (2009).

42. Ito, S. *et al.* Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333,** 1300–1303 (2011).

43. Maiti, A. & Drohat, A. C. Thymine DNA Glycosylase Can Rapidly Excise 5-Formylcytosine and 5-Carboxylcytosine: Potential implications for active demethylation of CpG sites. *J. Biol. Chem.* **286,** 35334–35338 (2011).

44. Zhang, L. *et al.* Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA. *Nat. Chem. Biol.* **8,** 328–330 (2012).

45. Ficz, G. *et al.* Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature* **473,** 398–402 (2011).

46. Al-Mahdawi, S., Virmouni, S. A. & Pook, M. A. The emerging role of 5-hydroxymethylcytosine in neurodegenerative diseases. *Front. Neurosci.* **8,** (2014).

47. Ficz, G. & Gribben, J. G. Loss of 5-hydroxymethylcytosine in cancer: Cause or consequence? *Genomics* **104,** 352–357 (2014).

48. Baylin, S. B., Belinsky, S. A. & Herman, J. G. Aberrant Methylation of Gene Promoters in Cancer—Concepts, Misconcepts, and Promise. *JNCI J. Natl. Cancer Inst.* **92,** 1460–1461 (2000).

49. Chen, H., Dzitoyeva, S. & Manev, H. Effect of aging on 5-hydroxymethylcytosine in the mouse hippocampus. *Restor. Neurol. Neurosci.* **30,** 237–245 (2012).

50. Coppieters, N. *et al.* Global changes in DNA methylation and hydroxymethylation in Alzheimer's disease human brain. *Neurobiol. Aging* **35,** 1334–1344 (2014).

51. Pfeifer, G. P., Xiong, W., Hahn, M. A. & Jin, S.-G. The role of 5-hydroxymethylcytosine in human cancer. *Cell Tissue Res.* **356,** 631–641 (2014).

52. Jin, S. G. *et al.* 5-Hydroxymethylcytosine Is Strongly Depleted in Human Cancers but Its Levels Do Not Correlate with IDH1 Mutations. *Cancer Res.* **71,** 7360–7365 (2011).

53. Haffner, M. C. *et al.* Global 5-hydroxymethylcytosine content is significantly reduced in tissue stem/progenitor cell compartments and in human cancers. *Oncotarget* **2,** 627–637 (2011).

54. Yang, H. *et al.* Tumor development is associated with decrease of TET gene expression and 5-methylcytosine hydroxylation. *Oncogene* **32,** 663–669 (2013).

55. Booth, M. J. *et al.* Quantitative Sequencing of 5-Methylcytosine and 5-Hydroxymethylcytosine at Single-Base Resolution. *Science* **336,** 934–937 (2012).

56. Yu, M. *et al.* Base-Resolution Analysis of 5-Hydroxymethylcytosine in the Mammalian Genome. *Cell* **149,** 1368–1380 (2012).

57. Han, D. *et al.* A Highly Sensitive and Robust Method for Genome-wide 5hmC Profiling of Rare Cell Populations. *Mol. Cell* **63,** 711–719 (2016).

58. Adey, A. *et al.* Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. *Genome Biol.* **11,** R119 (2010).

59. Li, W. *et al.* 5-Hydroxymethylcytosine signatures in circulating cell-free DNA as diagnostic biomarkers for human cancers. *Cell Res.* **27,** 1243–1257 (2017).

60. Le, T., Kim, K.-P., Fan, G. & Faull, K. F. A sensitive mass-spectrometry method for simultaneous quantification of DNA methylation and hydroxymethylation levels in biological samples. *Anal. Biochem.* **412,** 203–209 (2011).

61. Chen, M.-L. *et al.* Quantification of 5-Methylcytosine and 5-Hydroxymethylcytosine in Genomic DNA from Hepatocellular Carcinoma Tissues by Capillary Hydrophilic-Interaction Liquid Chromatography/Quadrupole TOF Mass Spectrometry. *Clin. Chem.* **59,** 824–832 (2013).

62. Shahal, T. *et al.* Spectroscopic Quantification of 5-Hydroxymethylcytosine in Genomic DNA. *Anal. Chem.* **86,** 8231–8237 (2014).

63. Song, C. X. *et al.* Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nat. Biotechnol.* **29,** 68–72 (2011).

64. Chowdhury, B., Cho, I.-H., Hahn, N. & Irudayaraj, J. Quantification of 5-methylcytosine, 5-hydroxymethylcytosine and 5-carboxylcytosine from the blood of cancer patients by an enzyme-based immunoassay. *Anal. Chim. Acta* **852,** 212–217 (2014).

65. Quantitative sequencing of 5-formylcytosine in DNA at single-base resolution. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4188980/. (Accessed: 28th October 2017)

66. Song, C.-X. *et al.* Genome-wide Profiling of 5-Formylcytosine Reveals Its Roles in Epigenetic Priming. *Cell* **153,** 678–691 (2013).

67. Lu, X. *et al.* Chemical Modification-Assisted Bisulfite Sequencing (CAB-Seq) for 5-Carboxylcytosine Detection in DNA. *J. Am. Chem. Soc.* **135,** 9315–9317 (2013).

68. Guo, F. *et al.* Active and Passive Demethylation of Male and Female Pronuclear DNA in the Mammalian Zygote. *Cell Stem Cell* **15,** 447–459 (2014).

69. Single-base resolution analysis of active DNA demethylation using methylase-assisted bisulfite sequencing | Nature Biotechnology. Available at: https://www.nature.com/articles/nbt.3073. (Accessed: 28th October 2017)

70. He, L. & Hannon, G. J. MicroRNAs: small RNAs with a big role in gene regulation. *Nat. Rev. Genet.* **5,** 522–531 (2004).

71. Mitchell, P. S. *et al.* Circulating microRNAs as stable blood-based markers for cancer detection. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 10513–10518 (2008).

72. Griffiths-Jones, S., Grocock, R. J., van Dongen, S., Bateman, A. & Enright, A. J. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.* **34,** D140–144 (2006).

73. Friedman, R. C., Farh, K. K.-H., Burge, C. B. & Bartel, D. P. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.* **19,** 92–105 (2009).

74. Laterza, O. F. *et al.* Plasma MicroRNAs as sensitive and specific biomarkers of tissue injury. *Clin. Chem.* **55,** 1977–1983 (2009).

75. Chen, X. *et al.* Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases. *Cell Res.* **18,** 997–1006 (2008).

76. Turchinovich, A., Weiz, L., Langheinz, A. & Burwinkel, B. Characterization of extracellular circulating microRNA. *Nucleic Acids Res.* **39,** 7223–7233 (2011).

77. Lu, J. *et al.* MicroRNA expression profiles classify human cancers. *Nature* **435,** 834–838 (2005).

78. de Planell-Saguer, M. & Rodicio, M. C. Detection methods for microRNAs in clinic practice. *Clin. Biochem.* **46,** 869–878 (2013).

79. Dong, H. *et al.* MicroRNA: Function, Detection, and Bioanalysis. *Chem. Rev.* **113,** 6207–6233 (2013).

80. Alhasan, A. H. *et al.* Circulating microRNA signature for the diagnosis of very high-risk prostate cancer. *Proc. Natl. Acad. Sci.* **113,** 10655–10660 (2016).

81. Dekker, C. Solid-state nanopores. *Nat. Nanotechnol.* **2,** 209–215 (2007).

82. Coulter, Wallace. Means for counting particles suspended in a fluid. (1953).

83. Li, J., Gershow, M., Stein, D., Brandin, E. & Golovchenko, J. A. DNA molecules and configurations in a solid-state nanopore microscope. *Nat. Mater.* **2,** 611–5 (2003).

84. Skinner, G. M., van den Hout, M., Broekmans, O., Dekker, C. & Dekker, N. H. Distinguishing single- and double-stranded nucleic acid molecules using solid-state nanopores. *Nano Lett.* **9,** 2953–2960 (2009).

85. Plesa, C. *et al.* Fast Translocation of Proteins through Solid State Nanopores. *Nano Lett.* **13,** 658–663 (2013).

86. Larkin, J., Henley, R. Y., Muthukumar, M., Rosenstein, J. K. & Wanunu, M. High-Bandwidth Protein Analysis Using Solid-State Nanopores. *Biophys. J.* **106,** 696–704 (2014).

87. Smeets, R. M. M., Kowalczyk, S. W., Hall, A. R., Dekker, N. H. & Dekker, C. Translocation of RecA-Coated Double-Stranded DNA through Solid-State Nanopores. *Nano Lett.* **9,** 3089–3095 (2009).

88. Song, L. Z. *et al.* Structure of staphylococcal alpha-hemolysin, a heptameric transmembrane pore. *Science* **274,** 1859–1866 (1996).

89. Kasianowicz, J. J., Brandin, E., Branton, D. & Deamer, D. W. Characterization of individual polynucleotide molecules using a membrane channel. *Proc. Natl. Acad. Sci. U. S. A.* **93,** 13770–13773 (1996).

90. Akeson, M., Branton, D., Kasianowicz, J. J., Brandin, E. & Deamer, D. W. Microsecond time-scale discrimination among polycytidylic acid, polyadenylic acid, and polyuridylic acid as homopolymers or as segments within single RNA molecules. *Biophys. J.* **77,** 3227–3233 (1999).

91. Clarke, J. *et al.* Continuous base identification for single-molecule nanopore DNA sequencing. *Nat. Nanotechnol.* **4,** 265–270 (2009).

92. Lieberman, K. R. *et al.* Processive Replication of Single DNA Molecules in a Nanopore Catalyzed by phi29 DNA Polymerase. *J. Am. Chem. Soc.* **132,** 17961–17972 (2010).

93. Cherf, G. M. *et al.* Automated forward and reverse ratcheting of DNA in a nanopore at 5-angstrom precision. *Nat. Biotechnol.* **30,** 344–348 (2012).

94. Derrington, I. M. *et al.* Nanopore DNA sequencing with MspA. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 16060–16065 (2010).

95. Soskine, M. *et al.* An Engineered ClyA Nanopore Detects Folded Target Proteins by Selective External Association and Pore Entry. *Nano Lett.* **12,** 4895–4900 (2012).

96. Fahie, M., Chisholm, C. & Chen, M. Resolved Single-Molecule Detection of Individual Species within a Mixture of anti-Biotin Antibodies Using an Engineered Monomeric Nanopore. *ACS Nano* **9,** 1089–1098 (2015).

97. Fologea, D., Uplinger, J., Thomas, B., McNabb, D. S. & Li, J. Slowing DNA Translocation in a Solid-State Nanopore. *Nano Lett.* **5,** 1734–1737 (2005).

98. Kowalczyk, S. W., Wells, D. B., Aksimentiev, A. & Dekker, C. Slowing down DNA Translocation through a Nanopore in Lithium Chloride. *Nano Lett.* **12,** 1038–1044 (2012).

99. Feng, J. *et al.* Identification of single nucleotides in MoS2 nanopores. *Nat. Nanotechnol.* **10,** 1070–1076 (2015).

100. Ma, J. *et al.* Detection of short single-strand DNA homopolymers with ultrathin Si3N4 nanopores. *Phys. Rev. E* **92,** 022719 (2015).

101. Sawafta, F., Carlsen, A. T. & Hall, A. R. Membrane thickness dependence of nanopore formation with a focused helium ion beam. *Sensors* **14,** 8150–61 (2014).

102. Li, J. *et al.* Ion-beam sculpting at nanometre length scales. *Nature* **412,** 166–169 (2001).

103. Storm, A. J., Chen, J. H., Ling, X. S., Zandbergen, H. W. & Dekker, C. Fabrication of solid-state nanopores with single-nanometre precision. *Nat. Mater.* **2,** 537–540 (2003).

104. Kwok, H., Briggs, K. & Tabard-Cossa, V. Nanopore Fabrication by Controlled Dielectric Breakdown. *PLoS ONE* **9,** e92880 (2014).

105. Yang, J. *et al.* Rapid and precise scanning helium ion microscope milling of solid-state nanopores for biomolecule detection. *Nanotechnology* **22,** 285310 (2011).

106. Sawafta, F., Clancy, B., Carlsen, A. T., Huber, M. & Hall, A. R. Solid-state nanopores and nanopore arrays optimized for optical detection. *Nanoscale* **6,** 6991–6996 (2014).

107. Yanagi, I., Ishida, T., Fujisaki, K. & Takeda, K. Fabrication of 3-nm-thick Si_3N_4 membranes for solid-state nanopores using the poly-Si sacrificial layer process. *Sci. Rep.* **5,** srep14656 (2015).

108. Larkin, J. *et al.* Slow DNA Transport through Nanopores in Hafnium Oxide Membranes. *ACS Nano* **7,** 10121–10128 (2013).

109. Garaj, S. *et al.* Graphene as a subnanometre trans-electrode membrane. *Nature* **467,** 190–U73 (2010).

110. Merchant, C. A. *et al.* DNA Translocation through Graphene Nanopores. *Nano Lett.* **10,** 2915–2921 (2010).

111. Schneider, G. F. *et al.* DNA Translocation through Graphene Nanopores. *Nano Lett.* **10,** 3163–3167 (2010).

112. MinION Analysis and Reference Consortium: Phase 1 data release and analysis - F1000Research. Available at: https://f1000research.com/articles/4-1075/v1. (Accessed: 9th November 2017)

113. Nanopore sequencing and assembly of a human genome with ultra-long reads | bioRxiv. Available at: https://www.biorxiv.org/content/early/2017/04/20/128835. (Accessed: 9th November 2017)

114. Ashton, P. M. *et al.* MinION nanopore sequencing identifies the position and structure of a bacterial antibiotic resistance island. *Nat. Biotechnol.* **33,** 296–300 (2015).

115. Quick, J. *et al.* Real-time, portable genome sequencing for Ebola surveillance. *Nature* **530,** 228–232 (2016).

116. Faria, N. R. *et al.* Establishment and cryptic transmission of Zika virus in Brazil and the Americas. *Nature* **546,** 406–410 (2017).

117. Storm, A. J. *et al.* Fast DNA Translocation through a Solid-State Nanopore. *Nano Lett.* **5,** 1193–1197 (2005).

118. Wanunu, M. *et al.* Nanopore analysis of individual RNA/Antibiotic complexes. *Acs Nano* **5,** 9345–9353 (2011).

119. Henley, R. Y. *et al.* Electrophoretic Deformation of Individual Transfer RNA Molecules Reveals Their Identity. *Nano Lett.* **16,** 138–144 (2016).

120. Kowalczyk, S. W., Hall, A. R. & Dekker, C. Detection of Local Protein Structures along DNA Using Solid-State Nanopores. *Nano Lett.* **10,** 324–328 (2010).

121. Chen, K. *et al.* Ionic Current-Based Mapping of Short Sequence Motifs in Single DNA Molecules Using Solid-State Nanopores. *Nano Lett.* **17,** 5199–5205 (2017).

122. Hanz, G. M., Jung, B., Giesbertz, A., Juhasz, M. & Weinhold, E. Sequence-specific Labeling of Nucleic Acids and Proteins with Methyltransferases and Cofactor Analogues. *JoVE J. Vis. Exp.* e52014–e52014 (2014). doi:10.3791/52014

123. Wanunu, M. *et al.* Rapid electronic detection of probe-specific microRNAs using thin nanopore sensors. *Nat. Nanotechnol.* **5,** 807–814 (2010).

124. Wang, Y., Zheng, D., Tan, Q., Wang, M. X. & Gu, L.-Q. Nanopore-based detection of circulating microRNAs in lung cancer patients. *Nat. Nanotechnol.* **6,** 668–674 (2011).

125. Riedl, J., Ding, Y., Fleming, A. M. & Burrows, C. J. Identification of DNA lesions using a third base pair for amplification and nanopore sequencing. *Nat. Commun.* **6,** ncomms9807 (2015).

126. Shim, J. *et al.* Detection and Quantification of Methylation in DNA using Solid-State Nanopores. *Sci. Rep.* **3,** (2013).

127. Wanunu, M. *et al.* Discrimination of Methylcytosine from Hydroxymethylcytosine in DNA Molecules. *J. Am. Chem. Soc.* **133,** 486–492 (2011).

128. Carson, S., Wilson, J., Aksimentiev, A., Weigele, P. R. & Wanunu, M. Hydroxymethyluracil modifications enhance the flexibility and hydrophilicity of double-stranded DNA. *Nucleic Acids Res.* gkv1199 (2015). doi:10.1093/nar/gkv1199

129. Globisch, D. *et al.* Tissue Distribution of 5-Hydroxymethylcytosine and Search for Active Demethylation Intermediates. *PLoS ONE* **5,** e15367 (2010).

2. Chapter 2 - Interpreting The Conductance Blockades Of DNA Translocations Through Solid-State Nanopores

Solid-State nanopore electronic signatures can be challenging to interpret, even for pure homogenous samples (dsDNA of a particular length). Multiple conductance blockade levels are observed for an individual sample, furthermore the mean conductance levels of these populations vary with applied voltage. This phenomenon is pronounced when using ultra-thin nanopore devices (i.e. membrane thickness of <10nm). While several groups have observed this and probable reasons for these have been proposed, no conclusive evidence has been provided for the actual cause. This report focuses on interpretation of the multiple conductance blockade populations observed while measuring DNA across a range of applied voltages. We investigate a 3kbp dsDNA molecule using a device with a 3.4nm pore diameter fabricated in a 4.5nm thick membrane. Across a range of applied voltages, we observe multiple conductance blockade (a total of 3) levels. These populations vary until high levels of voltage are applied (>400mV) after which only two discrete populations remain with non-varying mean conductance levels. To describe these observations, we provide a simple geometric model that gives insight into the translocation dynamics of the analyte through the nanopore. Each, low, mid and high conductance level corresponds to; the initial interaction of the DNA with the electric field on the trans end, its presence inside the pore and non-translocating interactions of the DNA, respectively. The experimental results match well with our model and can be applied to all solid-state nanopores in general. These results are an important step towards understanding the translocation process and will give insight into which levels correspond to actual translocation of the analyte. Furthermore, it would allow us to

determine exactly which ionic current population correspond to a particular target analyte, especially in a complex mixture of samples.

Peer-reviewed Publication: '*Interpreting the conductance blockades of DNA translocations through solid-state nanopores', A. T. Carlsen, O. K. Zahid, J. A. Ruzicka, E. W. Taylor, A. R. Hall, ACS Nano, 8 (5), 4754–4760, 2014*

Interpreting The Conductance Blockades Of DNA Translocations Through Solid-State Nanopores

Abstract: Solid-state nanopore electrical signatures can be convoluted and are thus challenging to interpret. In order to better understand the origin of these conductance changes, we investigate the translocation of DNA over a range of voltage. We observe multiple, discrete populations of conductance blockades that vary with applied voltage. To describe our observations, we develop a simple model that is applicable to solid-state nanopores generally. These results represent an important step towards understanding of the dynamics of the electrokinetic translocation process.

Keywords: Nanopore, DNA, translocation, conductance blockade, single-molecule

Solid-state nanopores[1, 2] are an emerging technology for rapid detection and characterization of biomolecules. In a typical measurement, an electric field is employed to drive individual molecules of RNA,[3] proteins,[4-7] and most often DNA[8, 9] through a single aperture in a solid-state membrane (Fig. 1a). The brief presence of the molecule within the opening is manifested as a shift in the measured electrical signal to one or more distinct levels of conductance, referred to as a blockade event. By virtue of their size, SS-nanopores are able to interrogate one or a few individual molecules at a time, and so they have proven to be an attractive possibility for a range of applications that require highly sensitive detection, perhaps most notably genetic sequencing.[10]

Although the operating principle of SS-nanopores is straightforward, the system is capable of exhibiting surprisingly complex behaviors that can make interpretation of the

measured electrical signal challenging. One source of this complexity is thought to be interactions with the access regions[11]- the sensing volume immediately surrounding each opening of the aperture. Here, we seek to shed light on SS-nanopore measurements in general by investigating double-strand (ds-) DNA conductance blockades systematically using a nanopore device with maximized contributions of the access regions to the sensing region. The access regions have long been an important consideration[12] in describing both cylindrical[13-19] and non-cylindrical[20-23] nanopore systems. However, for SS-nanopores in extremely thin membranes[19] (<10 nm), these regions take on increased significance. We therefore initiate our study of conductance depth (ΔG) by investigating these unconventional devices.

Results and Discussion

We begin by measuring 3 kbp dsDNA under high-ionic strength conditions with a single SS-nanopore (diameter 3.4 nm) fabricated in a 4.5 nm thick membrane. Previous work on devices with similar dimensions and under comparable solvent conditions demonstrated that dsDNA translocations produce deeper blockade events[19, 24] compared to the typical 1-2 nS depth measured in conventional (thick) membranes.[3, 8, 9] Our measurements confirm this observation in general (Figure 2-1b). However, we arrive at a substantially more complex picture when we investigate the dependence of ΔG on applied voltage. Figure 2-1c shows all-points histograms for (concatenated) events over the range of 50-400 mV, with Gaussian fits (grey lines) indicating the locations of discrete conductance levels.

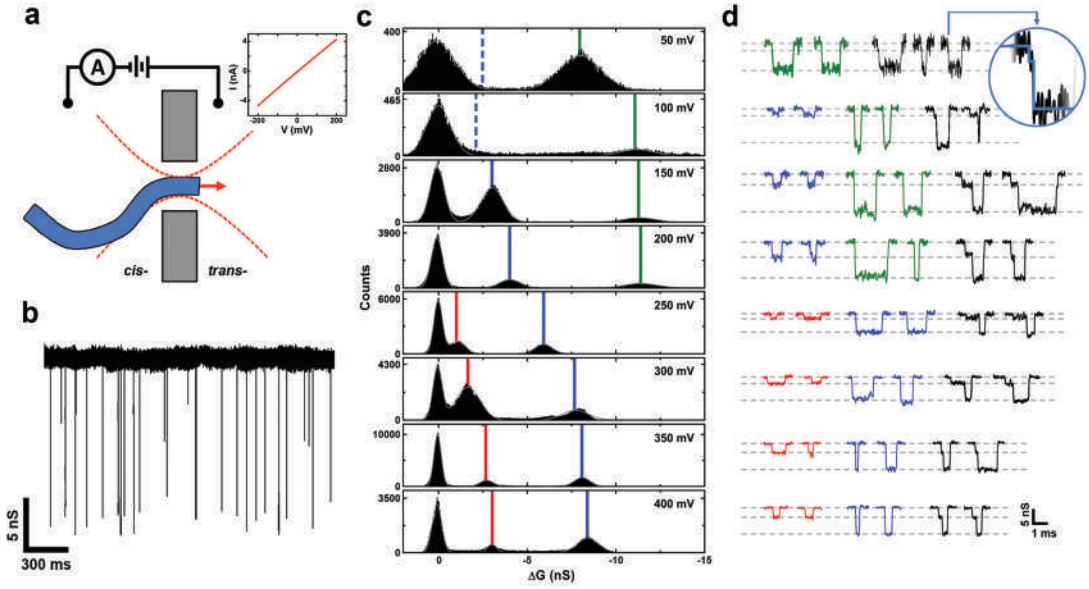

Figure 2-1 Conductance blockade measurements of dsDNA. (a) Schematic representation of electrokinetic translocation from the *cis*- side of a SS-nanopore membrane to the *trans*- side. Inset shows the linear I-V characteristics of the device used here. (b) Typical conductance trace measured for 3 kbp dsDNA using a 4.5 nm thick, 3.4 nm diameter SS-nanopore. Voltage is 400 mV. (c) All-points histograms of (concatenated) conductance blockades from 50 to 400 mV (low-pass filtered at 10 kHz). In each panel, the left-most peak corresponds to the baseline (open-pore) conductance. Vertical lines indicate the center of the Gaussian fit (gray line) and indicate the evolution of individual conductance populations designated by color. (d) Example event traces for each applied voltage in (c). Dashed lines in background designate the discrete populations from the histograms to the left. Trace colors indicate the conductance level population of the event from (c), except the black traces, which correspond to events containing more than one level. The inset offers a magnified view of the indicated combination event, highlighting the brief initial shallow level. Note that this level is not

resolvable in the 50 or 100 mV histograms (see text), but its position is indicated by dashed blue lines. All traces are low-pass filtered at 20 kHz. The scale bar applies to all traces.

From these data, we make two observations. First, we generally do not observe that the conductance blockade level resides exclusively at a single level, as may be expected for head-to-tail translocation of dsDNA through an aperture that is too narrow to allow the passage of folded molecules. Instead, we observe two, well-separated levels of conductance under most conditions. These levels are not mutually exclusive, however, as we note that combination events occur regularly (black traces in Figure 2-1d). Second, the ΔG levels themselves shift significantly as the applied voltage is increased. Examining the evolution of blockade levels as a function of voltage (Figure 2-1c) reveals that three distinct populations are detected, each of which appears to increase in depth as applied voltage is increased.

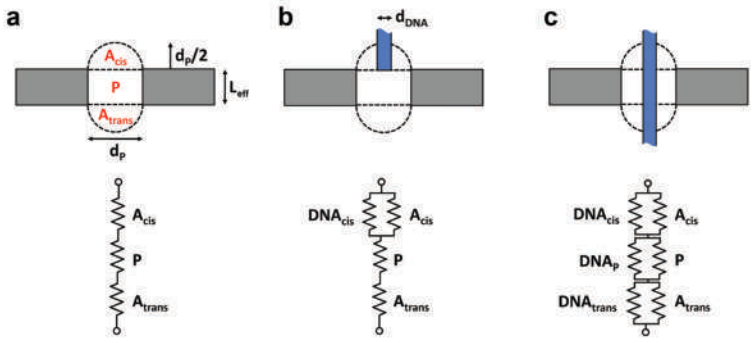

Figure 2-2 Components of the model. (a) Schematic of the three series conductances that form the sensing region of a SS-nanopore (gray): the pore (*P*) and the two hemispherical access regions (A_{cis} and A_{trans}, respectively). Models of DNA interaction with (b) one access region only (*case 1* from the text) and (c) all three regions (*case 2*

from the text). Below each diagram is an equivalent circuit (conductances shown as resistors) representing the contributions of the three sensing regions listed above as well as that of the presence of the DNA in A_{cis} (designated DNA_{cis}), P (designated DNA_P), and A_{trans} (designated DNA_{trans}). Not shown are contributions of counterions surrounding the DNA, which act as an additional parallel conductance in each of the three regions.

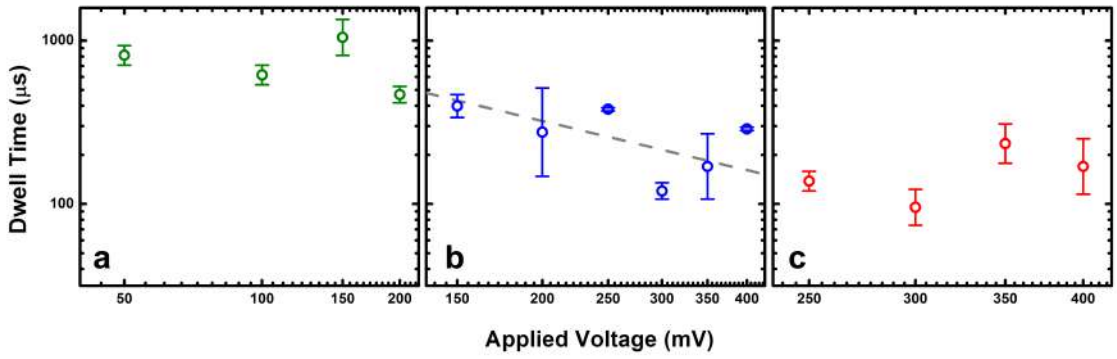

Figure 2-3 Event dwell times. Voltage dependence of mean dwell time for events observed in the three types of DNA-pore interactions described in the text: lateral blocking (a), translocation (b), and interaction with one access region (c). Only single-level events are considered. For measurements ≥250 mV, we frequently observe two populations in the translocation data, one of which yields a dwell time >1 ms. We attribute these long duration events to strong interactions between the DNA and the pore and/or the surrounding membrane, and so we plot only the shorter duration populations here. Data shown in (a) and (c) exhibit no voltage dependence[61] over the range studied, indicating that forces other than electrical (e.g. diffusive[26]) dominate. Dashed line in (b) is a 1/V fit to the data, similar to that reported elsewhere[27].

We address the first of these observations by hypothesizing that the plurality of conductance blockade levels are the result of non-translocative interactions with our device. This hypothesis is supported by the observed voltage dependence of mean event dwell time measurements (see Figure 2-3). Recent work from several groups has suggested that unexpected levels of ΔG measured under certain experimental conditions are a result of dsDNA entering the access region, either stochastically[25, 26] or as a precursor to translocation.[27-29] In order to describe the conductance blockades that may be expected from this type of interaction, we utilize a model in which the SS-nanopore sensing region is composed of three relevant sections- the interior of the nanopore itself and the two access regions on either side (Figure 2-2a), following past work.[14, 17] The conductance of each access region can be expressed simply[11] as

$$G_{0_{acc}} = 2\sigma d_p, \qquad \text{(Eq. 1)}$$

where d_p is nanopore diameter and σ is the conductivity of the solution, defined as $(\mu_{cation}+\mu_{anion})ne$. Here, n is the number density (proportional to concentration) of the ionic species, e is the elementary charge, and μ_{cation} and μ_{anion} are electrophoretic mobilities of the cation and anion, respectively. The conductance of the pore, meanwhile, is

$$G_{0_{pore}} = \frac{\pi d_p^2}{4L_{eff}}\left(\sigma + \frac{4S\mu_{cation}}{d_p}\right), \qquad \text{(Eq. 2)}$$

where L_{eff} is the effective thickness of the membrane and S is the surface charge density[30] of the nanopore walls (taken[31] as 0.06 C/m^2). We use the convention of Wanunu, *et al.*[19] who established experimentally that $L_{eff}=L/3$, where L is the initial membrane thickness,

to account for the non-cylindrical shape of the nanopore. Because these conductances are in series, the total open-pore conductance of the system can thus be written as

$$G_{0_{TOTAL}} = \left(\frac{1}{G_{0_{pore}}} + \frac{2}{G_{0_{acc}}} \right)^{-1} . \qquad \text{(Eq. 3)}$$

Generally, the presence of dsDNA in any of the sensing regions described above will act as a negative parallel conductance, displacing volume that would otherwise contribute to the total measured conductance. In order to quantify the effect, we consider two basic scenarios in relation to our data, taking into account that conductance can be expressed generally as $\sigma(A/l)$, where A is area and l is length.

In *case 1* (Figure 2-2b), the dsDNA is positioned coaxially with the mouth of the pore such that it interacts only with a single access region. Although the dsDNA could adopt a range of orientations with respect to this region, the geometry considered here can be considered a maximimum as it occupies the most space within the access region. In this scenario, the effect on the conductance of the occluded access region is

$$G_{acc_{DNA}} = G_{0_{acc}} - G_{DNA_{acc}} = G_{0_{acc}} - \sigma \frac{\pi d_{DNA}^2}{2 d_p} , \qquad \text{(Eq. 4)}$$

where d_{DNA} is the diameter of dsDNA, taken to be 2.2 nm. Note that in this case, the pertinent length of DNA, l, is the length of the access region ($d_p/2$). All other regions will be unchanged. As a result, the total change in conductance can be expressed as

$$\Delta G_{case1} = \left(\frac{1}{G_{0_{pore}}} + \frac{1}{G_{0_{acc}}} + \frac{1}{G_{acc_{DNA}}} \right)^{-1} - G_{0_{TOTAL}} . \qquad \text{(Eq. 5)}$$

In *case 2*, the dsDNA is present in all three regions of the system (Figure 2-2c). As such, both access regions are affected as described in Equation 4, and additionally, the conductance of the nanopore region is altered, as described by

$$G_{pore_{DNA}} = G_{0_{pore}} - G_{DNA_{pore}} = G_{0_{pore}} - \sigma\frac{\pi d_{DNA}^2}{4L_{eff}} \quad . \qquad \text{(Eq. 6)}$$

In total, this results in an expected conductance change for *case 2* of

$$\Delta G_{case2} = \left(\frac{1}{G_{pore_{DNA}}} + \frac{2}{G_{acc_{DNA}}}\right)^{-1} - G_{0_{TOTAL}} \, . \qquad \text{(Eq. 7)}$$

The intermediate case, in which the dsDNA resides only in the *cis*-side access region and the nanopore is transitional since passage to the *trans*-side access region (*i.e. case 2*) is almost immediate from this state. As such, we consider it unlikely to be observed.

Equations 5 and 7 can be applied to our data by incorporating the device dimensions (d_p=3.4 nm, L_{eff}=$L/3$=1.5 nm) and solvent conditions used in the experiment. Doing so yields for *case 1* a ΔG of -3.9 nS and for *case 2* a ΔG of -8.8 nS.

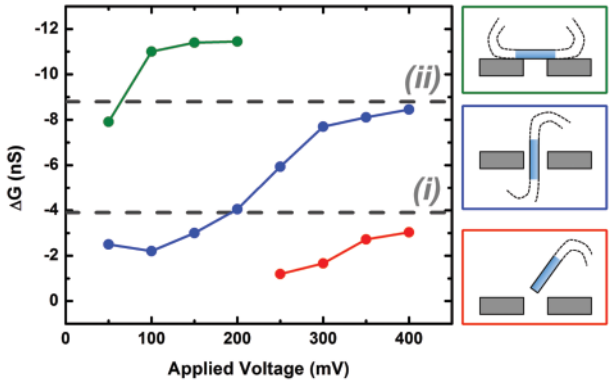

Figure 2-4 Analysis of dsDNA conductance blockades. (a) Mean conductance change *vs*. applied voltage. The dashed lines *(i)* and *(ii)* represent the calculated ΔG from Eqs. 5 (*case 1*, corresponding to non-translocative events) and 7 (*case 2*, corresponding to

translocation), respectively. Each point is the center of a Gaussian fit to the relevant histogram of all recorded events in Fig. 1c, except the blue points at 50 and 100 mV, which are Gaussian fit centers from an all-points histogram of events containing the lower (2-2.5 nS) level. Schematics to the right illustrate the DNA configuration we propose for each population. Colors match those in in Fig. 1c.

In Figure 2-4a, we plot the mean conductance blockade levels over all investigated voltages, showing that the measured ΔG of each population increases and then saturates. Strikingly, the conductance levels predicted by our simple model (dashed lines in Figure 2-4a) match very closely the saturation conductance observed for two of the event populations from the experimental data. This indicates that at high voltages (≥250 mV) for our device, the large ΔG level corresponds to true molecular translocations while the low ΔG level indicates non-translocative interactions with the access region. We note that this model can be used similarly to predict the apparent saturation conductance change for results on voltage dependence of ΔG published elsewhere[3, 32, 33] (see Appendix).

Our model accounts for two populations within our data, but what is the origin of the third? One possibility may be that this level of conductance blockade is caused by a molecule approaching the SS-nanopore such that it lays perpendicular to the axis of the aperture.[25, 27-29, 34] In this case, the stiffness of the molecule would prevent it from passing through the pore in a folded state, but the ion conductance would be blocked by its presence. The simplest approximation of this arrangement is that the area occupied by the dsDNA above the opening reduces the effective diameter of the SS-nanopore during its residence. If we assume a circular pore in the blocked case, this reduced effective pore

diameter, d^*_p, can be expressed geometrically (see Supplementary Information Fig. S-4) by the equation

$$d_p^* = \sqrt{\frac{2}{\pi}\left(d_p^2 \cos^{-1}\left(\frac{d_{DNA}}{d_p}\right) - d_{DNA}\left(d_p^2 - d_{DNA}^2\right)^{1/2}\right)}. \qquad \text{(Eq. 8)}$$

This assumes that the dsDNA sits directly across the center of the SS-nanopore. We arrive at an expected ΔG for the perpendicular case by simply calculating the difference between the conductance of the unoccluded pore with diameter d_p and the conductance of a pore with diameter d^*_p using Equation 3. For our experimental conditions, ΔG for lateral blocking of the pore is found to be -15 nS. Since this can be considered a maximum (*i.e.* the dsDNA may not interact symmetrically across the center of the pore), the value is in qualitative agreement with the maximum conductance blockade recorded for the uncategorized population in Figure 2-4a of about -12 nS. We therefore conclude that this population is likely to correspond to lateral, non-translocative interactions of the dsDNA with the SS-nanopore.

Identification of each conductance population presents further insight into the translocation process. As seen in Figure 2-1d, discrete conductance levels occur not only independently in single-level events, but also in combination to form two-level events. Interestingly, we observe that the shallow conductance level precedes the deep level for nearly all two-level events recorded across the entire investigated voltage range (543 out of 553, or 98.2%). At high voltages (≥250 mV), this ordering suggests an initial time period during which the end of the dsDNA is positioned in the access region of the SS-nanopore prior to threading through the aperture. The initial lag may correspond to repositioning or unfolding of the ensuing length of the molecule[35, 36] (see Appendix

Figure A2- 3). At low voltages (≤200 mV), the shallow-to-deep progression of conductance levels suggests instead that a portion of the dsDNA is threaded through the nanopore prior to lateral interactions between the remainder of the molecule and the aperture. We attribute these lateral interactions to the diminished capacity of the low-voltage electric field gradient to overcome the entropy of the dsDNA near the SS-nanopore.[26] Thus, the end of a threading molecule drags the entropic coil to the aperture *en bloc*, where it creates a deeper blockade as the translocation process continues. As voltage is reduced further, the likelihood of a molecular end being made available by the weaker electric field gradient is also reduced. Thus, at very low voltage, we would expect the shallow event level corresponding to translocation to be rare. This is indeed the case; the translocation ΔG level is uncommon at both 50 and 100 mV and, in conjunction with the low signal-to-noise ratio at these voltages, it is not distinguishable in an all-points histogram (see Figure 2-1c). However, the lower level can be resolved within individual blockade events (see Figure 2-1d, top panel). Besides serving as additional support for our interpretation, this observation also offers an explanation as to why the ΔG level corresponding to lateral interactions with the SS-nanopore is seen exclusively at low voltage: at voltages greater than 200 mV, the large electric field gradient and increased viscous drag act to uncoil the dsDNA in solution before it reaches the pore.[37] Note that the deep ΔG level does not necessarily preclude molecular translocation at low voltage. Indeed, the blockade caused by the entropic coil may be able to simply mask the signal of the simultaneous threading of dsDNA through the pore.

While our model explains a great deal of what we observe in experiment, one central question remains: why do the conductance blockade levels increase and then

saturate with applied voltage? Recently, several groups have reported similar behavior in conventional SS-nanopores,[3, 32, 33] but so far, an explanation has not been agreed upon. We suggest that the origin of this effect may be polarization of the dsDNA counterion cloud. The presence of positive charges surrounding the negatively-charged dsDNA backbone is known to counteract the conductance blockade by introducing additional carriers to the sensing region.[30] However, theoretical work by Mendel[38] and later refinement by Manning[39] and others[40, 41] has predicted that the local density of these counterions can be perturbed under extreme electric fields. Counterion polarization has since been observed through simulation[42-44] and experiment[45-47] and has recently been suggested as a potential factor in SS-nanopore measurements as well.[48, 49] Perturbation of the counterion cloud could remove charge carriers locally from the sensing region of a nanopore, resulting in a voltage-dependent conductance blockade. Note that this local perturbation does not contradict overall electroneutrality as has been observed in molecular dynamics simulations.[50, 51] Rather, the reduction in counterion density local to the sensing region of the nanopore would be accompanied by an equivalent buildup of charge outside the sensing region, as shown schematically in Figure 2-5 and in more detail in Appendix Figure A2- 4. In addition, because polarization will saturate at very high electric field strength,[39] the voltage-dependence would likewise saturate at high voltage. These two key expectations match our experimental results well.

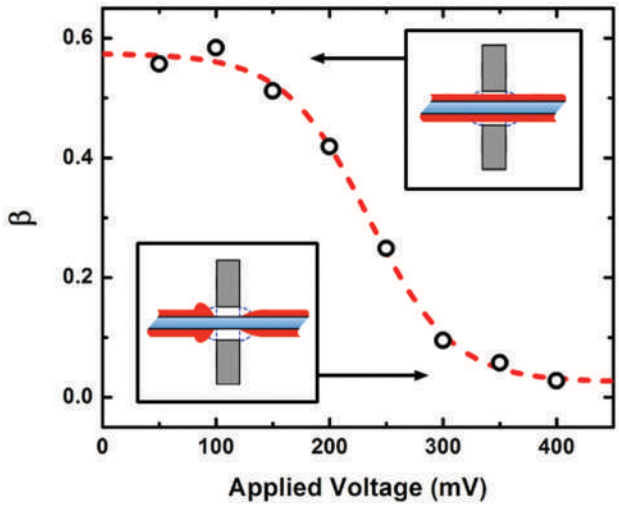

Figure 2-5 Counterion residence on translocating dsDNA. The fraction of Na counterion charge (relative to the charge density of the dsDNA backbone) β *vs*. applied voltage. Dashed line is a Boltzmann sigmoid fit to the data. Insets show schematic interpretation of the low-voltage case (top), where counterions remain bound to the DNA, and the high-voltage case (bottom), where counterions are displaced locally during translocation. In both images, DNA motion is towards the right. A more detailed schematic treatment is provided in Supplementary information Fig. S-6.

Counterion screening and polarization can be included in our model through a simple adjustment to Equations 4 and 6, as follows:

$$G_{acc_{DNA}} = G_{0_{acc}} - G_{DNA_{acc}} + G_{counter_{acc}} = G_{0_{acc}} - G_{DNA_{acc}} + \beta\frac{2q\mu_{cation}}{d_p}; \qquad \text{(Eq. 9)}$$

$$G_{pore_{DNA}} = G_{0\,pore} - G_{DNA_{pore}} + G_{counter_{pore}} = G_{0\,pore} - G_{DNA_{pore}} + \beta\frac{q\mu_{cation}}{L_{eff}}, \qquad \text{(Eq. 10)}$$

where q is the charge per unit length of dsDNA. We introduce β to denote the fractional effect of the new term relative to the zero-field counterion density, indicating polarization-induced depletion in the sensing region of the nanopore. When β=1, the charge of the counterion cloud is exactly equivalent to that of the dsDNA itself. When β=0, no counterions are present on the dsDNA in the sensing region and volumetric blocking is the only contribution to the conductance change (an alternative conceptualization is that β is indicative of a voltage-dependent shift in counterion mobility μ_{cation} rather than total counterion residence in the sensing region, though it is unclear by what mechanism this effect might saturate). So far, our model has assumed intrinsically that β=0, resulting in good agreement with measurements at high voltage where polarization fully depletes of counterions in the pore. The transition towards this state can be analyzed further by using Eqs. 9 and 10 to determine the β necessary to account for the measured voltage dependence of the ΔG. Such an analysis (Figure 2-5) suggests that the fractional residence of counterions around the dsDNA in the sensing region of the SS-nanopore decreases with voltage in a sigmoidal fashion. This data can be extrapolated to yield the zero-field value, which is β=0.57 for our system. This value may be indicative of the fraction of counterions relative to the total dsDNA charge in the sensing region that are bound tightly to the molecule (resident in the major and minor grooves,[52] for example) under our solvent conditions (Figure A2- 5). We note that β would represent an axial average along the length of the sensing region due to the axial inhomogeneity of the electric field in the nanopore. This approach may offer a general route towards probing the screening of polymers and biopolymers by arbitrary ionic species at various concentrations.

Conclusions

In conclusion, we have investigated dsDNA conductance blockades with a small diameter (3.4 nm) SS-nanopore formed in a thin (4.5 nm) membrane. We measured across a range of applied voltage and found that (*i*) three discrete levels of conductance change can be observed, and (*ii*) the ΔG associated with these levels becomes larger as the voltage is increased. We presented a simple model that takes into account the access regions of the SS-nanopore device and considers both the volume of the dsDNA and its accompanying countercharge layer as parallel conductances to that of the nanopore itself. We found that this model describes accurately the conductance blockade levels measured experimentally and additionally provides a possible explanation for the observed voltage dependence of ΔG. We proposed that the intensifying electric field that accompanies increasing voltage progressively removes the counterions surrounding the dsDNA in solution until eventually ion exclusion is the only contribution to the measured conductance change. Our results are widely applicable to a variety of experimental conditions and represent an important step toward understanding the meaning of SS-nanopore electrical signals in general. Indeed, our model can also be used to describe the observations of several previous studies using a variety of experimental conditions (Appendix). We note that while qualitative agreement is observed in all ionic conditions, quantitative agreement between our model and experimental work is currently limited to high-ionic strength solutions (Figure A2- 6). Extension to the low-ionic strength regime should be possible with refinement.

In conventional systems, the multiple conductance blockade populations described here will be subtle. For example, under typical solvent conditions (1 M KCl) for a SS-nanopore device with diameter and membrane thickness values of 20 nm each, the ΔG predicted for non-translocative interactions with the access region would be expected to have a maximum value of only -0.7 nS in the high-voltage (β=0) limit. Event duration would also be expected to be very brief under these high voltage conditions, and so as a result, reduction of the noise floor to a point where such events would be measurable is likely to filter them out entirely. This may explain why the effect has not been described previously. As device diameter or membrane thickness is reduced, however, the influence of the access regions will become more conspicuous. For this reason, the current trend of the field towards SS-nanopore devices with small diameters[53] or low dimensionality[54-58] will be especially aided by consideration of our findings in order to assess results accurately.

Materials and methods

Solid-state nanopore fabrication Silicon chips, each supporting a window of silicon nitride (24.5 nm thick as measured by ellipsometry), were purchased commercially (Norcada, Inc., Alberta, Canada). Thin SS-nanopores were produced with Helium ion microscope fabrication by first reducing the local membrane thickness controllably[59] and then milling material from the center of the thinned region using a timed exposure.[33] The region around the pore was processed to have a final thickness of 4.5 ± 0.6 nm, as judged by two separate calibration strategies.[59, 60] The precise diameter of the SS-nanopore was determined by applying the measured current-voltage characteristics of the device to

Equation 3 from the text and solving for d_p. The device exhibited a linear I-V curve and had a low-noise baseline conductance of 27.5 nS that varied less than 5% during the duration of the measurements.

DNA translocation measurements Solvent conditions used for the presented measurements were 900 mM NaCl, 10 mM tris, 1 mM EDTA. 3 kbp dsDNA was introduced to the *cis* side of the SS-nanopore at a concentration of ~10 ng/μL. Conductance blockade events were recorded at a bandwidth of 200 kHz and with a 100 kHz four-pole Bessel filter applied. An additional low-pass filter of 10-20 kHz was applied during analysis (as indicated in figure captions), which was performed using custom LabView software. Histograms from Fig. 1c (and scatter plots in Supplementary Information Fig. S-1) are composed of n = 96 (50 mV); 146 (100 mV); 426 (150 mV); 298 (200 mV); 437 (250 mV); 814 (300 mV); 542 (350 mV); and 616 (400 mV). Only events with durations between 100 μs and 2 ms were considered in our analysis.

Acknowledgements

ARH acknowledges startup funds from the Wake Forest University School of Medicine. EWT and JR acknowledge funding from The Dr. Arthur and Bonnie Ennis Foundation, Decatur, IL. We thank U. F. Keyser for critical reading of the manuscript and M. Wanunu for supplying data for analysis in Fig. S-5.

Supporting Information Available: Scatter plots of all measured blockade events, dwell time analysis, additional figures and text describing the application of the presented

model to published results, and derivation of lateral nanopore blocking expression (Eq. 8). This material is available free of charge *via* the Internet at http://pubs.acs.org.

References

1. Dekker, C. Solid-State Nanopores. *Nat. Nanotechnol.* 2007, 2, 209-215.
2. Wanunu, M. Nanopores: A Journey Towards DNA Sequencing. *Phys. Life Rev.* 2012, 9, 125-158.
3. Skinner, G. M.; van den Hout, M.; Broekmans, O.; Dekker, C.; Dekker, N. H. Distinguishing Single- and Double-Stranded Nucleic Acid Molecules Using Solid-State Nanopores. *Nano Lett.* 2009, 9, 2953-2960.
4. Firnkes, M.; Pedone, D.; Knezevic, J.; Doeblinger, M.; Rant, U. Electrically Facilitated Translocations of Proteins Through Silicon Nitride Nanopores: Conjoint and Competitive Action of Diffusion, Electrophoresis, and Electroosmosis. *Nano Lett.* 2010, 10, 2162-2167.
5. Fologea, D.; Ledden, B.; McNabb, D. S.; Li, J. Electrical Characterization of Protein Molecules By a Solid-State Nanopore. *App. Phys. Lett.* 2007, 91.
6. Han, A. P.; Schurmann, G.; Mondin, G.; Bitterli, R. A.; Hegelbach, N. G.; de Rooij, N. F.; Staufer, U. Sensing Protein Molecules Using Nanofabricated Pores. *App. Phys. Lett.* 2006, 88.
7. Plesa, C.; Kowalczyk, S. W.; Zinsmeester, R.; Grosberg, A. Y.; Rabin, Y.; Dekker, C. Fast Translocation of Proteins Through Solid-State Nanopores. *Nano Lett.* 2013, 13, 658-663.

8. Li, J. L.; Gershow, M.; Stein, D.; Brandin, E.; Golovchenko, J. A. DNA Molecules and Configurations in a Solid-State Nanopore Microscope. *Nat. Mater.* 2003, 2, 611-615.

9. Storm, A. J.; Chen, J. H.; Zandbergen, H. W.; Dekker, C. Translocation of Double-Strand DNA through a Silicon Oxide Nanopore. *Phys. Rev. E* 2005, 71.

10. Branton, D.; Deamer, D. W.; Marziali, A.; Bayley, H.; Benner, S. A.; Butler, T.; Di Ventra, M.; Garaj, S.; Hibbs, A.; Huang, X., *et al*. The Potential and Challenges of Nanopore Sequencing. *Nat. Biotechnol.* 2008, 26, 1146-1153.

11. Hall, J. E. Access Resistance of a Small Circular Pore. *J. Gen. Physiol.* 1975, 66, 531-532.

12. Willmott, G. R.; Smith, B. G. Comment on 'Modeling the Conductance and DNA Blockade of Solid-State Nanopores'. *Nanotechnology* 2012, 23, 088001.

13. Bacri, L.; Oukhaled, A. G.; Schiedt, B.; Patriarche, G.; Bourhis, E.; Gierak, J.; Pelta, J.; Auvray, L. Dynamics of Colloids in Single Solid-State Nanopores. *J. Phys. Chem. B* 2011, 115, 2890-2898.

14. Hyun, C.; Rollings, R.; Li, J. Probing Access Resistance of Solid-State Nanopores With a Scanning-Probe Microscope Tip. *Small* 2012, 8, 385-392.

15. Kowalczyk, S. W.; Grosberg, A. Y.; Rabin, Y.; Dekker, C. Modeling the Conductance and DNA Blockade of Solid-State Nanopores. *Nanotechnology* 2011, 22, 315101.

16. Sun, L.; Crooks, R. M. Single Carbon Nanotube Membranes: A Well-Defined Model for Studying Mass Transport Through Nanoporous Materials. *J. Am. Chem. Soc.* 2000, 122, 12340-12345.

17. Tsutsui, M.; Hongo, S.; He, Y.; Taniguchi, M.; Gemma, N.; Kawai, T. Single-Nanoparticle Detection Using a Low-Aspect-Ratio Pore. *ACS Nano* 2012, 6, 3499-3505.

18. Wang, J.; Ma, J.; Ni, Z.; Zhang, L.; Hu, G. Effects of Access Resistance on the Resistive-Pulse Caused by Translocating of a Nanoparticle Through a Nanopore. *RSC Adv.* 2014, 4, 7601-7610.

19. Wanunu, M.; Dadosh, T.; Ray, V.; Jin, J.; McReynolds, L.; Drndic, M. Rapid Electronic Detection of Probe-Specific MicroRNAs Using Thin Nanopore Sensors. *Nat. Nanotechnol.* 2010, 5, 807-814.

20. Heins, E. A.; Siwy, Z. S.; Baker, L. A.; Martin, C. R. Detecting Single Porphyrin Molecules in a Conically Shaped Synthetic Nanopore. *Nano Lett.* 2005, 5, 1824-1829.

21. Vogel, R.; Willmott, G.; Kozak, D.; Roberts, G. S.; Anderson, W.; Groenewegen, L.; Glossop, B.; Barnett, A.; Turner, A.; Trau, M. Quantitative Sizing of Nano/Microparticles With a Tunable Elastomeric Pore Sensor. *Anal. Chem.* 2011, 83, 3499-3506.

22. Willmott, G. R.; Vogel, R.; Yu, S. S. C.; Groenewegen, L. G.; Roberts, G. S.; Kozak, D.; Anderson, W.; Trau, M. Use of Tunable Nanopore Blockade Rates to Investigate Colloidal Dispersions. *J.Phys.: Condens. Matter* 2010, 22.

23. Stober, G.; Steinbock, L. J.; Keyser, U. F. Modeling of Colloidal Transport in Capillaries. *J. Appl. Phys.* 2009, 105.

24. Venta, K.; Shemer, G.; Puster, M.; Rodriguez-Manzo, J. A.; Balan, A.; Rosenstein, J. K.; Shepard, K.; Drndic, M. Differentiation of Short, Single-

Stranded DNA Homopolymers in Solid-State Nanopores. *ACS Nano* 2013, 7, 4629-4636.

25. van den Hout, M.; Krudde, V.; Janssen, X. J. A.; Dekker, N. H. Distinguishable Populations Report on the Interactions of Single DNA Molecules With Solid-State Nanopores. *Biophys. J.* 2010, 99, 3840-3848.

26. Wanunu, M.; Sutin, J.; McNally, B.; Chow, A.; Meller, A. DNA Translocation Governed by Interactions With Solid-State Nanopores. *Biophys. J.* 2008, 95, 4716-4725.

27. Kowalczyk, S. W.; Dekker, C. Measurement of the Docking Time of a DNA Molecule onto a Solid-State Nanopore. *Nano Lett.* 2012, 12, 4159-4163.

28. Rosenstein, J. K.; Wanunu, M.; Merchant, C. A.; Drndic, M.; Shepard, K. L. Integrated Nanopore Sensing Platform With Sub-Microsecond Temporal Resolution. *Nat. Meth.* 2012, 9, 487-U112.

29. Vlassarev, D. M.; Golovchenko, J. A. Trapping DNA Near a Solid-State Nanopore. *Biophys. J.* 2012, 103, 352-356.

30. Smeets, R. M. M.; Keyser, U. F.; Krapf, D.; Wu, M. Y.; Dekker, N. H.; Dekker, C. Salt Dependence of Ion Transport and DNA Translocation Through Solid-State Nanopores. *Nano Lett.* 2006, 6, 89-95.

31. Stein, D.; Kruithof, M.; Dekker, C. Surface-Charge-Governed Ion Transport in Nanofluidic Channels. *Phys. Rev. Lett.* 2004, 93.

32. Kowalczyk, S. W.; Dekker, C. Salt and Voltage Dependence of the Conductance Blockade Induced by Translocation of DNA and RecA Filaments Through Solid-

State Nanopores. In *Nanopores for Bioanalytical Applications*, Edel, J.; Albrecht, T., Eds. Royal Society of Chemistry: 2012.

33. Yang, J.; Ferranti, D. C.; Stern, L. A.; Sanford, C. A.; Huang, J.; Ren, Z.; Qin, L.-C.; Hall, A. R. Rapid and Precise Scanning Helium Ion Microscope Milling of Solid-State Nanopores for Biomolecule Detection. *Nanotechnology* 2011, 22, 285310.

34. Aksimentiev, A.; Heng, J. B.; Timp, G.; Schulten, K. Microscopic Kinetics of DNA Translocation Through Synthetic Nanopores. *Biophys. J.* 2004, 87, 2086-2097.

35. Lu, B.; Albertorio, F.; Hoogerheide, D. P.; Golovchenko, J. A. Origins and Consequences of Velocity Fluctuations During DNA Passage Through a Nanopore. *Biophys. J.* 2011, 101, 70-79.

36. Storm, A. J.; Storm, C.; Chen, J. H.; Zandbergen, H.; Joanny, J. F.; Dekker, C. Fast DNA Translocation through a Solid-State Nanopore. *Nano Lett.* 2005, 5, 1193-1197.

37. Chen, P.; Gu, J. J.; Brandin, E.; Kim, Y. R.; Wang, Q.; Branton, D. Probing Single DNA Molecule Transport Using Fabricated Nanopores. *Nano Lett.* 2004, 4, 2293-2298.

38. Mandel, M. The Electrical Polarization of Rod-Like Charged Macromolecules. *Mol. Phys.* 1961, 4, 489-469.

39. Manning, G. S. A Condensed Counterion Theory for Polarization of Polyelectrolyte Solutions in High Fields. *J.Chem. Phys.* 1993, 99, 477-486.

40. Mohanty, U.; Zhao, Y. Q. Polarization of Counterions in Polyelectrolytes. *Biopolymers* 1996, 38, 377-388.

41. Rau, D. C.; Charney, E. Polarization of the Ion Atmosphere of a Charged Cylinder. *Biophys. Chem.* 1981, 14, 1-9.

42. Elvingson, C. Computer Simlations of the Structure of DNA Molecules in an Electric Field. *Biophys. Chem.* 1992, 43, 9-19.

43. Netz, R. R. Polyelectrolytes in Electric Fields. *J. Phys. Chem. B* 2003, 107, 8208-8217.

44. Yoshida, M.; Kikuchi, K.; Maekawa, T.; Watanabe, H. Electric Polarization of Rodlike Polyions Investigated by Monte-Carlo Simulations. *J. Phys. Chem.* 1992, 96, 2365-2371.

45. Diekmann, S.; Hillen, W.; Jung, M.; Wells, R. D.; Porschke, D. Electric Properties and Structure of DNA Restriction Fragments From Measurements of the Electrical Dichroism. *Biophys. Chem.* 1982, 15, 157-167.

46. Diekmann, S.; Jung, M.; Teubner, M. On the Orientation Function of the Electric Dichroism of DNA. *J. Chem. Phys.* 1984, 80, 1259-1262.

47. Porschke, D. The Mechanism of Ion Polarization Along DNA Double Helices. *Biophys. Chem.* 1985, 22, 237-247.

48. Chang, H.; Venkatesan, B. M.; Iqbal, S. M.; Andreadakis, G.; Kosari, F.; Vasmatzis, G.; Peroulis, D.; Bashir, R. DNA Counterion Current and Saturation Examined by a MEMS-Based Solid State Nanopore Sensor. *Biomed. Microdevices* 2006, 8, 263-269.

49. Das, S.; Dubsky, P.; van den Berg, A.; Eijkel, J. C. T. Concentration Polarization in Translocation of DNA Through Nanopores and Nanochannels. *Phys. Rev. Lett.* 2012, 108.

50. Comer, J.; Dimitrov, V.; Zhao, Q.; Timp, G.; Aksimentiev, A. Microscopic Mechanics of Hairpin DNA Translocation Through Synthetic Nanopores. *Biophys. J.* 2009, 96, 593-608.

51. Luan, B.; Aksimentiev, A. Electro-Osmotic Screening of the DNA Charge in a Nanopore. *Phys. Rev. E* 2008, 78.

52. Yoo, J.; Aksimentiev, A. Competitive Binding of Cations to Duplex DNA Revealed Through Molecular Dynamics Simulations. *J. Phys. Chem. B* 2012, 116, 12946-12954.

53. Singer, A.; Rapireddy, S.; Ly, D. H.; Meller, A. Electronic Barcoding of a Viral Gene at the Single-Molecule Level. *Nano Lett.* 2012, 12, 1722-1728.

54. Garaj, S.; Hubbard, W.; Reina, A.; Kong, J.; Branton, D.; Golovchenko, J. A. Graphene as a Subnanometre Trans-Electrode Membrane. *Nature* 2010, 467, 190-U73.

55. Garaj, S.; Liu, S.; Golovchenko, J. A.; Branton, D. Molecule-Hugging Graphene Nanopores. *Proc. Natl. Acad. Sci. U. S. A.* 2013, 110, 12192-12196.

56. Merchant, C. A.; Healy, K.; Wanunu, M.; Ray, V.; Peterman, N.; Bartel, J.; Fischbein, M. D.; Venta, K.; Luo, Z. T.; Johnson, A. T. C.; *et al*. DNA Translocation Through Graphene Nanopores. *Nano Lett.* 2010, 10, 2915-2921.

57. Schneider, G. F.; Kowalczyk, S. W.; Calado, V. E.; Pandraud, G.; Zandbergen, H. W.; Vandersypen, L. M. K.; Dekker, C. DNA Translocation Through Graphene Nanopores. *Nano Lett.* 2010, 10, 3163-3167.

58. Traversi, F.; Raillon, C.; Benameur, S. M.; Liu, K.; Khlybov, S.; Tosun, M.; Krasnozhon, D.; Kris, A.; Radenovic, A. Detecting the Translocation of DNA Through a Nanopore Using Graphene Nanoribbons. *Nat. Nanotechnol.* 2013, 8, 939-945.

59. Marshall, M. M.; Yang, J.; Hall, A. R. Direct and Transmission Milling of Suspended Silicon Nitride Membranes With a Focused Helium Ion Beam. *Scanning* 2012, 34, 101-106.

60. Hall, A. R. *In Situ* Thickness Assessment During Ion Milling of a Free-Standing Membrane Using Transmission Helium Ion Microscopy. *Microsc. Microanal.* 2013, 19, 740-744.

61. Kasianowicz, J. J.; Brandin, E.; Branton, D.; Deamer, D. W. Characterization of Individual Polynucleotide Molecules Using a Membrane Channel. *Proc. Natl. Acad. Sci. U. S. A.* 1996, 93, 13770-13773.

3. Chapter 3 – Selective Detection and Quantification of Modified DNA with Solid-State Nanopores

Nanopore detection has been an extremely powerful molecular characterize technique. Its been utilized to detect DNA, RNA, proteins, determine protein-nucleic acid interactions and employed for a variety of applications including discerning epigenetic marks. Ultra-thin SS-Nanopores have previously been used to distinguish DNA molecules containing high-density 5mC and 5hmC epigenetic marks and more recently 5hmU. However, the ionic current profile generated by these molecules results in significant overlap of current signature with subtle differences that are difficult to resolve. We sought to develop a novel single-molecule nanopore assay that overcomes this problem by employing MS protein ligands that selectively bind to biotin tags on target dsDNA molecules. We anticipated binding of MS to target biotinylated DNA molecules would generate a unique ionic current signature due to the significantly larger structure of the nucleoprotein complex. To our surprise we observed events exclusively for the DNA-MS complexes whereas no events were observed for DNA or protein molecules probed individually. Interestingly, the capture rate increased gradually as MS was titrated against dsBioDNA, up to an equimolar ratio, after which the rate remained constant even in the presence of excess MS. Therefore, events are exclusively observed for bioDNA-MS complexes, presence of background unbound bioDNA, nonbioDNA or excess MS do not contribute to the events. These results thus conclusively demonstrate the basis of a unique binary nanopore detection assay that generates an output signal exclusively when biotinylated dsDNA target is present in any complex sample mixture. We further

demonstrated the ability to quantify amounts of biotinylated DNA present in two different samples at varying concentrations by comparing the capture rate of the unknown samples to a concentration standard. This novel binary detection assay forms the foundation of the following chapters, wherein it has been employed for a variety of applications, specifically selective detection of cancer biomarkers, including microRNAs and 5hmC content.

Peer-reviewed Publication: '*Selective detection and quantification of modified DNA with solid-state nanopores', A.T. Carlsen, O. K. Zahid, J. A. Ruzicka, E. W. Taylor, A. R. Hall, Nano Letters, 14(10):5488-92, 2014*

Selective detection and quantification of modified DNA with solid-state nanopores

ABSTRACT: We demonstrate a solid-state nanopore assay for the unambiguous discrimination and quantification of modified DNA. Individual streptavidin proteins are employed as high-affinity tags for DNA containing a single biotin moiety. We establish that the rate of translocation events corresponds directly to relative concentration of protein-DNA complexes and use the selectivity of our approach to quantify modified oligonucleotides from among a background of unmodified DNA in solution.

KEYWORDS: solid-state nanopores, proteins, DNA, DNA modifications, single-molecule, quantification

Immunoprecipitation and pull-down assays are workhorses in biochemistry. With the ability to discriminate specific substrates in heterogeneous mixtures, they play important roles in a wide range of fields, including proteomics[1-3], epigenomics[4-7], and transcriptomics[8, 9]. However, despite their broad utility, these well-established strategies have limitations. Besides requiring large sample sizes, they are labor-intensive and are not inherently quantitative, typically requiring subsequent PCR[10] or enrichment[11] for downstream analysis. For these reasons, quantitative technologies with single-molecule sensitivity may offer important advantages[5].

Hinging on the Coulter principle, solid-state (SS-) nanopores[12, 13] have been employed extensively for molecular resistive-pulse sensing. In these experiments, the temporary presence of a single molecule passing through a narrow pore alters the coincident flow of ions, thereby creating a unique electronic signature in the measured ionic current. Analysis of these signatures (or events) offers insight into molecular structure[14, 15], surface

charge[16, 17], and pore-molecule interactions[18, 19]. With potential applications ranging from DNA sequencing[20] to biomarker identification[21], this powerful detection platform has been used to probe a diverse set of materials, including nucleic acids[22-24], proteins[25-27], and non-biological nanoparticles[28-30].

Recently, several groups have reported progress towards SS-nanopore differentiation of structural[31-33] or chemical[34, 35] variations in nucleic acid molecules. However, such studies often rely on subtle shifts in event characteristics that can introduce uncertainty. In this paper, we present an assay that enables unambiguous identification of double-strand (ds-) DNA modified with a single biotin moiety. Using a high-affinity protein tag (monovalent streptavidin, MS) to label modified oligonucleotides, we determine that translocation event rate correlates with protein-DNA complex formation. We speculate on the mechanism underlying this finding and then exploit its specificity to quantify modified DNA directly in solution with unmodified oligonucleotides.

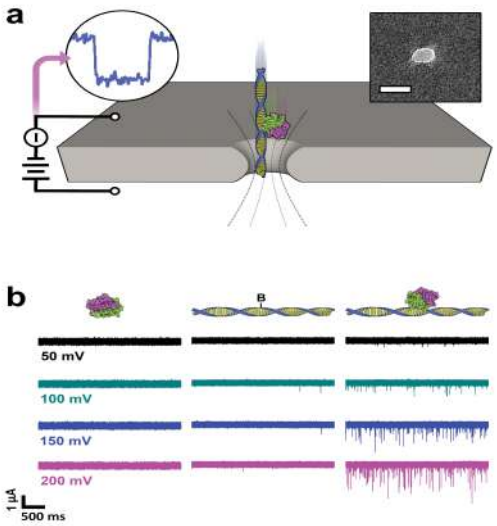

Figure 3-1 SS-Nanopore measurement. (a) Schematic diagram of the experimental system. A bias is applied across a SS-nanopore in a SiN membrane, inducing

electrokinetic translocation of protein-DNA complexes in electrolyte solution. Upper left shows the shape of a resistive pulse measured during a translocation event. Right inset: transmission electron micrograph of a typical SS-nanopore formed with conditions identical to those used here. Scale bar is 10 nm. (b) Raw current traces obtained for MS alone (left), bio90 alone (center), and MS incubated with bio90 at a molar ratio (MS:bio90) of 8:1 (right) over a range of voltages. Investigated molecules are shown schematically above each column ('B' represents the biotin moiety). Only the co-incubated material produces significant events.

SS-nanopore discrimination of monobiotinylated dsDNA

Figure 3-1a shows a schematic of the SS-nanopore measurement approach utilized here. An electrical bias is applied across a thin-film membrane with a single nanopore immersed in electrolyte solution (see Methods). This facilitates the electrokinetic translocation of molecules (or molecular complexes) through the pore, each of which can produce an ionic current event. We first use this technique to measure MS (

Figure 3-1b left) and monobiotinylated 90 bp dsDNA (bio90,

Figure 3-1b center) individually at concentrations of 8 μM and 1 μM, respectively. Over a range of 50-200 mV, few events can be identified for either molecule. However, when MS and bio90 are incubated together at a molar ratio of 8:1 (MS:bio90) prior to measurements, we observe a remarkable increase in the number of events per unit time (Figure 3-1b right). The event rate of the admixture is consistently more than an order of magnitude greater than that of either constituent molecule alone; at 200 mV applied voltage, for example, the MS-bio90 complex yields a rate of 23.3 ± 0.9 s-1, while the event rates of MS and bio90 individually are 0.09 ± 0.04 s-1 and 1.1 ± 0.2 s-1,

respectively. In order to verify that the MS-bio90 events correspond to actual translocations rather than stochastic interactions between the complex and the nanopore, we reverse the polarity of the applied voltage during our measurement and observe "recapture events"[36, 37] (see Figure 3-2). Translocations are further supported by the voltage dependence[38] observed for MS-bio90 event durations (See Appendix Figure A3-2).

What accounts for the extraordinary differences in the event rates that we observe? Although neither the MS nor bio90 alone yields a significant number of events over the investigated voltage range, we suggest differing explanations for each.

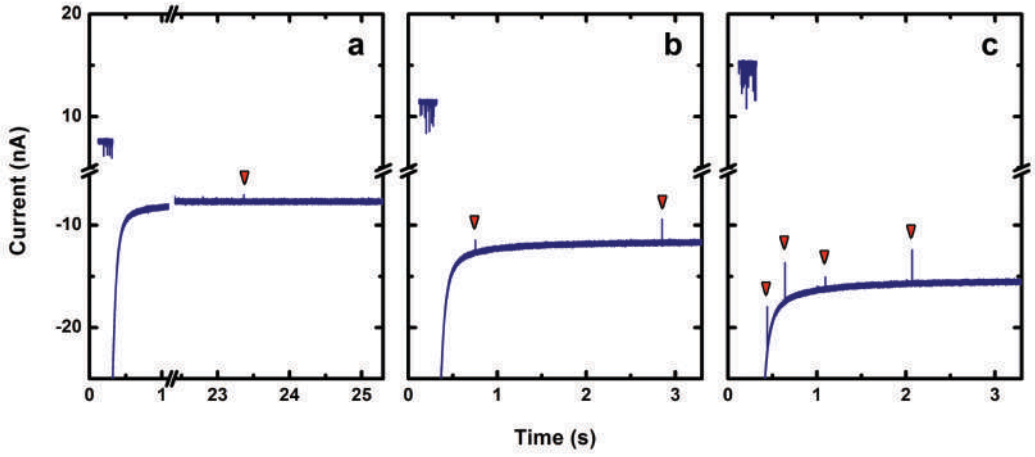

Figure 3-2 Recapture of translocated MS-bio90 constructs. Raw current traces showing translocation events of MS-bio90 at positive bias (initial 0.25 s of each trace) followed by subsequent re-capture of translocated material at negative bias. Voltages applied are ±200 mV (a), ±300 mV (b), and ±400 mv (c). Red arrows indicate recapture events. As voltage is increased, we observe a higher efficiency of re-capture. These data demonstrate that the MS-bio90 events measured in our experiments correspond to constructs passing through the SS-nanopore.

For the MS, recent work[26] comparing experimental data to the Smoluchowski theory[39] suggests that the translocation of proteins through SS-nanopores typically occurs too rapidly to be resolved by conventional systems like the one employed here. Indeed, Larkin, *et al*.[27] demonstrated that the enhanced time-resolution of a high-bandwidth system is capable of resolving far more protein translocation events than are detectable by standard electronics. The significant negative charge of our MS (see Methods) will contribute to a strong electrophoretic driving force, potentially reducing translocation time further. We therefore conclude that the absence of observed MS events results from the bandwidth limitations of our apparatus, which allow a significant number of protein molecules to translocate undetected. Conversely, we attribute the small number of events measured for bio90 to infrequent translocations. Storm, *et al*.[40] showed experimentally that dsDNA dwell times scale with molecular length as a power-law with exponent 1.27. Their results were obtained using SS-nanopores of similar diameter to those employed here and with a comparable voltage. Assuming that the scaling factor holds for small molecular lengths (Wanunu, *et al*.[19] reported similar scaling for short dsDNA), we estimate that 90 bp dsDNA should translocate in ~50 μs; a value that is resolvable by our electronics. Indeed, of the few events that we observe with bio90 alone, most yield a dwell time near this value (see Supplementary Fig. 4). Thus, we conclude that under the low-voltage conditions investigated here, the translocation event rate for bio90 is minimal.

The dichotomy between these two explanations offers a possible mechanism for the increase in event rate observed for the admixture of MS and bio90. When the molecules form a complex, the large net electrical force experienced by MS alone is

countered by a significant viscous drag imparted by the bio90. We therefore propose that the combination of these forces results in translocations that are slow in comparison to MS alone and are thereby resolvable with our apparatus. Importantly, this results in selective isolation of target DNA rather than exclusion determined, for example, by SS-nanopore dimensions[21].

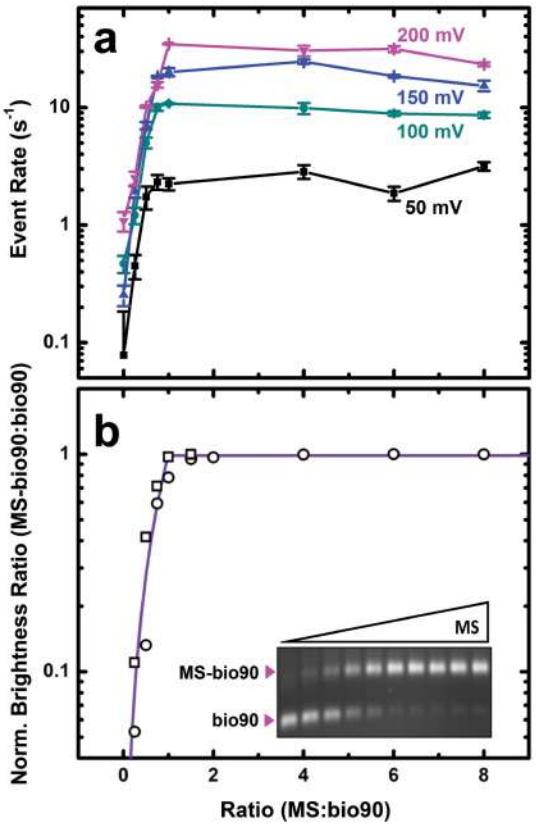

Figure 3-3 MS-bio90 titration experiments.

(a) Semilog plot showing stoichiometric dependence of SS-nanopore translocation event rate over a range of applied voltages. (b) Semilog plot (normalized) showing stoichiometric dependence of band intensity for MS-bio90 complex relative to bio90 measured from EMSA. Circles and squares are data from two separate assays. Inset shows an example gel (square symbols in plot).

In order to investigate the system further, we next perform a series of SS-nanopore measurements in which MS is titrated against a constant amount (1 μM) of bio90. Over all investigated voltages, we observe that the measured event rate rises dramatically up to a molar ratio of 1:1 (Figure 3-3a). However, from unity up to a molar ratio of 8:1 (MS:bio90), additional MS does not increase the event rate further. This is a result of the limited supply of dsDNA needed to form nucleoprotein complexes; the protein has an extremely low off rate[41] (~10-5 s-1) and each oligonucleotide contains only a single biotin moiety, so we expect that nearly all bio90 in solution will be bound at or above an equimolar concentration. Comparing our translocation results to an electromobility shift assay (EMSA) performed with MS and bio90 over the same stoichiometric range, we observe a strikingly similar trend (Figure 3-3b). These data support our assertion that virtually all observed translocation events for the admixture correspond to MS-bio90 complexes. Additional evidence of the high specificity of this approach is provided by control measurements in which non-biotinylated dsDNA incubated with MS yields a negligible event rate, equivalent to bio90 alone (insert into appendix see Supplemental Fig. 5).

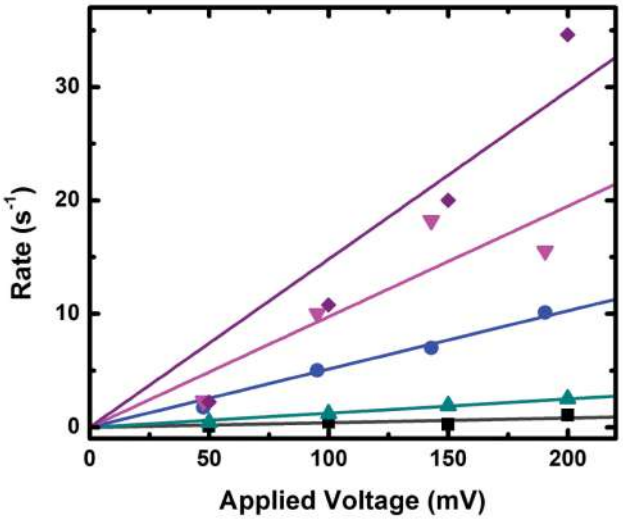

Figure 3-4 Voltage dependence of event rate. Event rates corresponding to specific MS-bio90 stoichiometries plotted over a range of applied voltages. From bottom to top, molar ratios (MS:bio90) are 0:1 (squares), 1:4 (upward triangles), 1:2 (circles), 3:4 (downward triangles), and 1:1 (diamonds). Solid lines are linear fits to each ratio set (color indicated).

Selective quantification of modified oligonucleotides In Figure 3-4, we examine admixture event rates up to a molar ratio of 1:1 and find a linear dependence on applied voltage. This implies that the capture process for the MS-bio90 complex is governed by diffusion rather than by interactions with the pore, in agreement with previous studies[36]. Importantly, the observed trend offers a route towards quantification of MS-bio90 complexes in solution. It has been demonstrated elsewhere[19, 26, 32] that event frequency varies with molecular concentration. Because nearly all events observed in our system can be attributed exclusively to the translocation of complexes, the linear fits in Figure 3-4 link the concentration of MS-bio90 in solution to specific event rates produced at a given voltage. The measurements described thus far have been performed in a protein-

limited regime (MS:bio90 < 1:1), and so the measured event rate has facilitated quantification of MS-bio90 complexes in a background of unconjugated bio90. However, the same approach could in principal be used to quantify biotinylated oligonucleotides in a heterogeneous solution with non-biotinylated DNA as well.

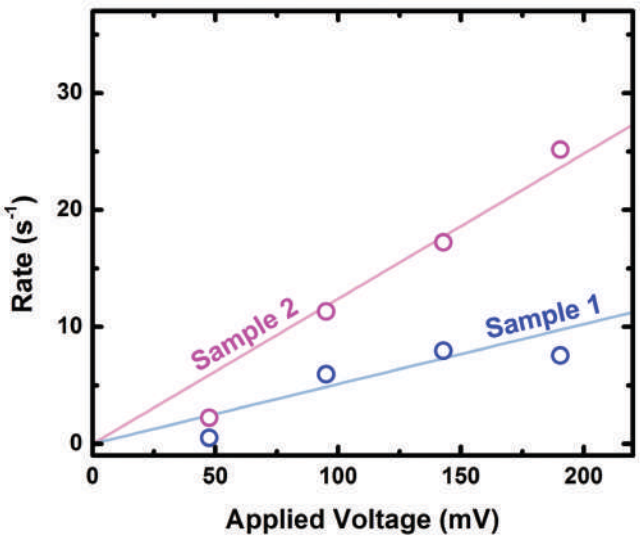

Figure 3-5 Analysis of blind samples. Measured event rate vs. applied voltage for the two prepared admixtures (see Fig. 4). Solid lines are linear fits to the data (i.e. values plotted in Fig. 4).

To investigate this possibility, we conduct a blind test on two samples prepared by a third party. Each of these samples contains a different mixture of biotinylated and non-biotinylated 90 bp dsDNA mixed to a total concentration of 1 μM (equivalent to that of the measurements described above). To ensure that all bio90 is conjugated, both solutions are incubated with MS at a concentration of 4 μM. As described in the previous sections, MS alone produces a negligible number of measurable events, and so excess protein does

not perturb the measurements. SS-nanopore analysis reveals a linear relationship between applied voltage and event rate for both samples, as expected (see Figure 3-5).

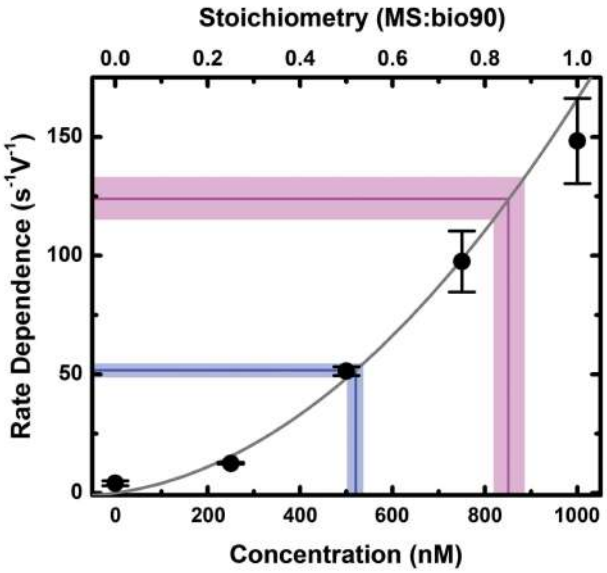

Figure 3-6 Selective determination of bio90 concentrations within a mixture. Solid data points indicate event rate dependence (i.e. slopes from Fig. 3) plotted against MS-bio90 concentration. Solid grey line is a second-order polynomial fit to the data. Experimentally-derived event rate values for samples prepared in blind experiments are shown as magenta (Sample 1) and blue (Sample 2) lines extending from y-axis. Shaded regions represent error.

Comparing the event rates obtained from the two blind samples to our prior measurements (Figure 3-6), we derive a value for the bio90 concentration in each: 850 ± 35 nM in Sample 1 and 520 ± 20 nM in Sample 2. Remarkably, these experimentally-determined concentrations are in excellent agreement with the prepared values of 800 ± 20 and 480 ± 20 nM, respectively. These results demonstrate that our SS-nanopore

approach is uniquely capable of quantifying DNA having single nucleotide biotin modifications selectively, even within a mixed sample.

Discussion In this study, we have demonstrated highly-specific detection and quantification of monobiotinylated dsDNA using SS-nanopores. We performed our detection through selective conjugation with a streptavidin protein containing a single biotin-binding domain. Under the experimental conditions investigated, we observed that the MS-bio90 complex produced a cascade of easily-resolved translocation events, in stark contrast to either constituent molecule individually. While the DNA constructs used here featured an internal biotin near the center of the molecule (see Methods), additional measurements indicate that the effect is not dependent on the position of the biotin moiety. The variation in translocation behavior was attributed to structure-dependent differences in net driving force, resulting in a dramatic increase in resolvable translocation events for the nucleoprotein complex as compared to DNA or protein alone. By studying stoichiometric effects, we showed that measured event rate correlated to the (partial) concentration of MS-bio90 conjugates in solution. Finally, we exploited the specificity of our measurement and the relationship between event rate and concentration to achieve selective quantification of modified oligonucleotides within a background of unmodified DNA. Upon examination of two different blind samples, we found excellent agreement with prepared values, validating our approach as a quantitative detection technique. An analog to conventional approaches like immunoprecipitation, our technique offers unique advantages, including molecular sensitivity and intrinsic quantification. We anticipate that our methodology will have direct impact on a variety of fields including diagnostic biomarker detection and gene profiling.

Methods

SS-nanopore device fabrication and electrical measurement Nanopores were fabricated using a technique described elsewhere[42]. Briefly, the beam of a scanning helium ion microscope (Carl Zeiss Orion Plus) was focused on a suspended silicon nitride thin film membrane (thickness 30 nm) in a silicon support chip. Calibrated exposure times were used to mill nanopores with diameters ranging from 7.3-7.7 nm. The support chip containing an individual pore was then positioned in a custom flow cell with fluid access to both sides of the membrane. Measurement solution (900 mM NaCl and 6 mM PBS buffer) was introduced on either side of the flow cell, and Ag/AgCl electrodes were immersed in the solution. Electrical measurements (Axopatch 200B) were used to verify that the device exhibited low RMS noise (typically <20 pA) and linear current-voltage characteristics that matched the calibrated nanopore diameter[32]. Translocation measurements were performed by replacing the solution on one side of the device with measurement solution containing biomolecules. Conductance was recorded at a bandwidth of 200 kHz and filtered at 100 kHz with a four-pole Bessel filter. Analysis was performed with custom software with which we applied an additional low-pass filter of 25 kHz to all measurements. The event threshold for analysis was set at 4 standard deviations and events with durations from 12-2000 μs were considered.

Biomolecules Bio90 oligonucleotides were purchased (Integrated DNA Technologies, Coralville, IA) with the sequence: TGT ATA CCA TGG CCA GGA TCC TGG GCC ATC TGG TA$\underline{T}^{B}$ GTA ATT CAT AAA GAA TTC TCA TTC TGC AGG TGC ACA

TGT TAA CAC TAG TCG TGA. The $\underline{T}^{B}$ represents a single internal biotinylated dT. The opposing strand (forming the dsDNA) contained no modified nucleotides. The non-biotinylated oligonucleotide used in the mixture (blind measurements) had the same sequence but with no biotin moiety. The streptavidin variant employed (SAe1D3) contained one active biotin-binding site[41] and was supplied by the Howarth lab (Oxford University). This mutant protein (54.5 kDa) retains binding affinity and stability similar to wild-type streptavidin and contains a hexaglutamate tag[43] used for isolation that imparts a negative charge of -17.1e under comparable pH conditions.

Electrophoretic Mobility Shift Assay MS was incubated in 1× PBS buffer with bio90 for 20 minutes at room temperature at molar ratios ranging from 0:1 to 8:1 (MS:bio90). The mixtures were then loaded onto a 1.5% agarose gel with ethidium bromide for visualization. The buffer reservoir of the electrophoresis unit was submerged in an ice bath to minimize dissociation of the protein-DNA complex.

Supporting Information Available: Additional figures depicting recapture of translocated MS-DNA constructs, scatter plots as a function of applied voltage and stiochiometry, dwell time analysis, control measurements with non-biotinylated DNA, and analysis of blind sample measurements. This material is available free of charge via the Internet at http://pubs.acs.org.

Corresponding Author

*To whom correspondence should be addressed: arhall@wakehealth.edu

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript. ‡These authors contributed equally.

Acknowledgements

We thank M. Marshall for helpful discussions. This work was supported by the North Carolina Biotechnology Center through Biotechnology Research Grant #2011-BRG-1201 and through start-up funds from Wake Forest University School of Medicine. JR and EWT acknowledge funding from the Dr. Arthur and Bonnie Ennis Foundation, Decatur, IL.

References

1. Peng, J.; Schwartz, D.; Elias, J. E.; Thoreen, C. C.; Cheng, D.; Marsischky, G.; Roelofs, J.; Finley, D.; Gygi, S. P. *Nat. Biotechnol.* **2003,** 21, (8), 921-926.

2. Rual, J.-F.; Venkatesan, K.; Hao, T.; Hirozane-Kishikawa, T.; Dricot, A.; Li, N.; Berriz, G. F.; Gibbons, F. D.; Dreze, M.; Ayivi-Guedehoussou, N.; Klitgord, N.; Simon, C.; Boxem, M.; Milstein, S.; Rosenberg, J.; Goldberg, D. S.; Zhang, L. V.; Wong, S. L.; Franklin, G.; Li, S.; Albala, J. S.; Lim, J.; Fraughton, C.; Llamosas, E.; Cevik, S.; Bex, C.; Lamesch, P.; Sikorski, R. S.; Vandenhaute, J.; Zoghbi, H. Y.; Smolyar, A.; Bosak, S.; Sequerra, R.; Doucette-Stamm, L.; Cusick, M. E.; Hill, D. E.; Roth, F. P.; Vidal, M. *Nature* **2005,** 437, (7062), 1173-1178.

3. Zhang, B.; Park, B.-H.; Karpinets, T.; Samatova, N. F. *Bioinformatics* **2008,** 24, (7), 979-986.

4. Jacinto, F. V.; Ballestar, E.; Esteller, M. *BioTechniques* **2008,** 44, (1), 35-43.

5. Laird, P. W. *Nat. Rev. Genet.* **2010,** 11, (3), 191-203.

6. Song, C.-X.; Yi, C.; He, C. *Nat. Biotechnol.* **2012,** 30, (11), 1107-1116.

7. Williams, K.; Christensen, J.; Pedersen, M. T.; Johansen, J. V.; Cloos, P. A. C.; Rappsilber, J.; Helin, K. *Nature* **2011,** 473, (7347), 343-348.

8. Hirai, M. Y.; Klein, M.; Fujikawa, Y.; Yano, M.; Goodenowe, D. B.; Yamazaki, Y.; Kanaya, S.; Nakamura, Y.; Kitayama, M.; Suzuki, H.; Sakurai, N.; Shibata, D.; Tokuhisa, J.; Reichelt, M.; Gershenzon, J.; Papenbrock, J.; Saito, K. *J. Biol. Chem.* **2005,** 280, (27), 25590-25595.

9. Zhao, J.; Ohsumi, T. K.; Kung, J. T.; Ogawa, Y.; Grau, D. J.; Sarma, K.; Song, J. J.; Kingston, R. E.; Borowsky, M.; Lee, J. T. *Mol. Cell* **2010,** 40, (6), 939-953.

10. Haring, M.; Offermann, S.; Danker, T.; Horst, I.; Peterhansel, C.; Stam, M. *Plant Methods* **2007,** 3, 11.

11. Selbach, M.; Mann, M. *Nat. Methods* **2006,** 3, (12), 981-983.

12. Dekker, C. *Nat. Nanotechnol.* **2007,** 2, (4), 209-215.

13. Wanunu, M. *Physics of Life Reviews* **2012,** 9, (2), 125-158.

14. Freedman, K. J.; Haq, S. R.; Edel, J. B.; Jemth, P.; Kim, M. J. *Sci. Rep.* **2013,** 3, 1638.

15. Fologea, D.; Brandin, E.; Uplinger, J.; Branton, D.; Li, J. *Electrophoresis* **2007,** 28, (18), 3186-3192.

16. Hoogerheide, D. P.; Garaj, S.; Golovchenko, J. A. *Phys. Rev. Lett.* **2009,** 102, (25), 256804.

17. Smeets, R. M. M.; Keyser, U. F.; Krapf, D.; Wu, M. Y.; Dekker, N. H.; Dekker, C. *Nano Lett.* **2006,** 6, (1), 89-95.

18. Larkin, J.; Henley, R.; Bell, D. C.; Cohen-Karni, T.; Rosenstein, J. K.; Wanunu, M. *ACS Nano* **2013,** 7, (11), 10121-10128.

19. Wanunu, M.; Sutin, J.; McNally, B.; Chow, A.; Meller, A. *Biophys. J.* **2008,** 95, (10), 4716-4725.

20. Branton, D.; Deamer, D. W.; Marziali, A.; Bayley, H.; Benner, S. A.; Butler, T.; Di Ventra, M.; Garaj, S.; Hibbs, A.; Huang, X. H.; Jovanovich, S. B.; Krstic, P. S.; Lindsay, S.; Ling, X. S. S.; Mastrangelo, C. H.; Meller, A.; Oliver, J. S.; Pershin, Y. V.; Ramsey, J. M.; Riehn, R.; Soni, G. V.; Tabard-Cossa, V.; Wanunu, M.; Wiggin, M.; Schloss, J. A. *Nat. Biotechnol.* **2008,** 26, (10), 1146-1153.

21. Niedzwiecki, D. J.; Iyer, R.; Borer, P. N.; Movileanu, L. *ACS Nano* **2013,** 7, (4), 3341-3350.

22. Li, J.; Gershow, M.; Stein, D.; Brandin, E.; Golovchenko, J. A. *Nat. Mater.* **2003,** 2, (9), 611-615.

23. Skinner, G. M.; van den Hout, M.; Broekmans, O.; Dekker, C.; Dekker, N. H. *Nano Lett*. **2009,** 9, (8), 2953-2960.

24. Storm, A. J.; Chen, J. H.; Zandbergen, H. W.; Dekker, C. *Phys*. *Rev*. *E Stat*. *Nonlin*. *Soft Matter Phys*. **2005,** 71, (5 Pt 1), 051903.

25. Fologea, D.; Ledden, B.; McNabb, D. S.; Li, J. *Appl*. *Phys*. *Lett*. **2007,** 91, (5), 053901.

26. Plesa, C.; Kowalczyk, S. W.; Zinsmeester, R.; Grosberg, A. Y.; Rabin, Y.; Dekker, C. *Nano Lett*. **2013,** 13, (2), 658-663.

27. Larkin, J.; Henley, R. Y.; Muthukumar, M.; Rosenstein, J. K.; Wanunu, M. *Biophys*. *J*. **2014,** 106, (3), 696-704.

28. Bacri, L.; Oukhaled, A. G.; Schiedt, B.; Patriarche, G.; Bourhis, E.; Gierak, J.; Pelta, J.; Auvray, L. *J*. *Phys*. *Chem*. *B* **2011,** 115, (12), 2890-2898.

29. Hall, A. R.; Keegstra, J. M.; Duch, M. C.; Hersam, M. C.; Dekker, C. *Nano Lett*. **2011,** 11, (6), 2446-2450.

30. Prabhu, A. S.; Jubery, T. Z. N.; Freedman, K. J.; Mulero, R.; Dutta, P.; Kim, M. J. *J*. *Phys*. *Cond*. *Mat*. **2010,** 22, (45), 454107.

31. Venta, K.; Shemer, G.; Puster, M.; Rodríguez-Manzo, J. A.; Balan, A.; Rosenstein, J. K.; Shepard, K.; Drndić, M. *ACS Nano* **2013,** 7, (5), 4629-4636.

32. Wanunu, M.; Dadosh, T.; Ray, V.; Jin, J.; McReynolds, L.; Drndić, M. *Nat*. *Nanotechnol*. **2010,** 5, (11), 807-814.

33. Zhao, Q.; Sigalov, G.; Dimitrov, V.; Dorvel, B.; Mirsaidov, U.; Sligar, S.; Aksimentiev, A.; Timp, G. *Nano Lett*. **2007,** 7, (6), 1680-1685.

34. Shim, J.; Humphreys, G. I.; Venkatesan, B. M.; Munz, J. M.; Zou, X.; Sathe, C.; Schulten, K.; Kosari, F.; Nardulli, A. M.; Vasmatzis, G.; Bashir, R. *Sci. Rep*. **2013,** 3, 1389.

35. Wanunu, M.; Cohen-Karni, D.; Johnson, R. R.; Fields, L.; Benner, J.; Peterman, N.; Zheng, Y.; Klein, M. L.; Drndic, M. *J. Am. Chem. Soc*. **2011,** 133, (3), 486-492.

36. Gershow, M.; Golovchenko, J. A. *Nat. Nanotechnol*. **2007,** 2, (12), 775-779.

37. Plesa, C.; Cornelissen, L.; Tuijtel, M. W.; Dekker, C. *Nanotechnology* **2013,** 24, (47), 475101.

38. Carlsen, A. T.; Zahid, O. K.; Ruzicka, J.; Taylor, E. W.; Hall, A. R. *ACS Nano* **2014**, DOI: 10.1021/nn501694n.

39. Smoluchowski, M. v. *Z. Phys. Chem*. **1917,** 92, 129-168.

40. Storm, A. J.; Storm, C.; Chen, J.; Zandbergen, H.; Joanny, J.-F.; Dekker, C. *Nano Lett*. **2005,** 5, (7), 1193-1197.

41. Howarth, M.; Chinnapen, D. J. F.; Gerrow, K.; Dorrestein, P. C.; Grandy, M. R.; Kelleher, N. L.; El-Husseini, A.; Ting, A. Y. *Nat. Methods* **2006,** 3, (4), 267-273.

42. Yang, J.; Ferranti, D. C.; Stern, L. A.; Sanford, C. A.; Huang, J.; Ren, Z.; Qin, L.-C.; Hall, A. R. *Nanotechnol*. **2011,** 22, (28), 285310.

43. Fairhead, M.; Krndija, D.; Lowe, E. D.; Howarth, M. *J*. *Mol*. *Bio*. **2014,** 426, (1), 199-214.

4. Chapter 4 – Sequence Specific Recognition of Nucleic Acids with Solid-State Nanopores

MicroRNAs are important biomarkers of cancer. Other short conserved nucleic acid sequences can be used to identify infectious disease agents. Most technologies employ PCR based amplification tests to detect these nucleic acid biomarkers. However at current length scales (<30 bases) PCR tests suffer from significant false-positive results. Therefore, novel and highly sensitive techniques are required to rapidly detect important biomarkers while avoiding biased results. We adapted our nanopore assay for selective detection of specific sequences amongst a background of complex oligonucleotides. Our assay recognizes and quantifies short biotinylated dsDNA molecules when bound to MS. Based on previous observations (See Chapter 3) we hypothesized that ssbioDNA would pass through the pore undetected, free or bound to MS. These differences arise due to variations in persistence lengths of ss- and dsDNA, which are significant for short (<150 base) nucleic acids. While dsDNA is essentially a stiff rod-like structure, ssDNA is flexible. Therefore, ssbioDNA-MS complex is structurally different from dsbioDNA-MS and does not interact with the pore walls as it passes through. Our results demonstrate and support our hypothesis, we first probe a ssbioDNA (34 nts) and its dsbioDNA counterpart individually and in the presence of MS, we observe events exclusively for the dsbioDNA-MS complex. Subsequently, we demonstrate the ability to selectively target a ssDNA sequence (target), in the presence of background oligonucleotides using a complementary ssbioDNA tag (bait). Events are observed exclusively when the target sequence is present in the sample. We employ a similar approach to detect miR-155 microRNA, a biomarker overexpressed in lung-cancer patients and demonstrated its selective detection across

range of concentrations. Our initial results were performed on inherently short ss-oligonucleotides (<40 nts), which isn't the native form of a number of nucleic acid biomarkers that are present within a longer target molecule (≥1,500 nts). For example, short conserved sequences in 16SrRNA that are biomarkers for bacterial detection. We developed an enzymatic assay to address these sequences, thus allowing us to detect any particular sequence with our nanopore assay from a long target nucleic acid molecule.

Peer-reviewed Publication: '*Sequence-specific recognition of microRNAs and other short nucleic acids with solid-state nanopores', O. K. Zahid, F. Wang, J. A. Ruzicka, E. W. Taylor, A. R. Hall, Nano Letters, 16 (3), 2033–2039, 2016*

4.1. Sequence-specific recognition of microRNAs and other short nucleic acids with solid-state nanopores

Abstract: The detection and quantification of short nucleic acid sequences has many potential applications in studying biological processes, monitoring disease initiation and progression, and evaluating environmental systems, but is challenging by nature. We present here an assay based on the solid-state nanopore platform for the identification of specific sequences in solution. We demonstrate that hybridization of a target nucleic acid with a synthetic probe molecule enables discrimination between duplex and single-stranded molecules with high efficacy. Our approach requires limited preparation of samples and yields an unambiguous translocation event rate enhancement that can be used to determine the presence and abundance of a single sequence within a background of non-target oligonucleotides.

Keywords: nanopore, sequence, microRNA, detection, cancer

The roles of short nucleic acid sequences are multiple and diverse. A prominent example is microRNA (miRNA), a family of 18-25 nt, non-coding RNAs that regulate a wide variety of cell functions and take part in post-transcriptional silencing of genes[1,2]. miRNAs are especially promising as biomarkers for cancer because anomalous levels have been identified in many tumor types[3]. In addition, they are known to exist as cell-free nucleic acids (cfNA) in serum and blood plasma[4], making them potentially easy to collect through non-invasive means. However, such sequences present unique challenges for conventional detection technologies. For example, quantitative real-time PCR is prone to amplification errors[5] while microarray assays require extensive design and labeling of

probes or targets to be validated. As a result, new approaches are highly desirable. We recently reported[6] a SS-nanopore approach for the selective detection of target DNA. In conventional SS-nanopore measurements, a nanometer-scale aperture formed in a thin membrane is positioned between two chambers of electrolyte solution and used for resistive pulse sensing of molecules as they thread electrically through it (Figure 4-1a). In our approach, a set of biomolecules is used that individually do not yield significant translocation signals (events): a monovalent variant of streptavidin[7] (MS) and a short, biotinylated double-stranded DNA (dsDNA).

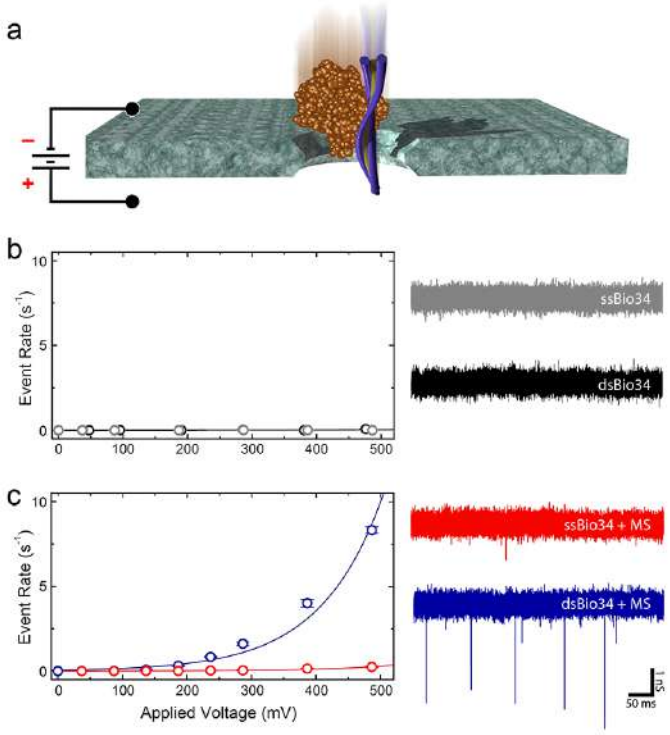

Figure 4-1 Selective detection of dsDNA.

(a) Schematic showing a short duplex DNA molecule bound to MS translocating electrically through a SS-nanopore. (b) Event rate vs. applied voltage for ssBio34 (grey) and dsBio34 (black) without MS bound (1 μM). Example trans-membrane ionic current

traces (colors matched) recorded at 500 mV are shown to the right. (c) Event rate vs. applied voltage for ssBio34 (red) and dsBio34 (blue) with MS bound (1 μM). All solid lines are exponential fits to the data. Example traces (colors matched) recorded at 500 mV are shown to the right. Scale bar (lower right) applies to all traces.

When the two bind, the construct yields a large increase in event rate that can be used for direct molecular quantification. Here, we apply this new assay to the high-fidelity detection of specific miRNAs and other short nucleic acid sequences in solution.

The mechanisms underlying our method are yet unestablished. It has been demonstrated that wild type streptavidin translocates rapidly[8] and is thus beyond the resolution of conventional electronics[9]. This is also true for the MS we use, which contains an additional hexaglutamate tag for isolation[10] and thus has a net charge of -17.1*e*, further increasing its electrophoretic velocity. However, the absence of events for short dsDNA, and thus the origin of the rate enhancement itself, is less clear. Our initial working hypothesis[6] was that nucleic acid attachment to the MS supplied an additional hydrodynamic drag to the fast-moving protein, thus slowing nucleoprotein translocation to a resolvable speed. However, this implied that the dsDNA itself experienced a net repulsive force, possibly caused by electroosmotic shear force counteracting elecrophoresis. While translocation physics at this length scale is not fully characterized and experimental factors like solvent conditions and SS-nanopore surface charge can alter expected dynamics[11], electrophoretic velocity and direction is understood to be independent of length[12], especially under high-salt conditions where the Debye layer is thin. In addition, a limited number of previous reports[13–16] have directly measured the translocation of short DNA through SS-nanopores, albeit under different experimental

conditions. Together, these considerations suggest that short oligonucleotides indeed translocate under a positive bias, potentially making our initial interpretation unphysical.

An alternative hypothesis is that the short dsDNA translocations are also rapid and thus challenge bandwidth limitations, much like those of MS. In this case, the subsequent rate enhancement could arise from transient interactions[17] with the SS-nanopore walls, facilitated by the bulkier nucleoprotein complex and resulting in protracted and resolvable event durations. While we observe that our measurements are highly stable, with individual nanopores supporting thousands of events without clogging, this effect does appear to be at play. Nonetheless, other considerations are likely important as well. For example, capture of dsDNA of length <8000 bp into a nanopore can be described well by an entropic barrier-limited model[18,19], dependent principally on molecular orientation, but also with weaker dependences on electrophoretic and conformational components. In the case of sub-persistence length DNA, the conformational term is irrelevant. However, extending to our nucleoprotein complex, the dominant orientational factor (rotational hydrodynamics) will be perturbed by the bound MS and the normally weak electrophoretic component may be influenced strongly by the significant charge of the protein, especially for very short DNA. These mechanistic details will be the subject of further studies.

Because our assay is able to selectively probe monobiotinylated dsDNA, we hypothesized that hybridization between target single-stranded (ss-) molecules and complementary synthetic biotinylated oligonucleotides could be used to identify sequences in solution. However, such an approach would be possible only under the condition that biotinylated ssDNA itself does not produce events, either individually or in

complex with MS. To investigate this, we focused on two short DNA constructs (see Table S1): a 34 nt ssDNA containing a single biotin moiety (ssBio34) and the same oligonucleotide hybridized to its (non-biotinylated) complementary sequence to form 34 bp dsDNA (dsBio34). When introduced to a SS-nanopore independently, neither molecule yielded translocation events (Figure 4-1b). Next, we repeated the measurements following incubation of each molecule with an excess of MS to form nucleoprotein complexes. For the ssBio34:MS, we again observed very few events across the entire investigated voltage range (Fig. 1c, red); for example, the rate at 500 mV was only 0.24 s^{-1}. However, under identical conditions, dsBio34:MS yielded a dramatic increase in translocation event rate that scaled exponentially[19] with applied voltage (Figure 4-1c, blue). For voltages above 200 mV, the rate was greater than an order of magnitude higher than that of ssBio34:MS.

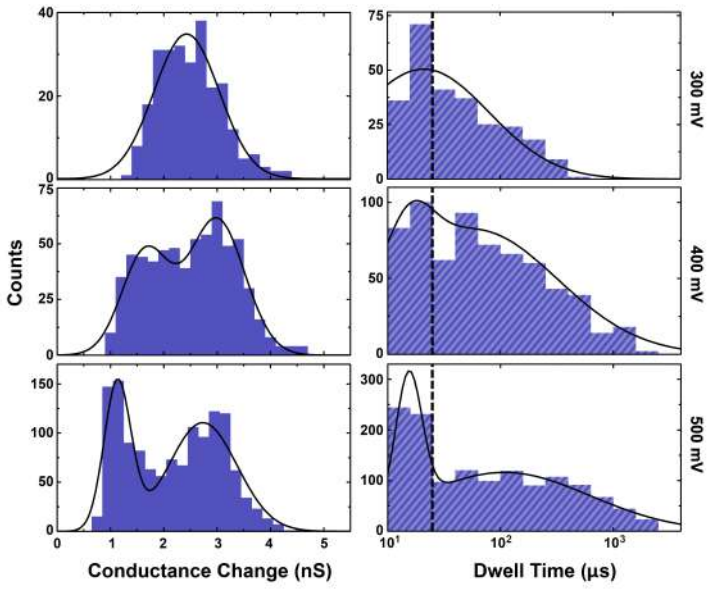

Figure 4-2 Analysis of dsDNA:MS events. Event depth (left) and duration (right) histograms of dsDNA:MS (1μM) from 300 (top) to 500 mV (bottom). Solid lines are

Gaussian fits to the data and the dashed line represents the temporal resolution of our system. Number of events represented are (T-B) 262, 652, and 1332, respectively.

The mean depth and duration of events (**Figure 4-2**) at lower voltages showed a single, deep (2.5-3 nS) blockade level and translocation durations at or below the temporal resolution of our system, similar to our previous report with 90 bp dsDNA[6]. As voltage was increased further, however, we observed the emergence of a bimodal distribution. This behavior is likely caused by the occurence of both translocation and collision events brought on by two discrete modes of molecular orientation during translocation[20] and made resolvable by the increased signal-to-noise ratio at high voltage. Regardless of the explanation, these experiments demonstrated the viability of sequence detection.

We next utilized our approach to recognize a specific sequence within a heterogeneous mixture (Figure 4-3). Here, we used ssBio34 as a "bait" sequence; by incubating it with a mixture of ssDNA, dsBio34 could be formed only if its complement (the "target" sequence) was present. To test this, we first prepared a mixture of three unlabeled ssDNA oligonucleotides (see Table S1) with low (~25%) homology to the bait or target to act as non-specific decoy sequences.

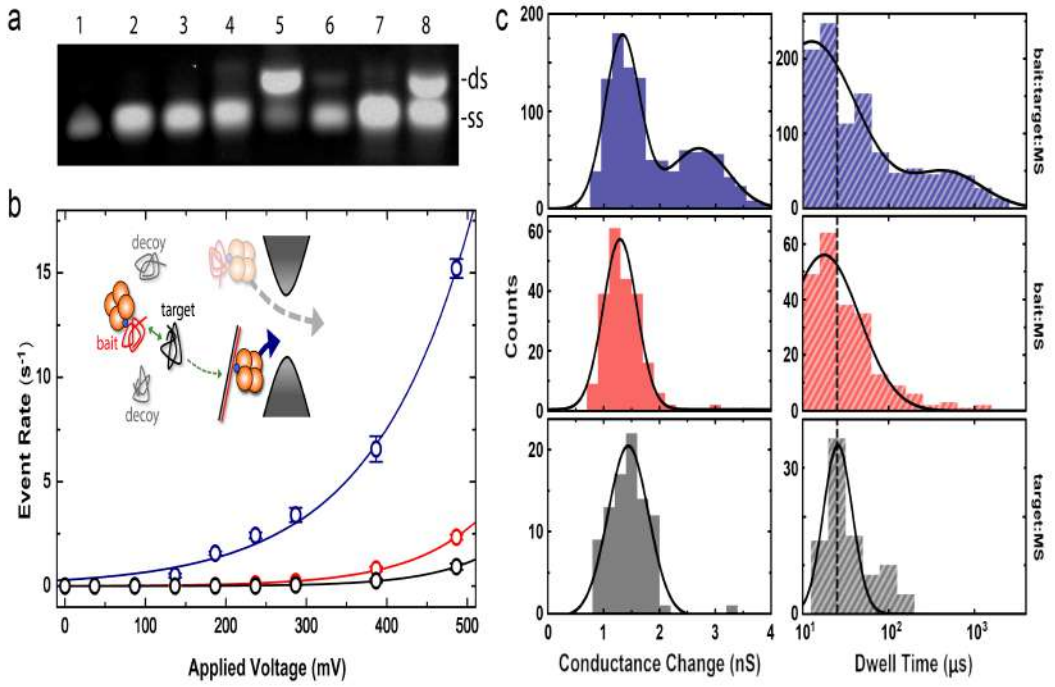

Figure 4-3 Sequence selection within a mixture.

(a) Agarose gel electrophoresis analysis of: bait (lane 1), decoys 1-3 (lanes 2-4), bait:target (equimolar ratio, lane 5), target (lane 6), bait:decoys (lane 7), bait:target:decoys (equimolar ratio, lane 8). ds and ss denote double- and single-stranded oligonucleotides, respectively. (b) Event rate vs. applied voltage for the target sequence (non-biotinylated complement to ssBio34, black), the bait sequence (ssBio34, red), and both the bait and target sequences (blue). All mixtures were incubated with MS among a background of three non-complementary decoy oligonucleotides. All molecules were supplied at a concentration of 1 μM. Solid lines are exponential fits to the data. Inset: schematic of sequence selection detection. (c) Mean event depth (left) and dwell time (right) histograms for each sample (colors match (b)). Solid lines are Gaussian fits to the data and the dashed line represents the temporal resolution of our system. Note that a second peak in the red and grey data sets could not be reliably fit due to the insignificant

number of deep and long events. Total number of events considered are (T-B) 1080, 223, and 89, respectively.

Combining these decoys with target sequence, performing a single thermal cycle to promote annealing, and incubating with an excess of MS yielded no significant SS-nanopore translocation events (Figure 4-3b, black). This followed our expectations since presumably only nonbiotinylated ssDNA molecules were present. An identical protocol with background sequences and ssBio34 bait similarly produced very few events (Figure 4-3b, red) because dsBio34 is absent. We did observe a minor increase in capture rate as compared to ssDNA alone or to ssBio34:MS alone, which we attributed to partial hybridization; intra-strand complexes may be sufficiently large to promote minor rate enhancement. It is likely that this effect could be minimized through manipulation of experimental conditions like solvent temperature to discourage low-energy hybridization.

When bait, target, and background sequences were co-annealed and incubated with MS, the resulting mixture again produced a large increase in the number of recorded events (Figure 4-3b, blue). Although the number of translocation events generated by bait:decoys:MS was not insignificant, bait-target coupling was marked by a relative enhancement of about an order of magnitude under applied voltages of >150 mV. Consequently, this result demonstrated that our assay could be used to discriminate a single sequence of interest from a heterogeneous mixture with high specificity. We note that the event rate for bait:target:MS among background sequences was somewhat higher than was observed for the same concentration of dsBio34:MS alone (Figure 4-1c). This may be caused by both the intra-strand interactions described above and minor pore-to-pore variations (e.g. local charge density or diameter). Figure 4-3c shows a representative

set of event depth and duration histograms for the three samples, collected at 500 mV. For the bait:target:MS sample, we found a bimodal distribution, similar to the high-voltage data in Figure 4-2. We again attribute this to the occurrence of both translocation and collision events. For the control samples, we observed a single, Gaussian distribution in both values, with most events occurring at or below the temporal resolution of our instrument. This suggests that the population to the right in each histogram (high conductance change and long duration) represents collisions, since they are far more prevalent for the larger nucleoprotein complex; only a few bait:decoy:MS events (>1%) fall within the same range. This distribution disparity may offer an additional metric by which to confirm detection of target sequences.

Having established the ability of the assay to selectively detect a target sequence, we finally demonstrated that the approach could be used to detect a specific miRNA. We used as a demonstration vehicle the miRNA designated hsa-mi-R155 (miR155), a 23 nt ssRNA that is an established biomarker of lung cancer[21,22]. To accomplish miRNA sequence detection, we employed a ssDNA bait construct (ssBio23, See Table 4-1) with a single internal biotin and complementarity to the miR155 sequence (Supplementary Fig. S1) [add to appendix]. When ssBio23 was annealed with miR155 target to form a DNA-RNA heteroduplex (Figure 4-4a), we observed an enhanced voltage-dependent event rate across a range of concentrations (Figure 4-4b).

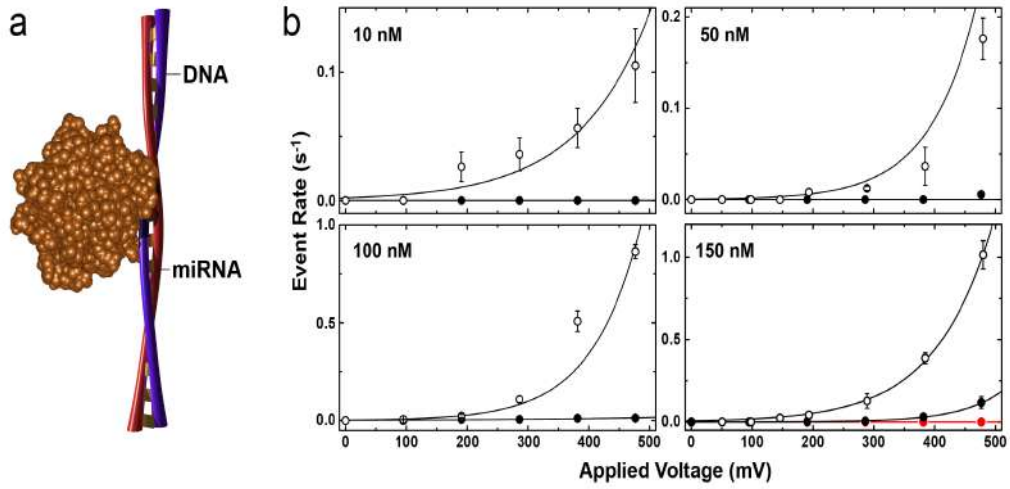

Figure 4-4 miRNA detection. (a) Schematic of monobiotinylated ssDNA bait oligonucleotide (blue) annealed to target miR155 ssRNA (red) to form a 23 bp DNA-RNA heteroduplex with bound MS (orange). (b) Typical plots of event rate vs. applied voltage for ssBio23:MS (closed circles) and ssBio23:miR155:MS (open circles), at 10, 50, 100, and 150 nM, respectively. The red data points in the 150 nM plot are measurements of miR155 alone. Solid lines are exponential fits to the data.

As before, measurements with ssBio23, either alone or incubated with a molar excess of MS, yielded few events under all investigated conditions. Unsurprisingly, the same was true of the miR155 alone (Figure 4-4b, red). Event rate for the heteroduplex bound by MS was found to vary linearly with concentration for several voltages (Figure 4-5a), in agreement with previous SS-nanopore measurements of short nucleic acids[13]. Analysis of mean event depth and duration (Figure 4-5b) showed less striking bimodal distributions than for the 34 bp constructs described above, likely due to the significantly smaller size of these molecules. Notably, the rate enhancement was sufficient to detect with high resolution as little as 10 nM ssBio23:miR155 heteroduplex, within the range of physiological miR155 concentrations[23], for example. Collectively, these results

confirmed both the viability of our assay down to the length scale of 23 nt as well as its applicability to RNA targets, and suggests its potential clinical utility.

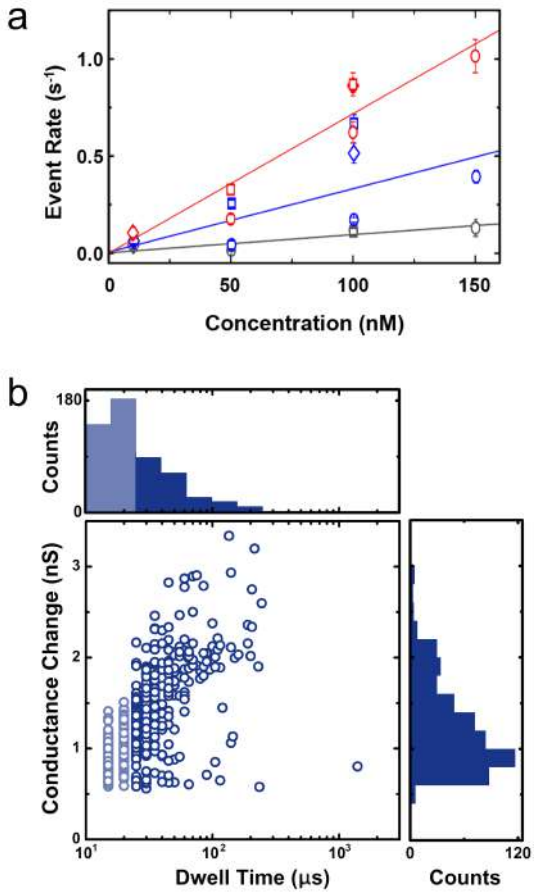

Figure 4-5 miRNA heteroduplex analyses.

(a) Event rate vs. concentration for ssbio23:miR155 heteroduplex with bound MS at applied voltages of 300 (grey), 400 (blue), and 500 (red). Different symbols represent data collected from different individual pores. Solid lines are linear fits to the data. (b) Scatter plot and accompanying histograms of mean event depth (conductance change) and duration (dwell time) for ssBio23:miR155:MS at 500 mV. Faded region (left) represents events below the temporal resolution of the system.

There have been only two prior reports of miRNA analysis with nanopore systems. Wanunu et al. pioneered SS-nanopore detection of miRNA, enabled by the use of ultrathin membranes[13]. They used their system to distinguish a single miRNA in rat liver extract. However, the approach required significant sample purification and concentration of miRNA target due to the non-selective nature of conventional SS-nanopore detection. A distinct difference between this and the present work is that our approach can resolve target sequences from among a background of ancillary molecules, potentially removing the need for extensive enrichment of the target sequence. Wang, et al. demonstrated sequence-specific detection of miRNA with the α-haemolysin protein nanopore[23], using the method to distinguish miR155 in patient serum. This important approach yielded information that is qualitatively similar to our results. However, the techniques differ in two key respects. First, the α-haemolysin system is built on lipid bilayers that are still challenging to integrate into stable device architectures, despite recent commercial progress[24]. Our system is built on silicon-based materials, which could promote easier integration and parallelization. Second, the previous measurement relied on differentiation between signals produced by target molecules as well as other background sequences, thus requiring some degree of thresholding and selection in analysis. In contrast, the present approach produces an unambiguous signal, essentially only yielding events when target sequences are present.

In conclusion, we have reported a SS-nanopore approach that uses hybridization with a biotinylated oligonucleotide to identify short nucleic acid sequences, including miRNA. The measurement is rapid and amenable to integration in wafer-scale device architectures. We first demonstrated that dsDNA could be differentiated from ssDNA

with high fidelity and subsequently that a single target sequence could be detected selectively. Indeed, identification from among a mixture of similarly-sized, non-complementary ssDNA was marked by a rate enhancement of more than an order of magnitude. Finally, we applied the technique to the *in vitro* sensing of a specific miRNA, miR155, that has importance as a lung cancer biomarker. We demonstrated detection of as little as 10 nM miRNA, validating the detection of physiologically-relevant concentrations[25]. We believe this resolution could be improved through the use of salt gradients[19,23] or further optimization of device dimensions, for example. This embodiment of SS-nanopore detection enables selectivity for arbitrary nucleic acid sequences and thus could be valuable for the sensitive analysis of biomarkers of disease[26], contamination[27], and bioterrorism agents[28].

4.2. Isolation and Detection of Specific Sequences from a Longer Targets

(Unpublished Results)

Nucleic acid sequences can be used to identify contamination or microbial infections. For example specific sequences can determine Influenza A viral strains[29] or an attractive biomarker for identification of bacterial species is the 16S rRNA conserved sequences[30]. The 16S rRNA is a ribosomal RNA molecule that is single-stranded and is approximately 1,500 bases in length, it's a component of 30S small subunit of the ribosome and is required for protein translation. There are conserved regions within the 16S rRNA sequence that are utilized for phylogenetic studies and can identify specific bacterial species accurately[31]. The most common techniques employed for this purpose are PCR or sequencing, while sequencing is highly sensitive, it can be expensive to perform. PCR analysis on the other hand suffers from false positives due to short lengths of the conserved target sequences.

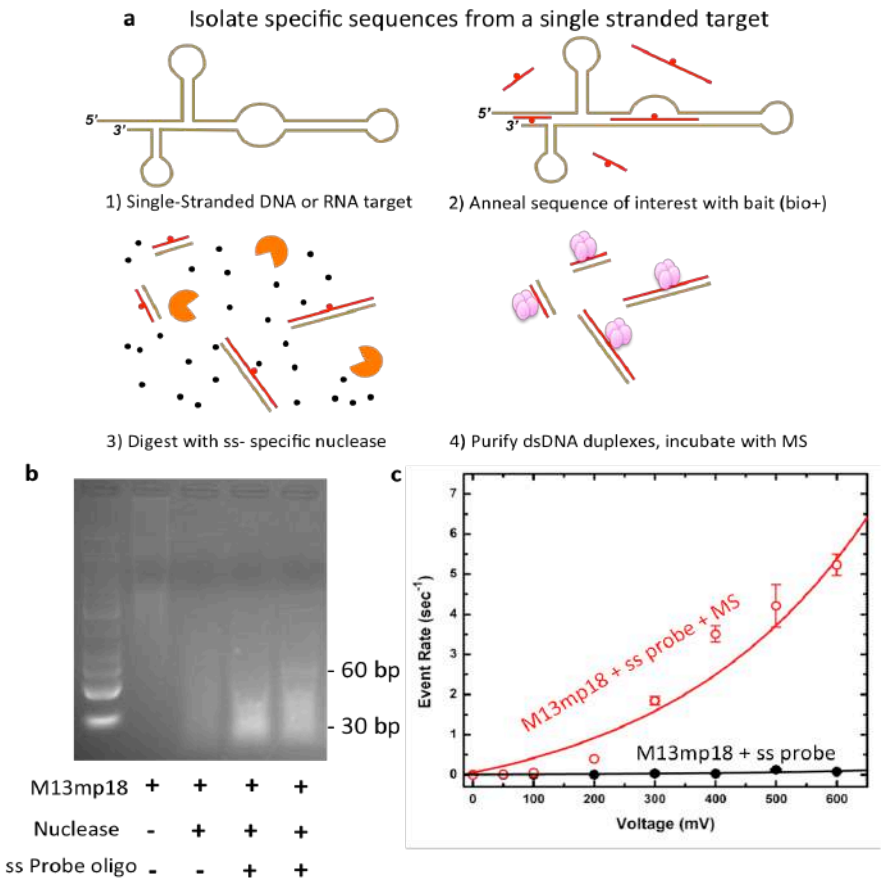

Figure 4-6 Isolation of specific target sequences from M13mp18 bacteriophage genome. (a) Schematic representation of individual steps involved in the sequence extraction protocol from a ss-target. (b) Isolation and detection of sequences from M13mp18 virus, analysis of samples on an agarose gel, individual components in each reaction is represented by symbol below. (c) Event rate graph of 60 bps sequence isolated from M13mp18, in the presence (red) and absence of MS (black) across a range of voltages.

We believe our selective nanopore assay can be adapted to rapidly identify infectious disease agents by targeting conserved sequences using complementary biotinylated probes. While this goal is beyond the scope of the current study, we have conducted

significant proof-of-principle measurements towards that objective. The idea is to be able to detect specific short (<250 bps) sequences from a long nucleic acid target that would allow us to identify a microbial target by our technology. Therefore the sample input must meet the assay specifications for operation; it must be relatively pure, short (<250bps), biotin-labeled, duplex oligonucleotide. As a proof-of-concept we have developed protocols necessary for isolating and converting the long nucleic acid target into detectable sequences for analysis. Only a few simple pretreatment steps will be necessary: (i) isolation of total genomic nucleic acid content (for example, total RNA including 16SrRNA) from a sample, (ii) annealing of one or more tagged probe sequences to the isolate, and (iii) removal of unannealed (single-stranded) nucleic acid regions to leave labeled duplex constructs only. We have developed each of these measures in turn. Since step (i) has been established in commercial kits and can be adapted as needed, we have focused on the remaining steps in our preliminary studies. Figure 4-6a shows our approach schematically, in which target sequences within long nucleic acid strands are annealed with biotin-labeled probes. A single-strand-specific nuclease, Mung Bean nuclease, is then used to digest all single-stranded material, leaving only target duplexes suitable for analysis. As a proof of principle, we tested this approach using M13mp18, a single-stranded viral genome. Once the sequence isolation protocol is established we should be able to target any region with a long single-stranded molecule of interest, design complementary biotinylated probes that can target those regions and detect them by our nanopore assay. After performing the treatment as described, we find that multiple, specific duplex sequences can be retrieved (Figure 4-7b). Subsequent nanopore analysis, using 60base ss-DNA probe in the presence of MS results in a positive

detection signal (Figure 4-6c), increase in event rate). This establishes the feasibility of the protocol as valid sample preparation steps for the assay.

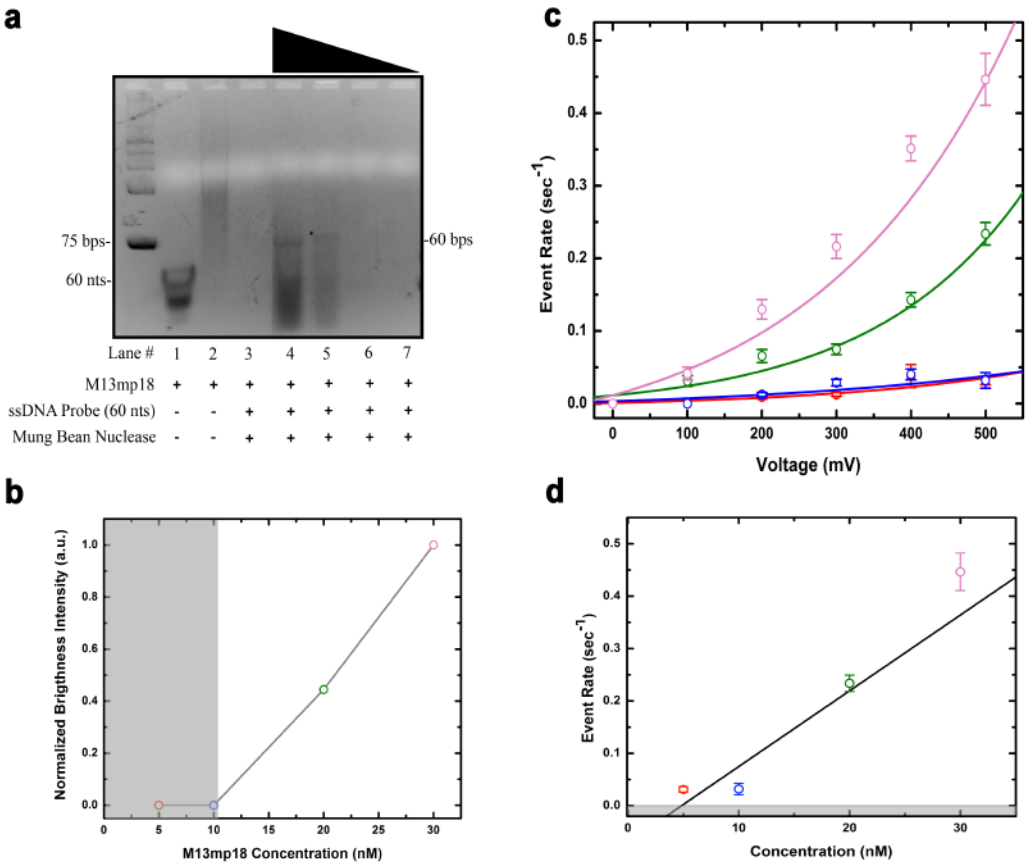

Figure 4-7 Titration of M13mp18 genomic DNA. (a) Agarose gel image of a M13mp18 titrated (50, 30, 20, 10 and 5nM) against ssbio60 probe, the key below gel represents the reaction components. Bands corresponding to 60bps dsDNA are observed in lanes 4-6, with no distinguishable bands in the final two lanes. (b) Normalized brightness intensity (lanes 8 to 5, i.e. low to high DNA concentration) plotted against the total M13mp18 genomic concentration in each reaction. Analysis indicates, below 10nM genomic sample content, sequence is not resolvable on a gel. (c) Event rate vs applied voltage results of isolated 60 base sequence from M13mp18 titration reactions post purification incubated with MS. (Pink, green, blue and red represents 30, 20, 10 and 5nM

of M13mp18 genomic DNA concentration). Solid lines are exponential fits to the data points. (d) Capture rate vs M13mp18 genomic DNA sample concentration at 500mV, measureable event rate is observed for both 5 and 10nM reaction concentrations as opposed to the gel results in (b). Solid line is a linear fit to the data points.

We also demonstrate ability to determine nucleic acid concentration by running another set of proof-of-principle nanopore measurements. M13mp18 genomic DNA was titrated against the ssbio60 probe and treated with our sequence isolation protocol. Initial gel electrophoresis analysis confirmed isolation of the target sequence (dsbio60) across the titrated concentrations, and the intensity of the bands gives a qualitative indication of M13mp18 concentration (Figure 4-7a). While a distinct band is observed for samples containing high concentration of M13mp18 genomic DNA, no isolated product was visible below 10nM concentration of the DNA sample (Figure 4-7a). This is mainly due to poor resolution of agarose gel electrophoresis, however these concentrations should readily be distinguished by the nanopore system. Subsequently we repeated the titration experiments, purified the samples using phenol-chloroform treatment followed by ethanol precipitation and investigated each sample by our nanopore assay (Figure 4-7c). We observed characteristic exponential rise in capture rate with applied voltage for all samples, including samples treated at low M13mp18 concentrations (5 and 10nM) which yielded no result by agarose gel analysis. Crucially, the capture rate increases for each sample with higher genomic M13mp18 content (Figure 4-7d), which can be compared to a concentration standard to determine quantify analyte concentration. These measurements are indicative that we can utilize our assay to detect specific target sequences of interest from a long ss-oligonucleotide fragment and determine its

concentration. The approach is modular since we should be able to target any ss-nucleic acid molecule for which we can determine unique sequence biomarkers, such as sequences in 16SrRNA[30,31] or M and HA genes[29] for identification and subtyping of bacterial and influenza infections, respectively.

Materials and methods

Biomolecule preparation All synthetic nucleic acids, including 34 nt biotinylated (bait) and non-biotinylated (target) ssDNA, low-homology background sequences, 23 nt biotinylated (bait) ssDNA, 23 nt miR155 (target) ssRNA, 30 nt, 60 nt and 90 nt monobiotinylated ssDNA probes complementary to M13mp18 bacteriophage sequence were obtained commercially (Integrated DNA Technologies, Coralville, IA). Sequences are provided in Table 4-1. All ssDNA molecules were resuspended in pure deionized water (Milli-Q) to a stock concentration of 200 μM and stored at −20° C prior to use. miR155 was aliquoted out in clean eppendorf tubes and stored to prevent contamination and degradation. Complementary oligonucleotides were hybridized by incubating the samples at a 1:1 molar ratio in pure Di H2O at 95° C for 10 minutes and gradually cooling to room temperature to generate duplex material (dsBio34 or 23 bp heteroduplex) at a final concentration of 8 μM, as confirmed by spectrophotometry. Hybridization was confirmed by gel electrophoresis. An identical hybridization reaction was conducted in the presence of three decoy ssDNA oligonucleotides (42, 40 and 33 bps in length respectively) for the selectivity measurement. Constructs were validated on a 4% agarose gel prepared in 1X TBE buffer with GelRed nucleic acid stain (Phenix Research Products, Candler, NC). All gel images were captured using a Gel Doc™ system (BioRad, Hercules, CA).

Oligonucleotide Label	Sequences (5'→ 3')
ssBio34	CAGTTGAGGATCCCCATAATGCGGCTGTTTTCTG
ssBio34 complement (i.e. 'target' sequence)	CAGAAAACAGCCGCATTATGGGGATCCTCAACTG
Decoy 1	CAGGGCTAGCAAAGGATTTTGCTATAACATGGGTGGCAAG
Decoy 2	TTTTTAAAACCTGATGAAACATTTGTACATTCAGGTTTTATC
Decoy 3	ATCAGGACCACATTCAAGAGGCCAGGAACCAAGACAGTGA
ssBio23	TCCCCTATCACGATTAGCATTAA
miR155	UUAAUGCUAAUCGUGAUAGGGGU
DNA miR155 homolog	TTAATGCTAATCGTGATAGGGGT

Table 4-1 Oligonucleotide Sequences. T signifies biotinylated thymine.

Sequence isolation protocol Complementary monobiotinylated probes (30, 60, & 90 nts) were incubated with M13mp18 (New England BioLabs, Ipswich, MA) at final concentrations of 4 μM and 20 nM, respectively. An annealing reaction was performed by incubating the samples at 95° C for 10 mins and gradually cooling the samples down to room temperature. Immediately after the annealing reaction, 20 units of Mung bean Nuclease and 10X buffer component (New England BioLabs, Ipswich, MA) were introduced to obtain a final 1X buffer concentration to the samples. The reactions were incubated in a water bath set at 30° C for 25 mins and 2 units of Proteinase K (New England BioLabs, Ipswich, MA) were added to inactivate the Mung Bean Nuclease protein. Subsequently reactions were loaded onto a 3% agarose gel prepared in 1X TBE buffer with GelRed nucleic acid stain (Phenix Research Products, Candler, NC) to determine whether the sequences were isolated.

M13mp18 titration reactions Genomic M13mp18 bacteriophage ssDNA at varying final concentrations (50, 40, 30, 10, and 5nM) were annealed to ssbio60 DNA probe at a fixed concentration of 4 μM, by incubating the samples at 95° C for 10 mins and gradually cooling the samples down to room temperature. Subsequently Mung Bean Nuclease buffer was added to achieve a final 1X buffer concentration along with 20 units of Mung Bean Nuclease enzyme. Samples were incubated in a 30 °C water bath for 25 mins; reactions were terminated by adding 2 units of Proteinase K. Samples were purified by phenol: chloroform treatment, and were subsequently treated with ethanol precipitation to remove any residual chloroform left over in the aqueous phase. Samples were resuspended in 20 μL of pure DiH_2O, their concentration determined by NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA) and stored at -20 °C until further use.

Binding reaction incubation MS, a 54.5 kDa streptavidin variant (SAe1D3) that contains one active biotin-binding site[7], was supplied by the Howarth lab (Oxford University). The protein contains a covalent hexaglutamate tag[10] used for isolation that imparts a net charge of −17.1*e* under pH conditions comparable to those used here. For all short sequence detection including miRNA experiments, 2.5 μL of prepared oligonucleotides at a stock concentration of 8 μM was incubated with 1 μL of MS (50 μM stock) in 1X PBS at room temperature for 10 minutes, and brought to a final salt concentration of 900 mM NaCl and 0.5X PBS. The resulting mixture contained 1 μM DNA and a 2.5X molar excess of MS. When necessary, further dilutions were performed using measurement buffer.

For M13mp18 titration reactions, 2 μL of the purified dsDNA (highest concentration < 0.5μM) from each sample was incubated with MS at a final concentration of 0.5 μM in 1X PBS at room temperature for 10 mins. All reactions were brought to a final salt concentration of 900 mM NaCl and 0.5X PBS.

Nanopore fabrication, detection and analysis Silicon chips (4.4 mm) containing 25 nm thick, free-standing silicon nitride membranes were obtained commercially (Norcada, Inc., Alberta, Canada). In each membrane, an individual nanopore (diameter 7.5-9.0 nm) was fabricated using a scanning helium ion microscope (Carl Zeiss Orion Plus) technique described elsewhere[32]. Nanopore chips were stored in a 50% ethanol solution until use. Directly before measurement, a chip was rinsed with deionized water and ethanol, dried under filtered air flow, and then exposed to air plasma (30 W) for 2 min on each side before being placed into a custom Ultem 1000 flow cell and immediately introduced with measurement buffer on both sides. A patch clamp amplifier (Axopatch 200B) was used to apply voltage and record current through Ag/AgCl electrodes. Each device was verified to exhibit a steady baseline and linear current vs. voltage characteristics that corresponded to intended pore diameter[33]. Current traces were collected at a rate of 200 kHz with a 100 kHz four-pole Bessel filter and analyzed with custom software, through which an additional 25 kHz low-pass filter was applied to all data. Devices were stable throughout the measurements, typically supporting >4000 events. An event was defined as having amplitude above a threshold of 4.5σ and duration between 12.5 to 1200 μs. Rate was determined by analyzing for each voltage an uninterrupted current trace of either 150 s (for 34 bp datasets) or 370-740 s (for 23 bp datasets). Data was saved in

increments of 3.2 s and the standard deviation of the rates was used as measurement error. DNA measurements were performed in triplicate. We display typical results from a single nanopore for consistency and to avoid effects of pore-to-pore variation. Additional data is presented in Fig. S2 to demonstrate the repeatability of the measurement.

Acknowledgements The authors thank the Howarth lab (Oxford University) for supplying monovalent streptavidin and F. Perrino and S. Harvey (Wake Forest School of Medicine) for assistance with RNA handling. This work was supported by NIH grant 1R21CA193067. E.W.T. and J.A.R. acknowledge funding from The Dr. Arthur and Bonnie Ennis Foundation, Decatur, IL. A.R.H. acknowledges support from the 3M Non-tenured Faculty Award program.

Competing financial interests A.R.H declares the following financial interest. He is listed as inventor on a patent covering the presented SS-nanopore assay for sequence detection. The other authors declare no competing interests.

References

1. He, L. & Hannon, G. J. MicroRNAs: small RNAs with a big role in gene regulation. *Nat. Rev. Genet.* **5,** 522–531 (2004).

2. Carthew, R. W. & Sontheimer, E. J. Origins and Mechanisms of miRNAs and siRNAs. *Cell* **136,** 642–655 (2009).

3. Lu, J. *et al.* MicroRNA expression profiles classify human cancers. *Nature* **435,** 834–838 (2005).

4. Mitchell, P. S. *et al.* Circulating microRNAs as stable blood-based markers for cancer detection. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 10513–10518 (2008).

5. Ruijter, J. M. *et al.* Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data. *Nucleic Acids Res.* **37,** e45 (2009).

6. Carlsen, A. T., Zahid, O. K., Ruzicka, J. A., Taylor, E. W. & Hall, A. R. Selective detection and quantification of modified DNA with solid-state nanopores. *Nano Lett.* **14,** 5488–5492 (2014).

7. Howarth, M. *et al.* A monovalent streptavidin with a single femtomolar biotin binding site. *Nat. Methods* **3,** 267–273 (2006).

8. Plesa, C. *et al.* Fast Translocation of Proteins through Solid State Nanopores. *Nano Lett.* **13,** 658–663 (2013).

9. Larkin, J., Henley, R. Y., Muthukumar, M., Rosenstein, J. K. & Wanunu, M. High-Bandwidth Protein Analysis Using Solid-State Nanopores. *Biophys. J.* **106,** 696–704 (2014).

10. Fairhead, M., Krndija, D., Lowe, E. D. & Howarth, M. Plug-and-Play Pairing via Defined Divalent Streptavidins. *J. Mol. Biol.* **426,** 199–214 (2014).

11. Firnkes, M., Pedone, D., Knezevic, J., Doeblinger, M. & Rant, U. Electrically facilitated translocations of proteins through Silicon Nitride nanopores: conjoint and competitive action of diffusion, electrophoresis, and electroosmosis. *Nano Lett.* **10,** 2162–2167 (2010).

12. Brochardwyart, F. & Degennes, P. Electrophoretic Mobility of a Fractal Object. *Comptes Rendus Acad. Sci. Ser. Ii* **307,** 1497–1500 (1988).

13. Wanunu, M. *et al.* Rapid electronic detection of probe-specific microRNAs using thin nanopore sensors. *Nat. Nanotechnol.* **5,** 807–814 (2010).

14. Rosenstein, J. K., Wanunu, M., Merchant, C. A., Drndic, M. & Shepard, K. L. Integrated nanopore sensing platform with sub-microsecond temporal resolution. *Nat. Methods* **9,** 487–U112 (2012).

15. Venta, K. *et al.* Differentiation of Short, Single-Stranded DNA Homopolymers in Solid-State Nanopores. *ACS Nano* **7,** 4629–4636 (2013).

16. Lee, M.-H. *et al.* A Low-Noise Solid-State Nanopore Platform Based on a Highly Insulating Substrate. *Sci. Rep.* **4,** 7448 (2014).

17. Wanunu, M., Sutin, J., McNally, B., Chow, A. & Meller, A. DNA translocation governed by interactions with solid-state nanopores. *Biophys. J.* **95,** 4716–4725 (2008).

18. Rowghanian, P. & Grosberg, A. Y. Electrophoretic capture of a DNA chain into a nanopore. *Phys. Rev. E* **87,** 042722 (2013).

19. Wanunu, M., Morrison, W., Rabin, Y., Grosberg, A. Y. & Meller, A. Electrostatic focusing of unlabelled DNA into nanoscale pores using a salt gradient. *Nat. Nanotechnol.* **5,** 160–165 (2010).

20. Carlsen, A. T., Zahid, O. K., Ruzicka, J., Taylor, E. W. & Hall, A. R. Interpreting the Conductance Blockades of DNA Translocations through Solid-State Nanopores. *ACS Nano* **8,** 4754–4760 (2014).

21. Yanaihara, N. *et al.* Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. *Cancer Cell* **9,** 189–198 (2006).

22. Donnem, T. *et al.* Prognostic impact of MiR-155 in non-small cell lung cancer evaluated by in situ hybridization. *J. Transl. Med.* **9,** 6 (2011).

23. Wang, Y., Zheng, D., Tan, Q., Wang, M. X. & Gu, L.-Q. Nanopore-based detection of circulating microRNAs in lung cancer patients. *Nat. Nanotechnol.* **6,** 668–674 (2011).

24. Ip, C. L. C. *et al.* MinION Analysis and Reference Consortium: Phase 1 data release and analysis. *F1000Research* (2015). doi:10.12688/f1000research.7201.1

25. Schmittgen, T. D. *et al.* Real-time PCR quantification of precursor and mature microRNA. *Methods San Diego Calif* **44,** 31–38 (2008).

26. Chen, X. *et al.* Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases. *Cell Res.* **18,** 997–1006 (2008).

27. Greisen, K., Loeffelholz, M., Purohit, A. & Leong, D. Pcr Primers and Probes for the 16s Ribosomal-Rna Gene of Most Species of Pathogenic Bacteria, Including Bacteria Found in Cerebrospinal-Fluid. *J. Clin. Microbiol.* **32,** 335–351 (1994).

28. Lim, D. V., Simpson, J. M., Kearns, E. A. & Kramer, M. F. Current and developing technologies for monitoring agents of bioterrorism and biowarfare. *Clin. Microbiol. Rev.* **18,** 583–+ (2005).

29. Gall, A. *et al.* Design and Validation of a Microarray for Detection, Hemagglutinin Subtyping, and Pathotyping of Avian Influenza Viruses. *J. Clin. Microbiol.* **47,** 327–334 (2009).

30. Clarridge, J. E. Impact of 16S rRNA Gene Sequence Analysis for Identification of Bacteria on Clinical Microbiology and Infectious Diseases. *Clin. Microbiol. Rev.* **17,** 840–862 (2004).

31. Srinivasan, R. *et al.* Use of 16S rRNA Gene for Identification of a Broad Range of Clinically Relevant Bacterial Pathogens. *PLOS ONE* **10,** e0117617 (2015).

32. Yang, J. *et al.* Rapid and precise scanning helium ion microscope milling of solid-state nanopores for biomolecule detection. *Nanotechnology* **22,** (2011).

33. Smeets, R. M. M. *et al.* Salt dependence of ion transport and DNA translocation through solid-state nanopores. *Nano Lett.* **6,** 89–95 (2006).

34. Wanunu, M., Morrison, W., Rabin, Y., Grosberg, A. Y. & Meller, A. Electrostatic focusing of unlabelled DNA into nanoscale pores using a salt gradient. *Nat. Nanotechnol.* **5,** 160–165 (2010).

5. Chapter 5 – Novel method to quantify mammalian 5-hydroxymethylcytosine content and development of the demethylation pathway labeling assay

5-hydroxymethylcytosine (5hmC), an oxidized form of 5-methylcytosine (5mC) was recently discovered in mammalian tissue. Since its discovery, research has been conducted to identify its biological function. While 5hmC appears as a stable epigenetic mark, it can be further oxidized by the Tet family of dioxygenases (Tet1, 2 and 3) to 5-formylcytosine (5fC) and finally to 5-carboxylcytosine (5caC). 5fC and 5caC are substrates of Thymine DNA Glysocylase (TDG), which removes these bases and the BER pathway can replace them with unmodified cytosine forming an active demethylation pathway. When compared to healthy cells, 5hmC content is significantly depleted in all forms of cancers. Therefore, 5hmC quantification can be used as a cancer screening biomarker. However, most methods that detect methylation modifications, including bisulfite sequencing are unable to distinguish 5mC from 5hmC. Therefore, novel methods like OxBS-Seq and TAB-Seq were recently developed that are performed in conjunction with traditional bisulfite treatment, these methods can determine the amount and position of 5hmC but aren't rapid. Similarly, HPLC while being the gold-standard technique for quantification requires sophisticated equipment and prolonged experimental runs to determine 5hmC content. Therefore, both techniques suffer from constraints that prevent them to be implemented for regular 5hmC monitoring. We sought to address this issue by employing our binary detection assay to selectively detect and quantify 5hmC genomic content.

Finally, we end the chapter with a novel assay that selectively labels all cytosine modifications involved in the demethylation pathway with a biotin-tag. Utilizing this

method we should be able to conduct highly specific affinity purification of genomic content containing any cytosine epigenetic modification and better understand the dynamics of demethylation in diseased states.

Peer-reviewed Publication: '*Zahid, O. K.; Zhao, B. S.; He, C.; Hall, A. R. Quantifying Mammalian Genomic DNA Hydroxymethylcytosine Content Using Solid-State Nanopores. Sci. Rep., 6, 29565, 2016*

5.1. Quantifying mammalian genomic DNA hydroxymethylcytosine content using solid-state nanopores

Abstract: 5-hydroxymethylcytosine (5hmC), the oxidized form of 5-methylcytosine (5mC), is a base modification with emerging importance in biology and disease. However, like most epigenetic elements, it is transparent to most conventional genetic techniques and is thus challenging to probe. Here, we report a rapid solid-state nanopore assay that is capable of resolving 5hmC with high specificity and sensitivity and demonstrate its utility in assessing global modification abundance in genomic DNA.

Introduction

Since its positive identification in the mammalian genome in 2009[1,2], the epigenetic DNA modification 5hmC has been the subject of intense investigation. Formed through the action of the ten-eleven translocation (TET) family of dioxygenases, 5hmC acts both as an intermediate in the active demethylation pathway and as an independent regulatory element, playing an important role in cell differentiation[3], development[4–6], aging and neurological disorders[7,8]. While positioning relative to the genetic sequence is critical[5], global levels of this modification have also been shown to be an important biomarker. For example, overall 5hmC abundance is significantly depleted across diverse cancers in comparison to healthy tissue[9]. As a result, quantification of 5hmC levels in cell-free DNA, for instance, could be a valuable tool for minimally-invasive monitoring of cancer initiation and progression. However, current methods for probing 5hmC either lack sensitivity or are too expensive or time-consuming to implement for regular clinical use, making the development of alternative approaches essential. The single-molecule approach of solid-state (SS-)nanopores[10] is an important candidate to address this

challenge, with precision and sensitivity that make it attractive for an array of applications that include analytical devices[11] and next-generation sequencing[12]. In the general approach, a nanometer scale aperture separates two chambers filled with electrolyte solution and a voltage applied between the chambers generates an ionic current (Figure 5-1a). The electrically-facilitated threading of molecules through the nanopore is marked by a transient change in this current (an "event"), through which properties of the translocating material can be obtained.

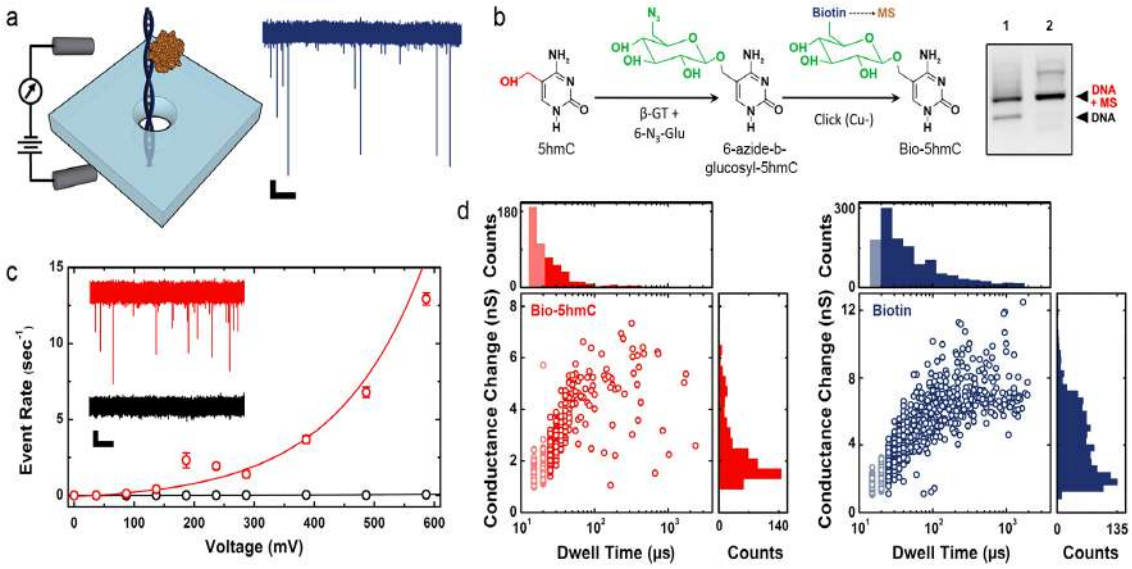

Figure 5-1 Nanopore detection of 5hmC modifications

(a) Schematic of measurement setup depicting a short biotinylated dsDNA bound by MS translocating through a SS-nanopore (left) and an example current trace for synthetic monobiotinylated 156 bp dsDNA measured at 400 mV (right). Scale bars are 1 nS and 250 ms. (b) 5hmC labeling scheme, consisting of glucosylation by βGT with 6-N_3-Glu to form 6-azide-b-glucosyl-5hmC and biotin functionalization by copper-free click chemistry. Right: EMSA showing biotin-labeled mono-5hmC (lane 1) and synthetic

monobiotinylated dsDNA (lane 2), both with excess MS. (c) Event rate vs. applied voltage for 156 bp mono-5hmC dsDNA (550 nM) biotin-labeled at 5hmC sites both with (red) and without (black) bound MS measured through a pore with diameter 8.5 nm. Solid lines are exponential fits to the data. Inset shows example current traces from each (400 mV, colors matched). Scale bars are 1 nS and 250 ms. (d) Scatter plots of mean event dwell time vs. mean conductance change collected and accompanying histograms for biotin-labeled mono-5hmC (red, diameter 8.5 nm) and synthetic monobiotinylated dsDNA (blue, diameter 7.8 nm) measured at 400 mV. Faded regions represent events below the resolution limit.

We have developed[13] a SS-nanopore assay that produces significant events only when short, double-stranded[14] DNA (dsDNA) binds to a chaperone protein (monovalent streptavidin, or MS[15]). Analyzed individually with a SS-nanopore of ~8 nm diameter, neither MS nor dsDNA generate significant events (See Chapter 3). The MS is both small and highly charged (-17.1*e*), owing to a covalent hexaglutamate tag used for isolation[15], and therefore passes through the pore at speeds beyond the resolution limits of conventional electronics[16]. While a few reports have demonstrated detection of short nucleic acids[17–20], these have all been conducted using either SS-nanopores fabricated specifically to do so[17–19], in ultra-thin membranes or with very small diameters, or using high bandwidth electronics to increase temporal resolution[20]. However, in more conventional systems, such molecules also translocate too rapidly to be detected. In contrast, the bulkier nucleoprotein complex interacts sterically with the SS-nanopore walls as it passes, resulting in prolonged, resolvable events. This intrinsic selectivity enables detection and quantification of dsDNA molecules containing a biotin moiety. We

recently applied an extension of this assay to the detection of specific nucleic acid sequences, including microRNAs[14]. Here, we extend the technique to probe 5hmC epigenetic modifications in genomic DNA.

Results

Assay demonstration with synthetic oligonucleotides

Because our assay is selective for biotinylated DNA, arbitrary base modifications can in principle be targeted for quantification through specific addition of a biotin linker to their structures. To study 5hmC, we utilize a strategy pioneered by Song, *et al*[5] that utilizes the T4 bacteriophage enzyme β-glycosyltransferase (β-GT) to transfer a glucose moiety specifically to 5hmC (Figure 5-1b). Through the use of a synthetic UDP-6 deoxy-6-azido-α-D-glucopyranoside (6-N_3-Glu), the incorporated azide group can subsequently be used for selective biotin functionalization through copper-free click chemistry. As an initial demonstration, we first employ the method to biotinylate a synthetic 156 bp dsDNA containing a single 5hmC nucleotide near its end (Supplementary Table 1 [insert appendix]), confirmed subsequently by electromobility shift assay (EMSA, Figure 5-1b, right). SS-nanopore investigation of this material yields an increase in capture rate of more than an order of magnitude compared to an unlabeled control (Figure 5-1c). This observation is comparable to data collected with 156 bp dsDNA synthesized with a biotin at the same position (Appendix A.5). We attribute the minor quantitative disparity to a difference in pore diameter (Table 5-1).

Sample Label (dsDNA)	Fig.	Pore Diameter (nm)	Measurement Duration (s)
156 bp monobiotinylated DNA (385 nM)	1a&d	7.8	200
156 bp 5hmC labeled DNA (550 nM)	1c-d	8.5	200
Monovalent streptavidin (2.5 μM)	S1	8.5	520
75 bp quantification standard	3a	8.3	320
Genomic DNA NP1 (475 nM)	3b	9.3	512
Genomic DNA NP2 (1 μM)	3b	9.1	640
Genomic DNA NP3 (475 nM)	S7	8.6	806

Table 5-1 SS-Nanopore dimensions and measurement times. List of nanopore diameters, as determined by ionic resistance, and total time duration of measured traces for each data point in the presented measurements.

Collectively, these results demonstrate two crucial points about the assay: first, the biotin label residing near the oligonucleotide end has the same effect as a center label[13], showing that modification position is not a major factor; and second, the bulky (605 Da) glucose linker used here does not significantly affect detection viability, showing the adaptability of the assay.

Having established an ability to assess 5hmC specifically with SS-nanopores, we next apply our technique to quantify modification levels in mammalian DNA using a methodology that is shown schematically in Figure 5-2a.

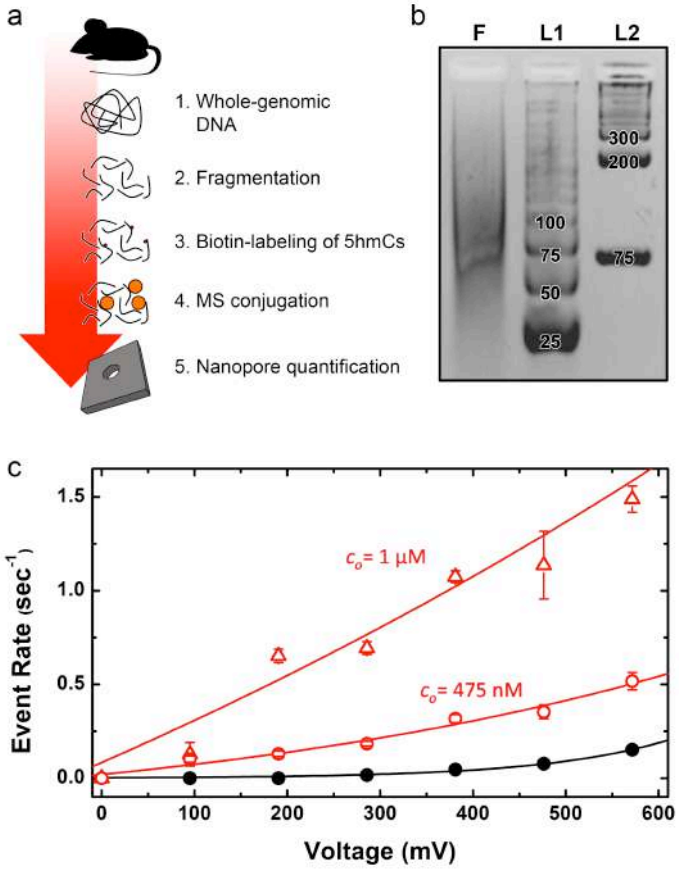

Figure 5-2 SS-Nanopore detection of 5hmC in mouse genomic DNA.

(a) Scheme showing preparation of genomic DNA for the SS-nanopore assay. (b) Gel electrophoresis of fragmented mouse genomic DNA ('F') and two low-molecular weight DNA ladders (L1 and L2, band labels are in bp). (c) Event rate vs. applied voltage measured for genomic DNA treated as described in (a) at total DNA concentrations c_o of 1 µM (red triangles) and 475 nM (red circles) in the presence of excess MS with 9.1 and 9.3 nm diameter nanopores, respectively. Black data points are a negative control of the same material (no MS). Solid lines are exponential fits to the data.

For this investigation, we focus on genomic DNA extracted from murine brain tissue because of the known abundance of 5hmC in functional elements of the brain[1] and its

suggested role in associated disorders[8,21]. The selectivity of our SS-nanopore approach requires short dsDNA because long molecules produce signals without MS binding[22]. Therefore, prior to assessment, the length of isolated whole genomic DNA must be reduced. For this, we utilize focused ultra-sonication to fragment the DNA to an average length of 75 bp (Figure 5-2b). We note that the variation in fragment length is approximately symmetric with respect to the mean; since the capture rate for our assay varies linearly with dsDNA length[22] from roughly 50 to 250 bp (Figure 5-3), SS-nanopore analysis is therefore representative of the mean substrate population. Fragmented DNA is subsequently subjected to the 5hmC biotinylation process, purified, and measured by SS-nanopore. We again observe a pronounced event rate enhancement for the labeled material in the presence of MS at two total DNA concentrations c_o (1 μM and 475 nM, Figure 5-2c) compared to the rate for a control measurement of fragmented DNA without MS.

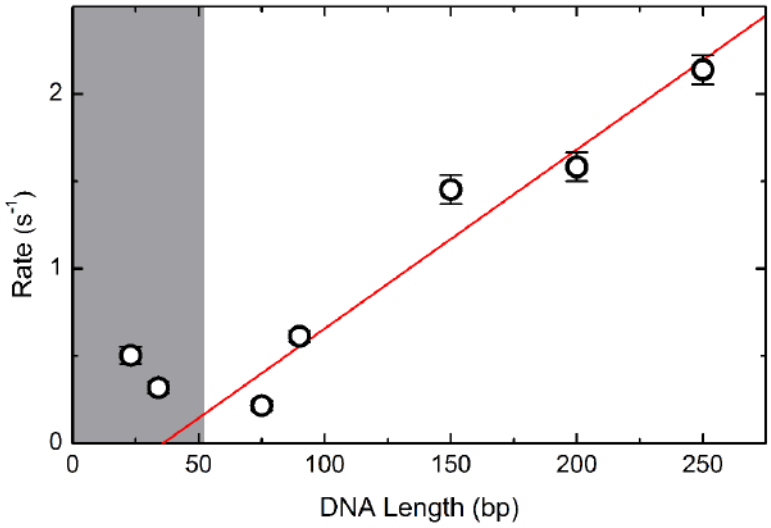

Figure 5-3 Event rate dependence of dsDNA length. Translocation event rate (200 mV) for synthetic monobiotinylated dsDNA constructs (1 μM) with bound MS as a

function of length. Solid line is a linear fit to data from 75 to 250 bp. Below about 50 bp (shaded region), an unexplained increase in rate is observed. This is the subject of further study but does not impact the present study, which focuses on constructs above this range. dsDNA without bound MS did not show any significant rate across the entire investigated range of lengths.

Determining hmC content in murine genomic DNA

To validate the measurement, we finally compare our results to quantification by the established technique of liquid chromatography-tandem mass spectrometry (LC-MS/MS). LC-MS/MS is a widely used and exquisitely sensitive analytical procedure that can be used to quantify 5hmC relative to unmodified cytosines (C)[23]. However, it is time consuming and relies on expensive and complex instrumentation, and so achieving similar results with a rapid, low cost system would be of tremendous benefit. To accomplish this with our SS-nanopore assay, we implement here a simple expression for the 5hmC/C ratio, *R*:

$$R = \frac{\alpha\left(c'/c_o\right)}{2Yl}, \tag{1}$$

where α is the fractional cytosine content of the genome, *c'* is the concentration of biotinylated fragments, *Y* is the biotin labeling yield, and *l* is the mean dsDNA fragment length in bp. For our experiments, we take α to be 0.21 based on the G+C content of the mouse genome[24] and *l* as 75 bp (Figure 5-2b). The total input DNA concentration, c_o, is determined by spectrophotometry. The yield, *Y*, for the labeling methodology has been reported[5] as high as 90%, but can vary based on experimental factors and therefore must be determined for each sample. For precision, we use here the LC-MS/MS analysis of

labeled and unlabeled DNA to find *Y* of 51% for our material. However, this variable can also be determined through less demanding approaches, including EMSA with simultaneously labeled synthetic oligonucleotides. Finally, the labeled fragment concentration, *c'*, can be determined through comparison of experimental SS-nanopore event rate data to a standard curve generated by measuring the rate of synthetic 75 bp monobiotinylated dsDNA through a similar pore (Figure 5-4a).

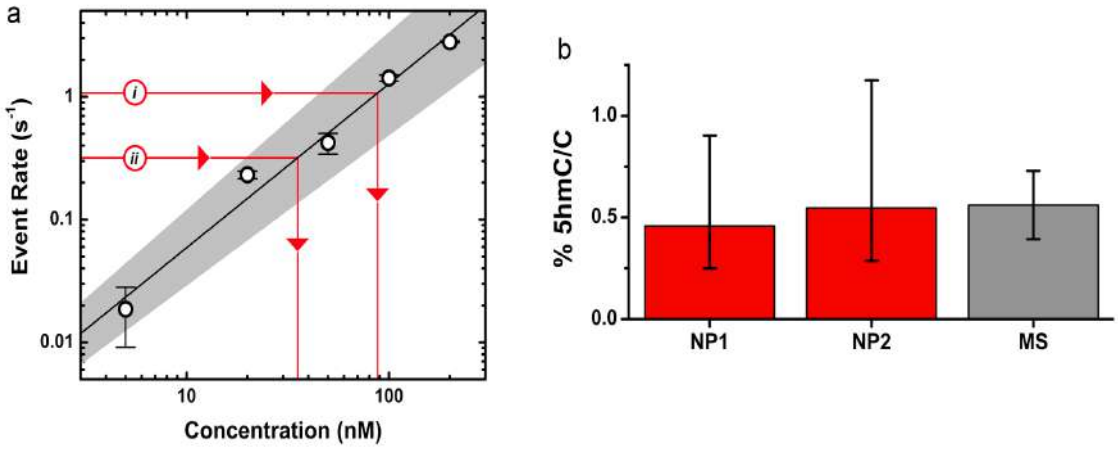

Figure 5-4 Quantification of 5hmC content by SS-Nanopore assay.

(a) Event rate vs. concentration (400 mV) for 75 bp monobiotinylated dsDNA with excess MS. Solid line is a linear fit to the data and shaded region represents error of fit. Solid red lines show experimentally-derived event rate values for mouse genomic DNA, biotin-labeled at 5hmC sites, and in the presence of excess MS, for *i*: NP1 (c_o = 1 μM, diameter 9.1 nm) and *ii*: NP2 (c_o = 475 nM, diameter 9.3 nm). (b) Quantification of 5hmC/C ratio *R* determined for the values in (a) and comparison with LC-MS/MS measurements of the same material. Error bars for nanopore measurements represent error of the fit from (a).

Figure 5-4b shows 5hmC/C quantification results from the two independent measurements of different total input DNA concentrations shown in Figure 5-2c. For the first device (NP1, c_o=1 μM), we find an R of 0.46 (+0.45, -0.21)%, while the second device (NP2, c_o=475 nM) yields an R of 0.55 (+0.63, -0.26)%. These values compare remarkably well to the ratio determined by LC-MS/MS of 0.56 ± 0.17%. The error in the SS-nanopore quantifications is taken as the error of the fit to the standard curve in Figure 3a. This is a conservative estimation that we consider as larger than the uncertainty in each of the variables from which R is derived. The error can be reduced with longer or additional measurements of synthetic standards to minimize uncertainty. We also note that our expression does not strictly quantify 5hmC relative to unmodified C, like LC-MS/MS, but rather to total (modified and unmodified) C content. However, the low cumulative levels of cytosine modifications, predominantly[6] 5mC, in mouse brain DNA would result in a difference of <2% in R; a value that is within the error of our measurements.

Our analysis assumes only that 5hmC is sparse- an argument that is supported generally by its established low abundance in genomic DNA (<1% in both mouse[25] and human[26] tissues). As a result, we expect that labeled DNA fragments will tend to feature either one or zero biotins. While CpG dinucleotides are known to often reside in close proximity to one another, forming "islands"[27] that may result in multiple 5hmC on a single fragment and potential undercounting, the characteristics of labeled genomic fragment events are similar to those of synthetic monobiotinylated 75 bp dsDNA ([appendix] Supplementary Fig. 6). This suggests that their respective conformations are

comparable and the quantitative agreement with LC-MS/MS further supports the validity of our assumption.

Discussion

SS-nanopores have been employed to detect epigenetic DNA modifications previously, through monitoring changes in translocation dynamics associated with modification-induced DNA structural differences[28,29] and through steric labeling of 5mC with methyl-binding proteins[30]. However, the minimum abundance of 5hmC necessary to produce resolvable structural variation in the former case (~3%) is considerably higher than physiological abundance and there are currently no known 5hmC-binding proteins for measurements analogous to the latter case. Additionally, neither previous approach is intrinsically selective, thus requiring significant analysis efforts and meticulous interpretation of the electrical signal for differentiation.

In response, we have combined a new SS-nanopore approach for detecting biotinylated oligonucleotides with an enzymatic technique for addition of biotin linkers to 5hmC sites, resulting in a novel assay capable of detecting and quantifying 5hmC content in DNA. First, we established the measurement with a synthetic dsDNA substrate (156 bp), demonstrating the detection of molecules containing a single 5hmC. We then used the approach to assess native modification abundance in genomic DNA derived from murine brain tissue. Comparison with the established technique of LC-MS/MS showed comparable quantification sensitivity. To our knowledge, this work represents the first measurement of a physiologically-relevant factor in mammalian genomic DNA using SS-nanopores. These results demonstrate the value of the platform in probing 5hmC

modifications with speed and precision, suggesting its possible future use in disease diagnostics and clinical screening. We anticipate that our approach can also be expanded to other modifications using alternative biotin-labeling strategies[31,32], offering potential modularity and targeting of additional biomarkers types.

5.2. Novel strategy to label epigenetic modifications of the demethylation pathway

(Unpublished results)

Intense research into DNA Epigenetic modifications have led to the discovery of 5hmC and its iterative oxidative forms, 5fC and 5caC. Tet family of dioxygenases oxidize 5hmC to 5fC and further to 5caC, these modifications are substrates for TDG, a glycosylase that cleaves these bases to generate an abasic site which is repaired by the BER. This cycle forms the basis of the demethylation pathway that results in removing any epigenetic marks from the cytosine base (Figure 5.2-1 a). Furthermore, its been established that while the levels of 5hmC, 5fC and 5caC are sparse they are stable and arise throughout the various genomic tissues. In order to probe these modifications only a handful of methods have been developed. All methods require BS treatment and sequencing in order to get relevant information, therefore they suffer from damaging the target DNA significantly. Therefore, novel methods are required that can be performed without BS treatment and can generate important biologically relevant results without conducting whole genome sequencing, similar to 5hmC-nano-seal (see Chapter 1.4.2).

We adapted a labeling technique recently developed in our lab (Figure 5.2- 1b) that labels damaged bases such as 8-Oxogaunine, Uracil, and Thymine: Guanine mismatches[38]. Our method utilizes glycosylase repair enzymes that selectively recognize these damaged bases and cleave them leaving an abasic site, subsequent treatment with an endonuclease primes the phosphate backbone to generate a 3' OH group[38,39]. A gap filling mutant T4 polymerase that lacks 3'-5' exonuclease activity (exo -) is introduced along with a 1biotinylated dNTP that can pair with the base present across the abasic site

(for eg biotinylated cytosine is used to treat a G:T mismatch). The technique is modular since we can address damaged bases, as selective glycosylases are available to repair them within the BER pathway enzymes.

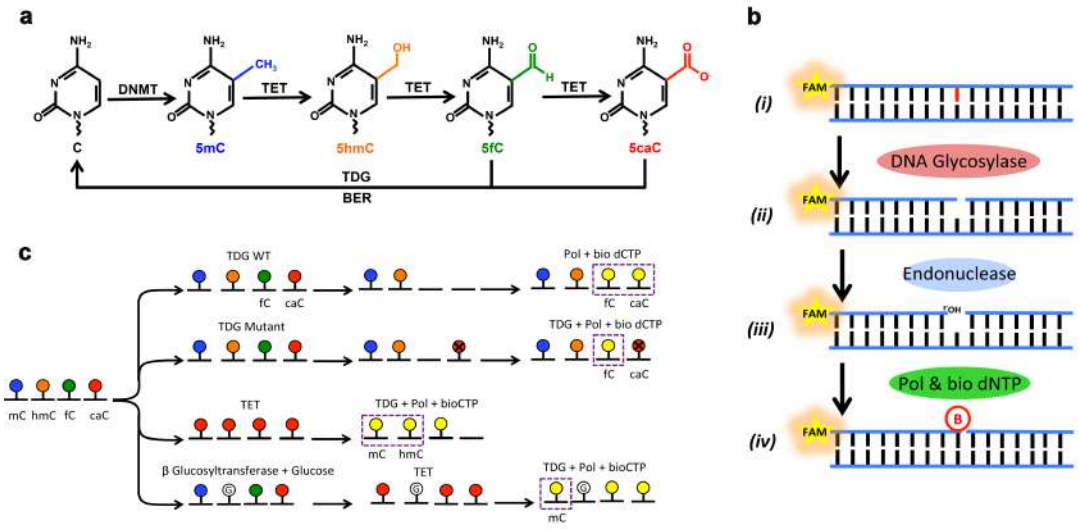

Figure 5-5 Demethylation pathway and labeling assay schemes (a) A schematic of the active Cytosine demethylation pathway. (b) A cartoon representation of the labeling protocol to replace damaged bases with a biotin tag, each step describes a treatment with a different enzyme. (c) A schematic demonstrating the adaption of the labeling scheme to target and label each Cytosine DNA epigenetic modification (blue, orange, green and red depict 5mC, 5hmC, 5fC and 5caC, respectively).

Recently we demonstrated selectively repair of U: G, 8-OxoG, T: G and 1,N^6-ethenoadenine, an analog of methyladenine using this approach. In the case of T: G mismatches we employed the Thymine DNA Glycosylase (TDG), interestingly TDG's substrate include fC and caC as well[40–42] , these bases are cleaved by TDG and are replaced with C forming the basis of the active DNA demethlylation pathway (Figure 5.2-1 a). We hypothesized that we could address all four Cytosine epigenetic modifications using our novel-labeling assay[38] with slight modifications. Since TDG

works on either 5fC or 5caC, essentially we would follow the same protocol developed to address T:G mismatches shown in Figure 5.2- 1b, i.e. (i) cleavage of 5fC or 5caC base leaving an abasic site (ii) introduce AP endonuclease 1 (APE1) to disengage TDG from the dsDNA molecule (iii) treatment with Endonuclease IV (ENDO IV) in order to cleave the phosphodiester backbone 5' to the abasic site, generating a 3' OH (iv) addition of a biotinylated cognate base by T4 polymerase (exo-). Its important to use a gap filling polymerase with no 3'-5' exonuclease activity otherwise the polymerase can possibly remove the biotin base or misincorporate it at a non-specific site We performed these labeling steps on FAM-labeled oligonucleotides containing either an 5fC or 5caC modification, Figure 5-6b shows when treated with TDG-WT, as expected the bases are cleaved off leaving a 22 nt strand (all modifications were designed to be the 23rd base). Subsequent treatment with T4 polymerase (exo-) in the presence of bio-dNTP, the band shifts up, indicating that the biotinylated base has replaced the epigenetic modification. Since TDG-WT acts on both 5fC and 5caC, we cannot differentiate between them. To distinguish 5fC and 5caC we utilized another form of TDG, a TDG N191A mutant[42] (TDG-102 Mut) that cleaves 5fC exclusively and is inactive on 5caC. We successfully demonstrated selective labeling of these two modifications in Figure 5-6b, TDG-WT works on both 5fC and 5caC as described above. As anticipated when the two substrates (5fC or 5caC) were treated with TDG-102 Mut, only the target DNA strand containing 5fC is labeled where as 5caC remains unaffected, as observed in Figure 5-6b. Conducting these two types of reactions in tandem on a single sample, would allow us to select for fragments that are 5fC or 5fC+5caC labeled. Subtracting the 5fC content from the sample labeled for both 5fC+5caC would give us information on the 5caC content.

To address 5mC and 5hmC, the modifications need to be converted to either 5fC or 5caC, in order to achieve this we employed one of the Tet family of enzymes that conduct step-wise oxidation of 5mC (or 5hmC) to 5fC and finally to 5caC. We utilized a truncated form of human Tet2 protein, the Tet2 crystal structure (Tet2-CS, residues 1129–1936) consisting the main regions that interact with DNA and are necessary for its catalytic activity[44,45]. We treated DNA substrates containing either 5mC or 5hmC by Tet2-CS and immediately subjected them to the labeling process with TDG-WT that allowed us to tag 5mC and 5hmC modifications as well (Figure 5-6a). In order to differentiate between the two, 5hmC substrate was pre-treated with β-GT and regular UDP-glucose, this covalently attaches UDP-glucose to 5hmC, thereby preventing its oxidation to 5caC by Tet2-CS. Therefore, any sample treated with β-GT will result in oxidation and labeling of 5mC while 5hmC will remain protected, as can be observed in Figure 5-6b (far right), no shift in the band is observed for 5hmC through every step of the labeling process.

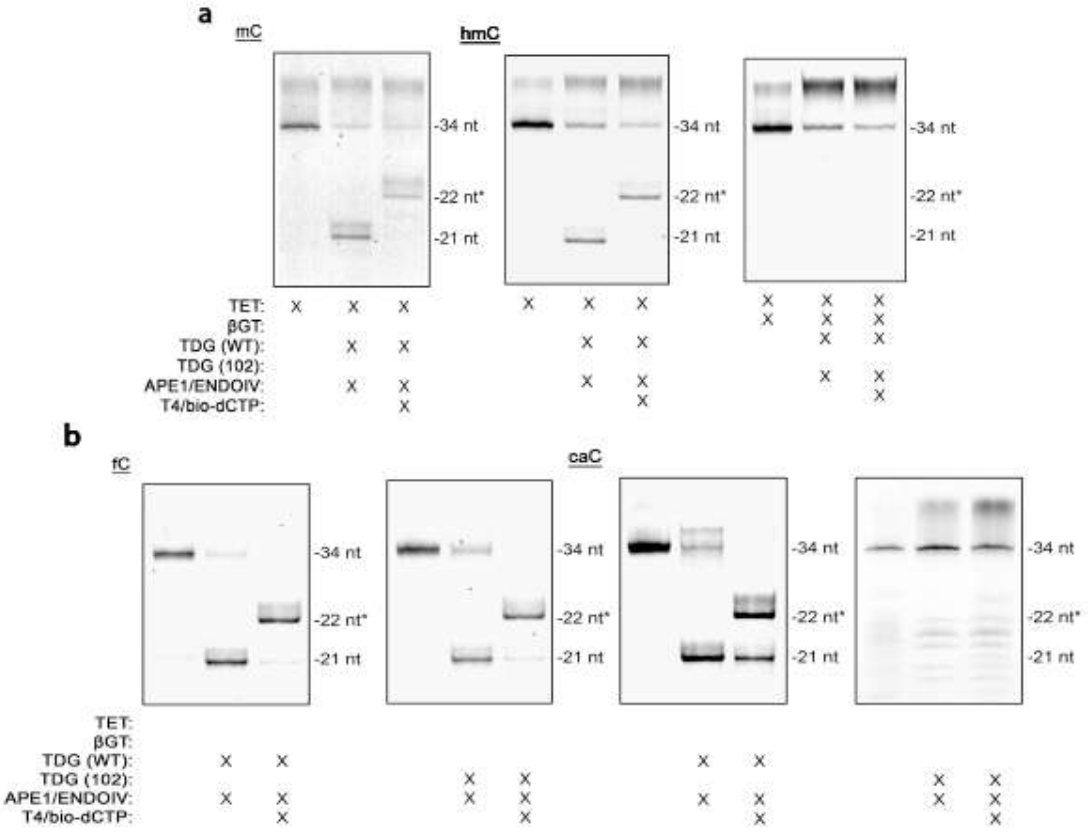

Figure 5-6 Denaturing gel analysis of demethylation pathway labeling assay.

(a) Denaturing gel results demonstrating labeling of 5mC and 5hmC base modifications. The hmC modification results are represented by two gels, the dataset demonstrating no shift of bands (far, right) indicates 5hmC modification was blocked by pretreatment with UDP-glucose. (b) 5fC and 5caC labeling results, wherein the first and second gel represents treatment of each modification with WT TDG or N191A MUT TDG (referred in the key as TDG 102), respectively. The key at the bottom of the gels represents the enzymes (not in order of treatment protocol) used in the whole labeling schematic, and the x refer to whether that particular enzyme was used in the treatment for that particular base modification.

We realize each labeling pathway (total of four) has to be conducted in tandem in order to get a comprehensive analysis of each type of modification. However, these proof-of-principle measurements demonstrate a novel epigenetic labeling assay capable of tagging all cytosine epigenetic modifications. Furthermore, since it's an affinity labeling technique, it should be able to focus on regions that contain the epigenetic modifications, unlike most other techniques like TAB-Seq, OxBS-Seq and redBS-Seq, that treat the entire genome and then have to conduct whole-genome sequencing. Our assay is essentially similar to nano-hmC-Seal the only biotin based affinity assay developed demonstrated ability to tag and get pertinent differential expression profiles of 5hmC gene body regions of genomic DNA[46]. The method was recently used to demonstrate selective hmC gene expression profiles from liquid biopsies of patients suffering from different forms of cancer. The authors also demonstrated the 5hmC gene content in liquid biopsies is identical to cancer genomic DNA isolated from tumor biopsies and can be used as a strong predictive biomarker for different types of cancers. However, the latter method only addresses 5hmC, whereas, ours demonstrates the potential to address all of the cytosine epigenetic modifications.

Methods

Biomolecule preparation Synthetic dsDNA oligonucleotides (156 bp or 75 bp) containing either a single 5hmC or biotin modification were prepared by PCR using a modified primer (Table 5-2) and genomic bacteriophage Lambda DNA (New England BioLabs, Ipswich, MA) as a template.

Oligonucleotide	Sequence (5'→ 3')
Biotinylated Reverse Primer	CAGTTGAGGATCCCCATAAT^{B}GCGGCTGTTTTCTG
5hmC Reverse Primer	CAGTTGAGGATCCCCATAATGhmCGGCTGTTTTCTG
156 bp Forward Primer	AACAACTGTTTCAGCCACTGCTTC
75 bp Forward Primer	GCAGCCGGACGTGAACGCGCAG
FAM mC	CAGTTGAGGATCCCCATAATGmCGGCTGTTTTCTG
FAM hmC	CAGTTGAGGATCCCCATAATGhmCGGCTGTTTTCTG
FAM fC	CAGTTGAGGATCCCCATAATGfCGGCTGTTTTCTG
FAM caC	CAGTTGAGGATCCCCATAATGcaCGGCTGTTTTCTG
Epi Complement	CAGAAAACAGCCGCATTATGGGGATCCTCAACTG

Table 5-2 Oligonucleotide sequences. Sequences used for PCR of synthetic DNA constructs using Lambda DNA as a template. T^{B} signifies biotinylated thymine and sequences used for epigenetic labeling protocol. Cytosine modifications are highlighted in red.

Following the PCR reaction, dsDNA products were purified using QIAquick purification kit (Qiagen, Inc., Valencia, CA) and eluted in pure deionized water (EMD Millipore, Billerica, MA). PCR products were quantified with a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA). All synthetic oligonucleotides were obtained commercially (Integrated DNA Technologies, Coralville, IA), suspended in deionized water to a stock concentration of 200 μM, and stored at -20° C prior to use. All PCR and annealing products were confirmed by gel electrophoresis, wherein samples were loaded on a 2.5% agarose gel prepared in 1X TBE with GelRed nucleic acid stain (Phenix

Research Products, Candler, NC) and imaged using a Gel Doc™ system (BioRad, Hercules, CA).

5hmC labeling reaction 5hmC sites were labeled following the protocol developed by Song *et al*. The labeling reaction was performed at total concentrations of 50 mM HEPES buffer (pH 7.9), 25 mM $MgCl_2$, 1mM DDT, 250 μM UDP-azide-gluocose, 0.5 μM of T4-βGT, and 100 ng/μL DNA (either synthetic hmC oligonucleotides or fragmented mouse genomic DNA) and incubated at 37° C for 2 hours. Subsequently, 350 μM sulfo-dibenzocyclooctyne-biotin was added to the solution and incubated at 37° C overnight. The labeled DNA samples were then purified using QIAquick purification kit and eluted in pure deionized water. Reaction components were taken from the Hydroxymethyl Collector™ Kit (Active Motif, Carlsbad, CA) except T4-βGT, purchased from New England BioLabs, and sulfo-dibenzocyclooctyne-biotin purchased from Sigma-Aldrich (St. Louis, MO).

Mouse genomic DNA preparation Mouse genomic DNA (1 μg/μL) extracted from brain tissue was obtained commercially (Zyagen, San Diego, CA) and stored at -20° C prior to use. Sample fragmentation was performed by suspending 10 μg of genomic DNA in pure deionized water to a final volume of 50 μL in a microTUBE AFA fiber snap-cap (Covaris, Woburn, MA) following by shearing with a Covaris S220 focused-ultrasonicator operated at the following settings: peak incident power = 175 W; duty factor = 10%; cycles per burst = 200; total treatment time = 280 s. Fragment length was confirmed by gel electrophoresis on 4% agarose gel as described above. For control

material, fragmented genomic DNA was purified immediately using QIAquick purification kit (Qiagen, Valencia, CA), eluted in pure deionized water, and stored at -20° C prior to nanopore measurement.

Electromobility shift assay (EMSA) Synthetic monobiotinylated or biotin-labeled mono-5hmC dsDNA (156 bp, 0.3 μM) was incubated in 1X PBS with 1.5 μM MS for 10 min at room temperature. Samples were subsequently loaded on a 2.5% agarose gel with GelRed nucleic acid stain for visualization and run in an ice bath to prevent excessive diffusion of bands during migration. Imaging was performed as described above.

Binding reaction incubation MS, a 54.5 kDa mutant streptavidin variant (SAe1D3) containing a single biotin-binding site, was supplied by the Howarth lab (Oxford University). For all experiments, 2.5 μM MS was incubated at molar excess with DNA target in 1X PBS buffer for 10-15 min at room temperature. After incubation the mixture was brought to a final salt concentration of 900 mM NaCl and 0.5X PBS.

Nanopore fabrication, detection and analysis 4.4 mm silicon chips containing a 25 nm thin free-standing SiN membrane were obtained commercially (Norcada, Inc., Alberta, Canada). In each membrane, an individual nanopore 7.8-9.5 nm in diameter was fabricated with an Orion Plus helium ion microscope (Carl Zeiss, Peabody, MA) using a technique described elsewhere. Following fabrication, chips were stored in 50% ethanol solution until use. Before measurement, a chip was rinsed with deionized water and ethanol, and dried by filtered air. After exposure by air plasma (30 W) for 2 min on each side, the chip was placed in a custom Ultem 1000 flow cell and measurement buffer was

added immediately to both sides. Voltage was applied and current measured through Ag/AgCl electrodes using an Axopatch 200B patch clamp amplifier (Molecular Devices, Sunnyvale, CA). Pores were verified by measuring linear IV characteristics with a steady baseline and resistance commensurate with intended diameter. Current traces were collected at a rate of 200 kHz with a 100 kHz four-pole Bessel filter and analyzed with custom software, with which an additional 25 kHz low-pass filter was applied to all data. An event was defined as having amplitude above a threshold of 4.5σ and duration between 12.5 to 2400 μs. Below the temporal resolution limit (c.f. Fig. 1d), event metrics may be skewed but can still be identified as events. Rate was determined by measuring uninterrupted current traces of identical time durations for each data set (See Supplementary Table 2 for details). Data was saved in increments of 3.2 s and the standard deviation of the rates was used as measurement error.

Liquid chromatography-tandem mass spectrometry DNA was boiled at 95° C for 5 min and digested by 1 U nuclease P1 (Wako Chemicals, Richmond, VA) in 25 μl of buffer containing 10 mM NH_4OAc (pH 5.3) at 42°C overnight. Next, NH_4HCO_3 (1 M, 3 μl) and 0.001 U venom phosphodiesterase (Sigma) were added and incubated at 37 °C for 2 h before a final treatment with 0.5 U alkaline phosphatase (Sigma) at 37 °C for 2 h. After the digestion steps, samples were diluted to 50 μl and filtered (0.22 μm pore size, 4 mm diameter; EMD Millipore), and 10 μl of the solution was injected into LC-MS/MS for analysis. Nucleosides were separated by reverse phase ultra-performance liquid chromatography on a C18 column (#927700-902; Agilent Technologies, Santa Clara, CA) at 35 °C with on-line mass spectrometry detection using an Agilent 6410 QQQ

triple-quadrupole LC mass spectrometer in positive electrospray ionization mode. Soluble phase A was dionized water with 0.1% formic acid and phase B was methanol with 0.1% formic acid. The nucleosides were quantified by using the nucleoside to base ion mass transitions of 228 to 112 (C) and 258 to 142 (5hmC). Quantification was performed by comparison to a curve obtained with pure nucleoside standards running on the same batch of samples. The 5hmC/C ratio was calculated based on the calibrated concentrations.

TDG Protein Expression from E. coli

We followed a protocol adapted by Liu et al.[47] from earlier work[43] with minor modifications. Expression plasmids for both human WT and mutant N191A TDG were transformed into BL21(DE3) cells and grown in 1 L LB broth at 37 °C. Once the cultures reached OD600 = 0.6, they were gradually cooled to 16 °C, induced with 0.25 mM IPTG and grown overnight. Cells were harvested by centrifugation, resuspended in 20 mL of TDG lysis buffer (50 mM sodium phosphate, pH 8.0, 300 mM NaCl, 25 mM imidazole) with protease inhibitors, and lysed by two passes through an EmulsiFlex-C5. The lysate was cleared by centrifugation at 20 000 g for 20 min, loaded onto a 1 mL column of HisPur cobalt resin (Fisher Scientific) equilibrated with TDG lysis buffer, and bound by two applications of the lysate to the column under gravity flow. The column was washed with 20 mL of TDG lysis buffer and subsequently eluted in 1 mL aliquots by a linear gradient of 100–500 mM imidazole. Elutions were analyzed by SDS-PAGE, and fractions containing the protein were pooled and dialyzed overnight at 4 °C against TDG storage buffer (20 mM HEPES, pH 7.5, 100 mM NaCl, 1 mM DTT, 0.5 mM EDTA, 1% v/v glycerol). Dialyzed proteins were concentrated using 10 kDa MWCO centrifugal spin

filter columns. The final TDG concentration was determined with the Bio-Rad Protein Assay, and aliquots were stored at –80 °C prior to use.

APE1 D308A Protein Expression

APE1 D308A plasmid (provided by the Demple Lab, Stony Brook University) was transformed into BL21*(DE3) cells and grown in 1 L LB broth at 37 °C. After bacterial cell cultures reached OD600 = 0.6, expression was induced with 0.5 mM isopropyl β-d-thiogalactopyranoside (IPTG). The cultures were then incubated for another 90 min before being harvested by centrifugation, resuspended in 50 mM HEPES-KOH (pH 7.5), 100 mM KCl, 1 mM EDTA, 0.1 mM DTT, and 10% (v/v) glycerol, and lysed by two passes through an EmulsiFlex-C5 (Avestin, Ottawa, Canada). The lysate was cleared by centrifugation at 20 000 g for 20 min, loaded onto a 15 mL SP Sepharose column (GE Healthcare, Pittsburgh, PA), and eluted with a linear gradient of 100–750 mM KCl and elutions were analyzed by SDS-PAGE. Fractions containing the protein were pooled and dialyzed overnight at 4 °C against APE1 storage buffer (50 mM HEPES-KOH (pH 7.5), 200 mM KCl, 1 mM EDTA, 0.1 mM DTT, 10% (v/v) glycerol) and concentrated using 10 kDa MWCO centrifugal spin filter columns (EMD Millipore, Billerica, MA). The final protein concentration was determined with the Bio-Rad Protein Assay (Bio-Rad), and aliquots were stored at –20 °C prior to use.

TET reactions in vitro.

Custom 34 base 5'-FAM labeled (Sequence: CAGTTGAGGATCCCCATAATGXGGCTGTTTTCTG, X = mC, hmC, fC or caC) was

annealed to a complementary sequence (containing an unmodified CpG across the target base substrate) by mixing both at an equimolar ratio, incubating in deionized water at 95 °C for 10 min, and cooling to room temperature over 1 h. A mixture of 55 nM duplexed DNA (containing either mC, hmC or hmC-UDP blocked) was suspended in a final buffer concentration of 50 mM HEPES, pH 6.5, 100 mM NaCl, 1 mM α-ketoglutarate, 1 mM DTT, and 2 mM sodium ascorbate was pre-warmed to 37 °C. Immediately before the reaction, fresh ammonium iron(II) sulfate (Sigma) was added to 75 μM and TET2-CS [cite] was added to a final concentration of 54 ng/μL in reaction volumes of ~60-70 μL. After incubation at 37 °C for 1 hour, the reactions were terminated by the addition of 2 units Proteinase K (New England Biolabs, Ipswich, MA) and incubated in the water bath for another 15 mins. Reaction products were purified using the QIAquick Nucleotide Removal Kit (Qiagen, Valencia, CA), eluted in LC–MS grade H_2O, quantified with NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA) and stored at -20 °C until further use.

Epigenetic Labeling

Custom 34 nt oligonucleotides with a 5' FAM (sequence: Add sequence X, where X is either fC or caC) was annealed to a complementary sequence by mixing both at an equimolar ratio, incubating in deionized water at 95 °C for 10 min, and cooling to room temperature over 1 h. Duplex DNA samples containing either Tet2-CS oxidized (mC, hmC and hmC-UDP blocked) or unmodified fC and caC substrates were incubated with WT hTDG[43], at a final concentration of 95ng/μL, 40 fg of APE1 (D308A mutant[48]), 1 μg of BSA, in 1× HEMN.1 buffer (20 mM HEPES (pH 7.3), 100 mM NaCl, 2.5 mM MgCl2,

0.2 mM EDTA) in 40 μL reaction volumes at 37 °C for 1 h. The mixture was purified with a QIAquick nucleotide removal kit. Then, 40 U of EndoIV and 1 μg of BSA were added to a total volume of 30 μL in 1× NEB2 buffer and incubated at 37 °C for 30 min. 1.5 nmol of biotinylated dCTP (PerkinElmer, Waltham, MA) and 0.12 U of T4(exo-) were added to a final volume of 40 μL in 1× NEB2 buffer, and the mixture was further incubated at 37 °C for 30 min. Finally, the mixture was subjected to a second purification to remove proteins and excess nucleotides. To demonstrate labeling selectivity between fC and caC, DNA substrate containing unmodified fC or caC were both treated with hTDG N191A[42] mutant instead of WT at identical buffer and protein concentration reaction conditions mentioned above.

Gel Electrophoresis

To prepare the denaturing gel, 70 mL of 23% gel matrix (22% acrylamide, 1% bis-acrylamide, 7 M urea in 1× tris/borate/EDTA (3:1:1) (TBE) buffer), 240 μL of 25% ammonium persulfate, and 42 μL of tetramethylethylenediamine were mixed thoroughly. The gel mixture was cast and allowed to polymerize for 30 min before running samples with dye in 1× TBE (3:1:1) at 55 W for 90 min. Labeling yields were approximated by measuring product band intensity relative to intermediates in the final lane using ImageJ analysis software[49].

Acknowledgements

The authors wish to thank the Howarth Lab (Oxford University) for supplying monovalent streptavidin. This work was supported by NIH grant 1R21CA193067. O.K.Z. acknowledges support from the Sigma Xi Grants-in-Aid of research program (Grant ID

No. 489441). A.R.H. acknowledges support from the 3M Non-tenured Faculty Award program.

Author Contributions

O.K.Z. performed all SS-nanopore measurements, analyzed data, and wrote the manuscript. B.S.Z and C.H. performed the LC-MS/MS measurements and made material contributions to the project. A.R.H. oversaw the project, analyzed data, and wrote the manuscript. All authors reviewed the manuscript.

Competing Interests

A.R.H declares the following financial interest. He is listed as inventor on a patent covering the presented SS-nanopore assay. The other authors declare no competing interests.

References

1. Kriaucionis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **324,** 929–930 (2009).

2. Tahiliani, M. *et al*. Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. *Science* **324,** 930–935 (2009).

3. Ficz, G. *et al*. Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature* **473,** 398–402 (2011).

4. Münzel, M. *et al*. Quantification of the Sixth DNA Base Hydroxymethylcytosine in the Brain. *Angew. Chem. Int. Ed.* **49,** 5375–5377 (2010).

5. Song, C. X. *et al*. Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nat. Biotechnol.* **29,** 68–72 (2011).

6. Lister, R. *et al*. Global Epigenomic Reconfiguration During Mammalian Brain Development. *Science* **341,** 1237905 (2013).

7. Szulwach, K. E. *et al*. 5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging. *Nat. Neurosci.* **14,** 1607–1616 (2011).

8. Villar-Menéndez, I. *et al*. Increased 5-Methylcytosine and Decreased 5-Hydroxymethylcytosine Levels are Associated with Reduced Striatal A2AR Levels in Huntington's Disease. *NeuroMolecular Med.* **15,** 295–309 (2013).

9. Jin, S. G. *et al*. 5-Hydroxymethylcytosine Is Strongly Depleted in Human Cancers but Its Levels Do Not Correlate with IDH1 Mutations. *Cancer Res.* **71,** 7360–7365 (2011).

10. Dekker, C. Solid-state nanopores. *Nat. Nanotechnol.* **2,** 209–215 (2007).

11. Miles, B. N. *et al*. Single molecule sensing with solid-state nanopores: novel materials, methods, and applications. *Chem. Soc. Rev*. **42,** 15–28 (2012).

12. Branton, D. *et al*. The potential and challenges of nanopore sequencing. *Nat. Biotechnol*. **26,** 1146–1153 (2008).

13. Carlsen, A. T., Zahid, O. K., Ruzicka, J. A., Taylor, E. W. & Hall, A. R. Selective Detection and Quantification of Modified DNA with Solid-State Nanopores. *Nano Lett*. **14,** 5488–5492 (2014).

14. Zahid, O. K., Wang, F., Ruzicka, J. A., Taylor, E. W. & Hall, A. R. Sequence-Specific Recognition of MicroRNAs and Other Short Nucleic Acids with Solid-State Nanopores. *Nano Lett*. (2016). doi:10.1021/acs.nanolett.6b00001

15. Fairhead, M., Krndija, D., Lowe, E. D. & Howarth, M. Plug-and-Play Pairing via Defined Divalent Streptavidins. *J. Mol. Biol*. **426,** 199–214 (2014).

16. Plesa, C. *et al*. Fast Translocation of Proteins through Solid State Nanopores. *Nano Lett*. **13,** 658–663 (2013).

17. Wanunu, M. *et al*. Rapid electronic detection of probe-specific microRNAs using thin nanopore sensors. *Nat. Nanotechnol*. **5,** 807–814 (2010).

18. Carson, S., Wilson, J., Aksimentiev, A. & Wanunu, M. Smooth DNA Transport through a Narrowed Pore Geometry. *Biophys. J*. **108,** 331a (2015).

19. Venta, K. *et al*. Differentiation of Short, Single-Stranded DNA Homopolymers in Solid-State Nanopores. *ACS Nano* **7,** 4629–4636 (2013).

20. Rosenstein, J. K., Wanunu, M., Merchant, C. A., Drndic, M. & Shepard, K. L. Integrated nanopore sensing platform with sub-microsecond temporal resolution. *Nat. Methods* **9,** 487–U112 (2012).

21. Kraus, T. F. J. *et al*. Low values of 5-hydroxymethylcytosine (5hmC), the 'sixth base,' are associated with anaplasia in human brain tumors. *Int. J. Cancer* **131,** 1577–1590 (2012).

22. Wanunu, M., Morrison, W., Rabin, Y., Grosberg, A. Y. & Meller, A. Electrostatic focusing of unlabelled DNA into nanoscale pores using a salt gradient. *Nat. Nanotechnol.* **5,** 160–165 (2010).

23. Ito, S. *et al*. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333,** 1300–1303 (2011).

24. Chinwalla, A. T. *et al*. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420,** 520–562 (2002).

25. Globisch, D. *et al*. Tissue Distribution of 5-Hydroxymethylcytosine and Search for Active Demethylation Intermediates. *PLoS ONE* **5,** e15367 (2010).

26. Li, W., Liu, M., Li, W. & Liu, M. Distribution of 5-Hydroxymethylcytosine in Different Human Tissues, Distribution of 5-Hydroxymethylcytosine in Different Human Tissues. *J. Nucleic Acids J. Nucleic Acids* **2011, 2011,** e870726 (2011).

27. Gardiner-Garden, M. & Frommer, M. CpG Islands in vertebrate genomes. *J. Mol. Biol.* **196,** 261–282 (1987).

28. Wanunu, M. *et al*. Discrimination of Methylcytosine from Hydroxymethylcytosine in DNA Molecules. *J. Am. Chem. Soc.* **133,** 486–492 (2011).

29. Carson, S., Wilson, J., Aksimentiev, A., Weigele, P. R. & Wanunu, M. Hydroxymethyluracil modifications enhance the flexibility and hydrophilicity of double-stranded DNA. *Nucleic Acids Res.* gkv1199 (2015). doi:10.1093/nar/gkv1199

30. Shim, J. *et al*. Detection and Quantification of Methylation in DNA using Solid-State Nanopores. *Sci. Rep*. **3,** (2013).

31. Zhang, L. *et al*. TET-mediated covalent labelling of 5-methylcytosine for its genome-wide detection and sequencing. *Nat. Commun*. **4,** 1517 (2013).

32. Song, C.-X. *et al*. Genome-wide Profiling of 5-Formylcytosine Reveals Its Roles in Epigenetic Priming. *Cell* **153,** 678–691 (2013).

33. Quantitative sequencing of 5-formylcytosine in DNA at single-base resolution. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4188980/. (Accessed: 28th October 2017)

34. Lu, X. *et al*. Chemical Modification-Assisted Bisulfite Sequencing (CAB-Seq) for 5-Carboxylcytosine Detection in DNA. *J. Am. Chem. Soc*. **135,** 9315–9317 (2013).

35. Guo, F. *et al*. Active and Passive Demethylation of Male and Female Pronuclear DNA in the Mammalian Zygote. *Cell Stem Cell* **15,** 447–459 (2014).

36. Single-Base Resolution Analysis of 5-Formyl and 5-Carboxyl Cytosine Reveals Promoter DNA Methylation Dynamics: Cell Reports. Available at: http://www.cell.com/cell-reports/fulltext/S2211-1247(15)00009-1?_returnURL=http%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS2211124715000091%3Fshowall%3Dtrue. (Accessed: 28th October 2017)

37. Single-base resolution analysis of active DNA demethylation using methylase-assisted bisulfite sequencing | Nature Biotechnology. Available at: https://www.nature.com/articles/nbt.3073. (Accessed: 28th October 2017)

38. Wang, F. *et al*. Solid-State Nanopore Analysis of Diverse DNA Base Modifications Using a Modular Enzymatic Labeling Process. *Nano Lett*. (2017). doi:10.1021/acs.nanolett.7b03911

39. Riedl, J., Ding, Y., Fleming, A. M. & Burrows, C. J. Identification of DNA lesions using a third base pair for amplification and nanopore sequencing. *Nat. Commun*. **6,** ncomms9807 (2015).

40. Maiti, A. & Drohat, A. C. Thymine DNA Glycosylase Can Rapidly Excise 5-Formylcytosine and 5-Carboxylcytosine: Potential implications for active demethylation of CpG sites. *J. Biol. Chem*. **286,** 35334–35338 (2011).

41. Zhang, L. *et al*. Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA. *Nat. Chem. Biol*. **8,** 328–330 (2012).

42. Divergent Mechanisms for Enzymatic Excision of 5-Formylcytosine and 5-Carboxylcytosine from DNA - Journal of the American Chemical Society (ACS Publications). Available at: http://pubs.acs.org/doi/abs/10.1021/ja406444x?journalCode=jacsat. (Accessed: 29th October 2017)

43. Excision of 5-Halogenated Uracils by Human Thymine DNA Glycosylase. Available at: http://www.jbc.org/content/282/38/27578.full. (Accessed: 29th October 2017)

44. Hu, L. *et al*. Crystal Structure of TET2-DNA Complex: Insight into TET-Mediated 5mC Oxidation. *Cell* **155,** 1545–1555 (2013).

45. Hu, L. *et al*. Structural insight into substrate preference for TET-mediated oxidation. *Nature* **527,** 118–122 (2015).

46. Han, D. *et al*. A Highly Sensitive and Robust Method for Genome-wide 5hmC Profiling of Rare Cell Populations. *Mol. Cell* **63,** 711–719 (2016).

47. Liu, M. Y. *et al*. Mutations along a TET2 active site scaffold stall oxidation at 5-hydroxymethylcytosine. *Nat. Chem. Biol*. **13,** 181–187 (2017).

48. Rapid Dissociation of Human Apurinic Endonuclease (Ape1) from Incised DNA Induced by Magnesium. Available at: http://www.jbc.org/content/273/46/30360.abstract. (Accessed: 29th October 2017)

49. Collins TJ. ImageJ for microscopy. *BioTechniques* **43,** 25–30 (2007).

50. Booth, M. J. *et al*. Quantitative Sequencing of 5-Methylcytosine and 5-Hydroxymethylcytosine at Single-Base Resolution. *Science* **336,** 934–937 (2012).

Appendix

A.1. Fabrication of Nanopore devices using Helium Ion Microscopy

Loading the Chips:

- For nanopore fabrication there is specific sample holder that needs to be used.
- Use gloves for the entire process. Place chips (pit side facing down) and making sure they are placed such that the pit side is in the center (therefore the etched window is in the center)
- Once the chips have been placed on the sample holder and the sample holder is all ready to be used, place the entire holder along with the chips into a plasma cleaner to remove any organic impurities. The plasma cleaner is located in the 'Sample Prep' room down the hall from the HIM room. When using the plasma cleaner, first place the sample holder into the cleaning chamber, turn on the power strip switch (which supplies power to the plasma cleaner and the vacuum pump). There are two gas cylinders (Argon and Oxygen, respectively) placed next to the cleaner that are connected to it. Turn on the main gas valve for the oxygen tank, make sure **NOT TO TOUCH** the external pressure gauges and valves that control the gas supply to the plasma cleaner, that are all at ideal settings. Press the vacuum button on the plasma cleaner, it will place the cleaning chamber under vacuum, watch the gauge and make sure it drops to around 300 mTorr, turn on the button for the oxygen supply, there will sudden increase in the gas pressure, turn the oxygen supply button off watch until the pressure gauge drops again down to around 200-300 mTorr and turn it the oxygen supply again, this is to purge the cleaning chamber and ensure it mainly contains oxygen and all atmospheric gas has been removed. Finally keep the oxygen supply button on, and you will notice the pressure gauge will

have a reading of 200-250 mTorr, as soon as the chamber reads 250 mTorr, set up the time duration that you would want treat the sample under plasma (ideally in the range os 3-5 mins) and turn on the RF power supply, you will notice the chamber will light with a 'bright white light' indicative of an oxygen plasma. Once the plasma treatment duration is over, the RF power supply will shut off automatically, turn off the oxygen supply button, turn off the vacuum pump button, and turn on the chamber vent button. Once the chamber is vented, you will be able to open the sample chamber and can remove the sample holder from the chamber. Turn off the vent button, turn off the plasma cleaner, close the oxygen valve and turn off the main power strip.

- Place the sample holder in the loading chamber. The sample holder base has grooves that match the instruments 'arm' on which the holder rests. This looks like 'Y', when placing the holder into the chamber on the arm be careful to before placing the holder that these align perfectly and place the holder gently onto the arm (prevent any quick movement/jostling of the holder) as it may mis-align the chips.
- Close the load chamber
- The load sample window on the instrument should already be up, click on 'load sample', next option is to clean the sample with plasma. Use short cycle, cleaning generally since the chips and the sample holder is clean post plasma treatment, however to remove any contaminants that may have been introduced during the time the sample holder was removed from the plasma cleaner and placed into the HIM.

Imaging & Fabrication:

- In the mean time that the sample is loading, we can upload the image of the sample holder, this is used as a reference to manipulate the stage.

• ‘**STAGE MOTION’** tab on the bottom right of the screen (click), there is a tiny tab ‘A square tab with a Circle in the center’ select that. It brings up the window and in there choose the ‘load image’ tab, we need to browse through the Adam H folder and choose the folder ‘Holder images’, choose the sample image ‘holder_top.jpg’. Align image.

• Once the sample is done loading, we need to manipulate the stage and bring it close enough to the aperture. For that we need to move the stage in the Z axis therefore make sure that only the Z axis box is checked while the others are unchecked (X, Y, R, & tilt). Also make sure ‘Relative’ is unchecked when you enter the number (mm) you want the stage to move to. For this particular holder 26-28 mm is ideal. (Can also ‘STOP’ the stage motion at any time to prevent crashing of the sample holder into the detector/aperture.

• At this stage the trimer should be formed and therefore we can go in straight ahead and start the fabrication process. Make sure the instrument is operational:

i) ‘System Operating Mode’ → drop down menu & click on ‘Operating’

ii) Check everything is operational: click on all the tabs on the bottom left hand side of the program → Green on the “Column Control”

iii) Bottom right hand side → choose ‘Primary’ see everything is functional, ‘Control tab’ at top & “detector”.

iv) Check the trimer → to see that if the trimer is fine: Click on ‘Column Control’ tab → Mode → SFIM allows to observe/image trimer. Similarly in the ‘Basic Controls’ tab → view mode → choose either ‘Sample’ or ‘Source’ → source will allow to observe trimer. If the ‘cross-hair’ is not perfectly aligned to the center of the brightest atom. Turn on the

'Beam Tilt' option on the main panel and use the two knobs to align the cross hair to the center of the atom. **USE ONLY FOR MINOR ADJUSTMENTS**

- Pause the beam and check if the Blanker Current is at least 5 – 5.5 pA. This is the ideal current we need for milling/pore fabrication. If the current is lower than that, in order to increase it:

i) Under **'BASIC CONTROL'** Select the 'Spot Control' drop down tab, reduce the number and that will increase the 'Blanker Current' **NEVER REDUCE THE SPOT CTRL number TO 3 or BELOW**

ii) Can also increase the 'He ion pressure' make sure the exponent is correct (in the -7 to -6 range)

- Focus on a tiny feature on the metal surface, ideally a feature that is extruding out of the surface and is also 'circular' (helps in adjusting the beam stigmation):

i. Under Basic Controls, select 'Enable Wobble', make sure the number is + 0.001. Click on the 'Beam Align' tab on the main panel and Use X & Y stigmators dials to adjust image formation (technically adjusting the shape of the beam). Adjust only in one axis X or Y and then move on the next axis until the image that's formed makes it look like the feature is only pulsing and not diagonally moving across in either direction. Turn off 'Beam align' and 'Enable Wobble'

ii. Align Stigmators once again focusing on a feature (zoomed in) → adjust stigmation by X then Y dials and then use Focus to adjust focus (multiple times, at least twice) to make sure the beam is perfectly aligned and circular and the focus is perfect too.

Patterning:

• Column Control → FOV (field of view) to bring up the window to adjust the focus which in turn aligns it with the image pixels use for patterning.

• Before focusing on the chip, try on the metal surface, click in the FOV tab and zoom into the metal surface to 400nm by typing in 0.4. On the bottom left of the of the user interface app there are an array of functions, click drop down menu of the first tab, and select 'Spot Mode'. Leave it for 5 to 6 secs, it should form a circular pit (etches material away), click back on 'Continuous' tab and try to get best focus using the focus knob on the main panel.

i. Pattern tab on the bottom left of the program screen. Click on pattern type, browse and choose 'FS_modified – 5x5 array', repeat should be 1 (that is beam raster), increase the dwell time to 1000 usec, go to FOV (Field of View) 0.512nm, (a) Pattern, then reduce the rate of image capture slightly (i.e. increases beam exposure time) (1 or 2 clicks clockwise), click the grab tab on the control panel of the instrument, to take an image. The patterns should look circular if they aren't, move over to a new spot on the surface by double clicking on the edge of the screen couple of times to move the stage over. Adjust the focus by a few clicks either (clockwise or counter clockwise0 count the number of clicks to know exactly how much you've adjusted, the pattern may look worse and therefore you might have to either go back to original spot or readjust from the there again. Repeat steps from (i) (a) until the patterns look circular.

ii. Move on the chip surface, find the SiN window (should appear as a tiny dark black square), repeat earlier mentioned steps along the edge of the window in (i) (make sure you are far enough from the window so as to not accidently expose the beam to it and thereby rendering the device useless) and once the patterned circles appear to be

round and ideal for pore fabrication choose the pattern type for fabrication of the pore. That is either ‘single shot fabrication’ method, that refers to pattern ‘One dot-One pixel’ in the ‘Patterns folder’ or if you’d like to make a nanopore with multiple exposures by rastering the beam over a region multiple times in discrete steps until total desired dose is reached to fabricate a nanopore i.e. commonly referred to as ‘Fill fabrication’ method. In that case depending on the desired nanopore diameter, go into the ‘Pattern folder’ and choose either: ‘6nm Fill’, ‘8nm Fill’, ‘10nm Fill’ or any other desired nanopore diameter pattern.

iii. Calculate out the dose for the desired diameter using the ‘calibration curve’ for that particular chip batch. Once the pattern has been loaded (filled circle or single shot pattern) depending on that adjust the ‘number of repeats’ or the ‘dwell time’ to match/obtain the total calculated dose. For example, in the case of ‘single shot’ fabrication mode, your exposure repeats should be 1 and in that case adjust the dwell time to achieve the desired total dose you’d like to use. For the ‘Fill’ method, switch the repeat number to 10, and then adjust the dwell time to reach the total desired dose.

iv. Zoom out from the surface to about 20uM using FOV and then move the cross hair into the center of the chip window. Use the fastest rate to ‘grab image’ to make sure you are in the center of the chip window → Zoom into 0.512 uM using the FOV and then click on pattern.

v. Move on to the next chip and repeat the steps from step i for patterning

Unload the samples once the fabrication is complete. Select ‘Sample’ tab on the top left of the screen, it should only you to choose ‘Unload Sample’, click on that and you will

have to wait until the sample holder is unloaded. Go ahead and remove the sample from the sample exchange chamber. Once you are done ***make sure sample load-dock is under vacuum*** at the end, the program will prompt to ask **'Leave Load Dock Vented'** or **'Pump Load Dock'** choose **'Pump Load Dock'** even if you have to continue using the instrument for fabricating another batch of pores. The time duration for you to unload your fabricated chips and then remount another set of devices will take considerable time. That can cause the instrument to be in a state where it takes a significant period of time to pump out the load dock (for you second set of samples >20-30mins) to reach to a pressure level where it can actually interact with the main instrument chamber (at extremely low pressure) to drop off your sample holder. Furthermore, if you forget to Pump the load dock it can introduce impurities into the instrument chambers that can affect the instrument performance. **Finally**, make sure ***the 'System Operating' mode is in Standby*** before you leave the room.

A.2. Chapter 2 - Interpreting the Conductance Blockades of DNA Translocations Through Solid-State Nanopores

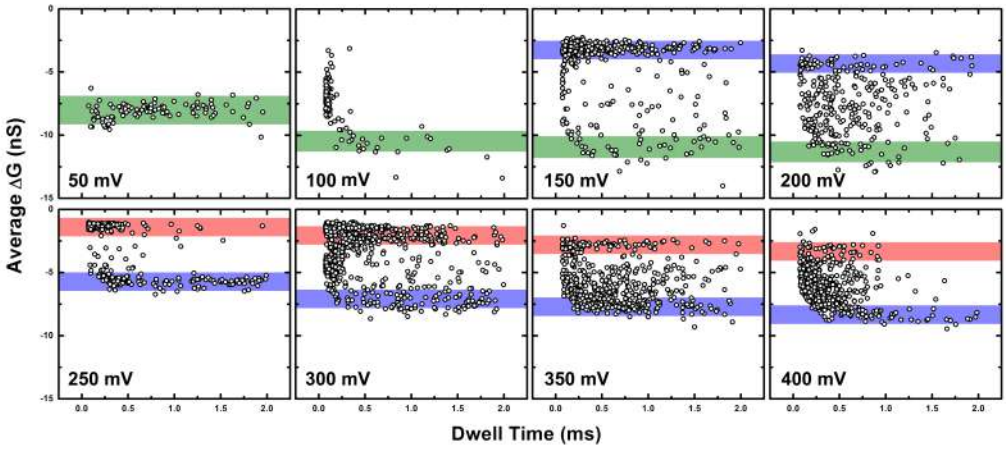

Figure A2- 1 Scatter plots of blockade events. Event depth (ΔG) vs. dwell time (Δt) for all recorded events at each voltage (same data as shown in Fig. 1c of the main text). Shaded regions correspond to the three interaction types described in the text, with the same color scheme as in Figs. 1 and 3. Events consisting of multiple levels reside between the shaded regions.

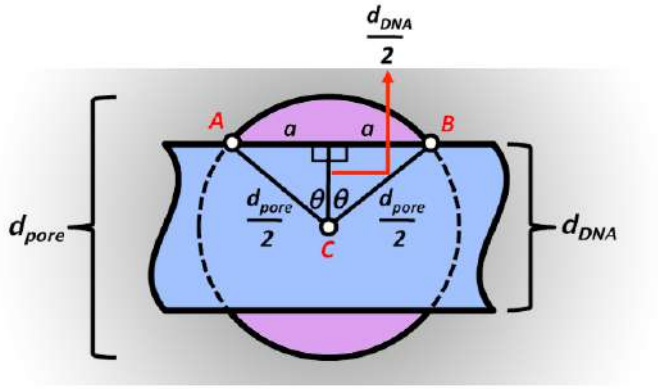

Figure A2- 2 Geometric model of lateral blocking of SS-nanopore. Schematic of dsDNA (blue) laying laterally across a SS-nanopore (area enclosed by black circle). The areas shaded in purple represent the remaining accessible area of the blocked nanopore.

S2. Geometric expression of DNA blocking a SS-nanopore laterally The unoccluded area of the laterally-blocked SS-nanopore is composed of two segments of the circular pore, defined by the intersection of the dsDNA with the pore circumference (purple shaded regions in Fig. S-4). The area of one such segment, A_{seg}, can be expressed geometrically as

$$A_{seg} = A_{arc} - 2A_T, \qquad \text{Eq. S-1}$$

where A_{arc} is the area of the sector ABC and A_T is the area of one of the two right triangles defined by the same three points (blue regions in Fig. S-4). In terms of the known experimental quantities (nanopore diameter, d_p, and dsDNA diameter, d_{DNA}), A_{arc} can be written as

$$A_{arc} = \frac{d_p^2}{4}\theta = \frac{d_p^2}{4}\left(\cos^{-1}\left(\frac{d_{DNA}}{d_p}\right)\right). \qquad \text{Eq. S-2}$$

Line segment a at the base of the two right triangles is defined as $\frac{1}{2}(d_p^2 - d_{DNA}^2)^{\frac{1}{2}}$, and so A_T can be expressed as

$$A_T = \frac{1}{8} d_{DNA}\left(d_p^2 - d_{DNA}^2\right)^{1/2}. \qquad \text{Eq. S-3}$$

Treating the unoccluded area of the laterally blocked pore as a circular region of area A_p^* and diameter d_p^*, we can write

$$A_p^* = \frac{\pi}{4} d_p^{*2} = 2A_{seg} = \frac{d_p^2}{2}\cos^{-1}\left(\frac{d_{DNA}}{d_p}\right) - \frac{1}{2} d_{DNA}\left(d_p^2 - d_{DNA}^2\right)^{1/2}, \qquad \text{Eq. S-4}$$

and therefore

$$d_p^* = \sqrt{\frac{2}{\pi}\left(d_p^2 \cos^{-1}\left(\frac{d_{DNA}}{d_p}\right) - d_{DNA}\left(d_p^2 - d_{DNA}^2\right)^{1/2}\right)}. \qquad \text{Eq. S-5}$$

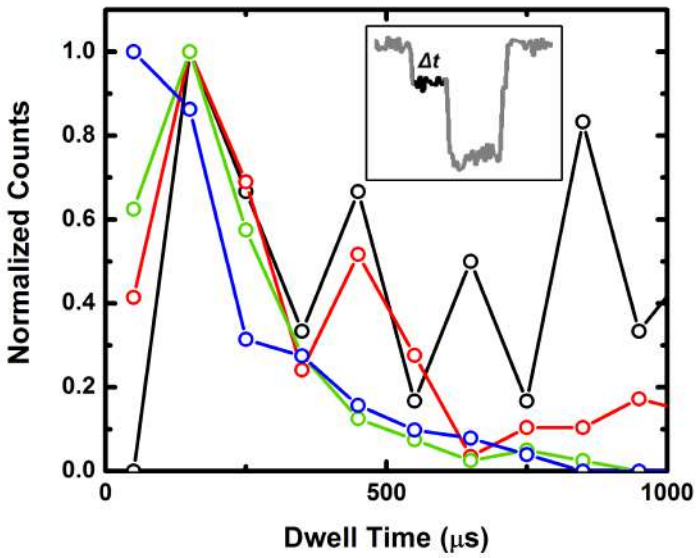

Figure A2- 3 Shallow-level dwell time distributions for two-level events. Normalized dwell time distributions for the first (non-translocative) level in two-level events measured at 250 (black), 300 (red), 350 (green) and 400 mV (blue), respectively. As applied voltage rises, the dwell time distribution narrows and the mean dwell time is reduced. Inset highlights the section considered (black region in trace).

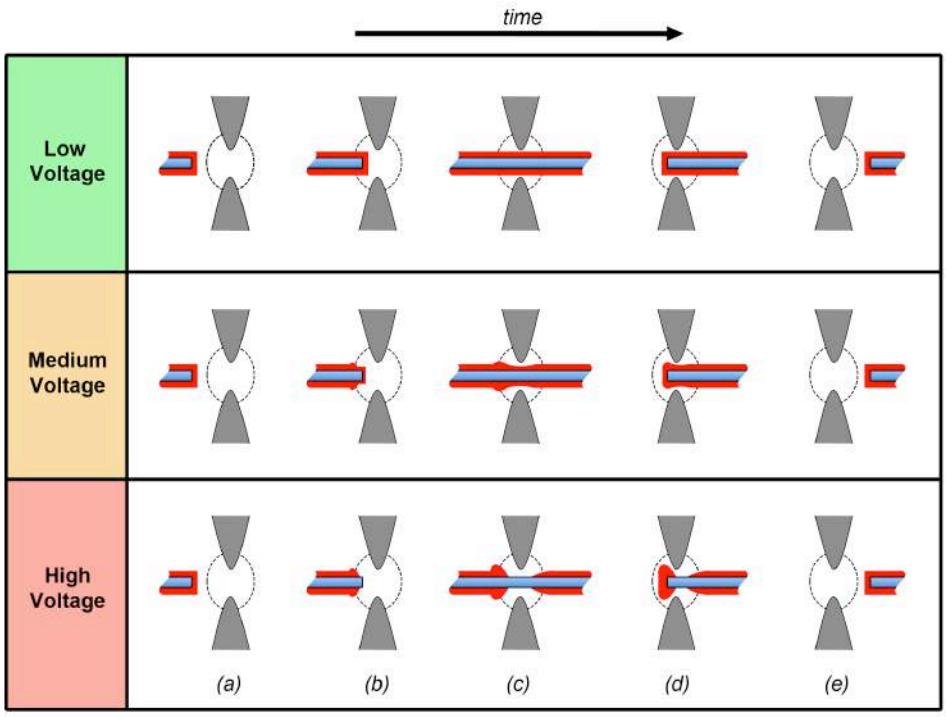

Figure A2- 4 Proposed mechanism of local counterion depletion by field polarization. Schematic representation of proposed counterion configuration (red) surrounding DNA (blue) during translocation at low, medium, and high voltage (i.e. electric field). DNA is shown sequentially before entry (a), during entry (b), fully threaded (c), exiting (d) and fully ejected (e). As field strength increases, polarization causes depletion of counterions *local to the sensing region* (dashed lines); the buildup of charge outside this region preserves electroneutrality in the system as a whole. We note that this model suggests a significant increase in counterion density at the trailing end of the DNA (d), but this effect would be difficult to resolve experimentally due to temporal limitations[8].

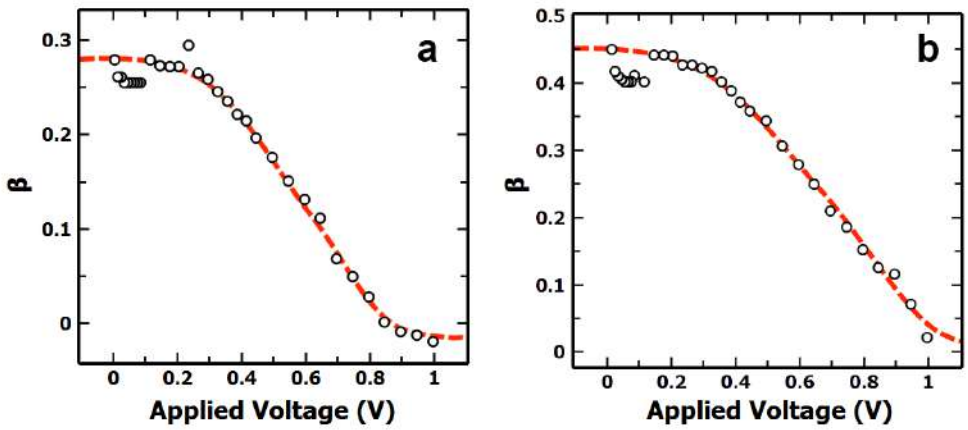

Figure A2- 5 Estimation of local counterion density on dsDNA at zero field. Fractional counterion density β vs. applied voltage for both the 1 M KCl data (a) and the 1 M NaCl data (b) from Supplementary Reference 1. β is adjusted in our model to account for the shifting ΔG in the data, yielding a sigmoidal relation. The low-voltage limit appears to fall at about 0.28 for 1 M KCl and about 0.45 for 1 M NaCl. The latter is near the NaCl data from the main text (Fig. 4), which yields a limit of roughly 0.57. The deviation may result from lower ionic strength used in our experiments (900 mM) compared to these data. Dashed lines are Boltzmann sigmoidal fits.

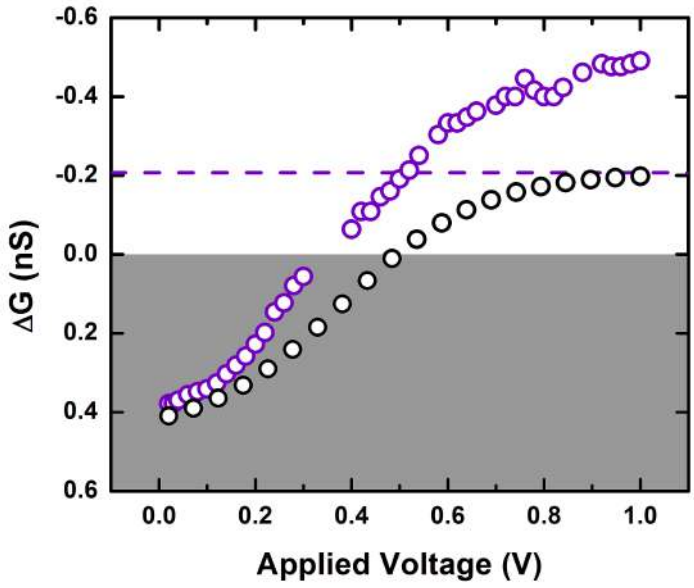

Figure A2- 6 Application of model interpretation to measurements in low ionic strength solvent. Purple circles are data for 0.1 M KCl from Supplementary Reference 1. The shaded region represents conductance *increases*. The dashed line is the value predicted from our model for β=0. Black circles represent expectations from the model with β ranging from 0.22 to 0. Adjusting β with applied voltage based on a Boltzmann sigmoidal fit to the experimental data, we demonstrate that, in spite of quantitative disagreement with the experimental saturation level (~0.5 nS), the model captures a major qualitative feature of the low-ionic strength data: a switch from conductance blockades to enhancements as voltage is reduced.

Supplementary References

1. Kasianowicz, J. J.; Brandin, E.; Branton, D.; Deamer, D. W. Characterization of Individual Polynucleotide Molecules Using a Membrane Channel. *Proc. Natl. Acad. Sci. U. S. A.* 1996, 93, 13770-13773.

2. Wanunu, M.; Sutin, J.; McNally, B.; Chow, A.; Meller, A. DNA Translocation Governed By Interactions with Solid-State Nanopores. *Biophys. J.* 2008, 95, 4716-4725.

3. Kowalczyk, S. W.; Dekker, C. Measurement of the Docking Time of a DNA Molecule Onto a Solid-State Nanopore. *Nano Lett.* 2012, 12, 4159-4163.

4. Kowalczyk, S. W.; Dekker, C. Salt and Voltage Dependence of the Conductance Blockade Induced By Translocation of DNA and RecA Filaments Through Solid-State Nanopores. In *Nanopores for Bioanalytical Applications*, Edel, J.; Albrecht, T., Eds. Royal Society of Chemistry: 2012.

5. Skinner, G. M.; van den Hout, M.; Broekmans, O.; Dekker, C.; Dekker, N. H. Distinguishing Single- and Double-Stranded Nucleic Acid Molecules Using Solid-State Nanopores. *Nano Lett.* 2009, 9, 2953-2960.

6. Yang, J.; Ferranti, D. C.; Stern, L. A.; Sanford, C. A.; Huang, J.; Ren, Z.; Qin, L.-C.; Hall, A. R. Rapid and Precise Scanning Helium Ion Microscope Milling of Solid-State Nanopores For Biomolecule Detection. *Nanotechnol.* 2011, 22, 285310.

7. Wanunu, M.; Dadosh, T.; Ray, V.; Jin, J.; McReynolds, L.; Drndic, M. Rapid Electronic Detection of Probe-Specific MicroRNAs Using Thin Nanopore Sensors. *Nat. Nanotechnol.* 2010, 5, 807-814.

8. Das, S.; Dubsky, P.; van den Berg, A.; Eijkel, J. C. T. Concentration Polarization in Translocation of DNA Through Nanopores and Nanochannels. *Phys. Rev. Lett.* 2012, 108, 138101.

9. van den Hout, M.; Krudde, V.; Janssen, X. J. A.; Dekker, N. H. Distinguishable Populations Report On the Interactions of Single DNA Molecules With Solid-State Nanopores. *Biophysical Journal* 2010, 99, 3840-3848.

A.3. Chapter 3 - Selective detection and quantification of modified DNA with solid-state nanopores

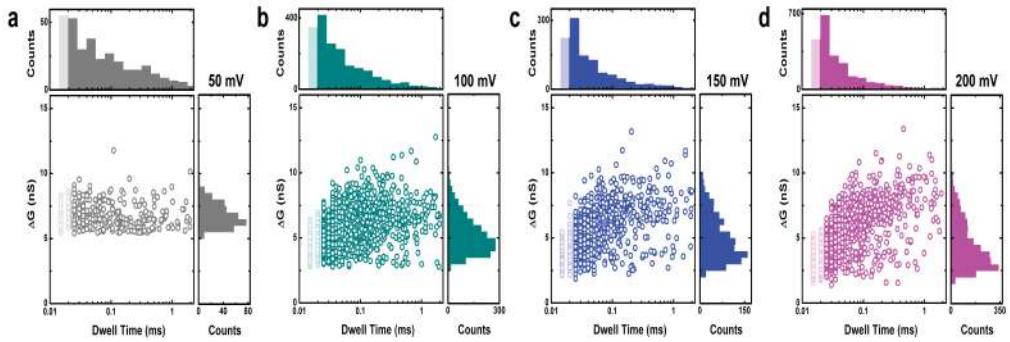

Figure A3- 1 Voltage dependence of scatter plots. Conductance blockage vs. dwell time for events measured at a stoichiometry of 1:1 (MS:bio90) and across a range of applied voltage (indicated at upper right of each plot). Total numbers of events considered are 332 (a), 1852 (b), 1129 (c), and 2193 (d). The faded population at the left of each plot represents events with duration below the resolution limit (25 μs). The Δ*G* histogram profile change may indicate multiple conformatios of the MS-bio90 construct during translocation at higher voltage.

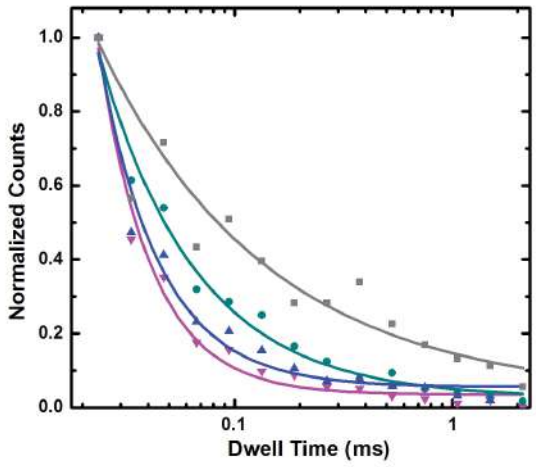

Figure A3- 2 Voltage dependence of MS-bio90 translocations. The normalized abundance of MS-bio90 dwell times, plotted from the resolution limit of 25 μs up to 2 ms

(from the upper histograms in Supplementary Figure 2). Data displayed (T-B) are for applied voltages of 50 mV (grey squares), 100 mV (cyan circles), 150 mV (blue upward triangles), and 200 mV (magenta downward triangles), respectively. Solid lines are exponential fits to the (log-normal) data, demonstrating that event durations decay faster as applied voltage increase.

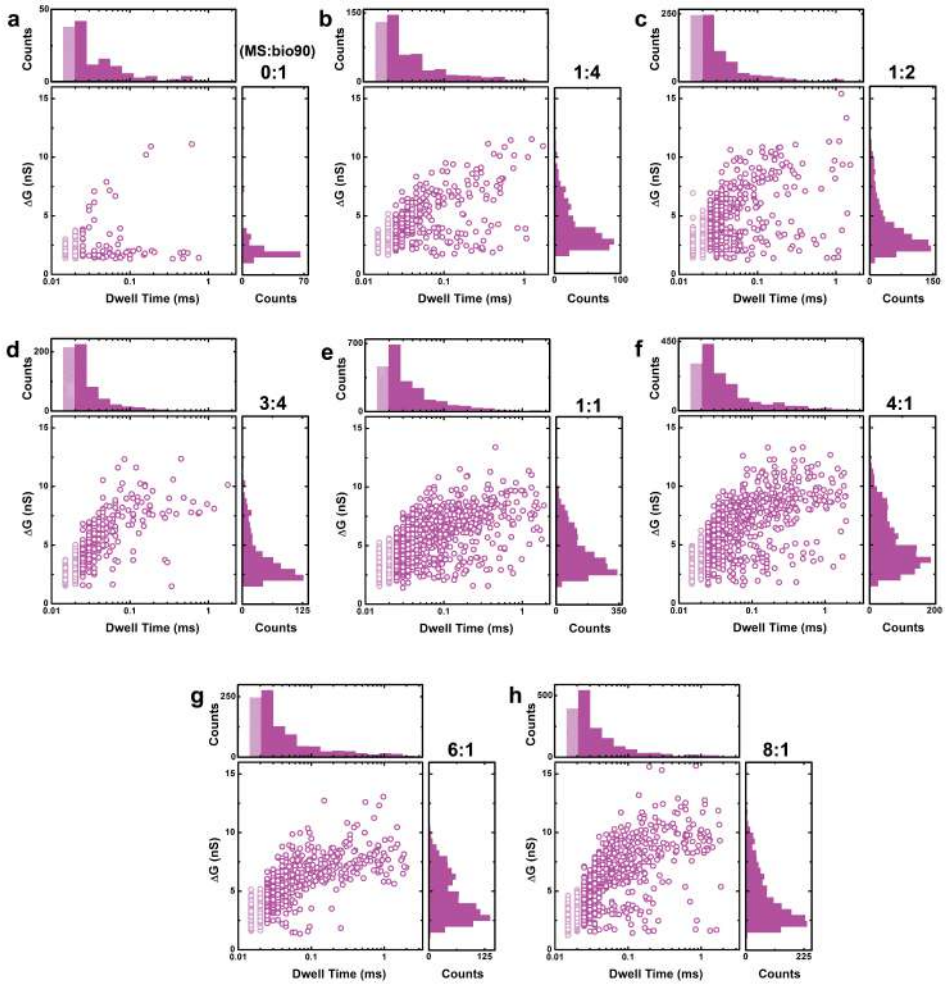

Figure A3- 3 Stoichiometry dependence of scatter plots. Conductance blockage vs. dwell time for events measured at an applied bias of 200 mV and across a range of stoichiometries (indicated at upper right of each plot). Total numbers of events considered are 139 (a), 521 (b), 816 (c), 642 (d), 2193 (e), 1528 (f), 982 (g), and 1652 (h).

The faded population at the left of each plot represents events with duration below the resolution limit (25 μs).

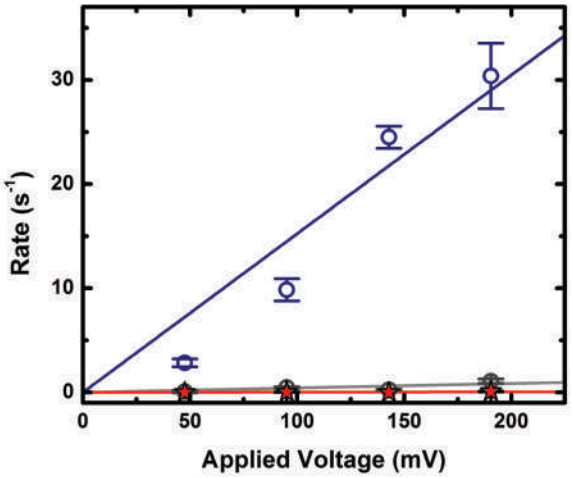

Figure A3- 4 Control measurements on non-biotinylated dsDNA. Measured event rate vs. applied voltage for MS mixed with non-biotinylated 95 bp dsDNA at a molar ratio (MS:DNA) of 4:1 (red stars). The mixture of MS and non-biotinylated DNA yields an extremely low event rate, more than an order of magnitude lower than bio90 mixed with MS at the same ratio (blue) and equivalent to that of bio90 with no added MS (grey). Solid lines are linear fits to the data.

A.4. Chapter 4 - Sequence-specific recognition of microRNAs and other short nucleic acids with solid-states nanopores

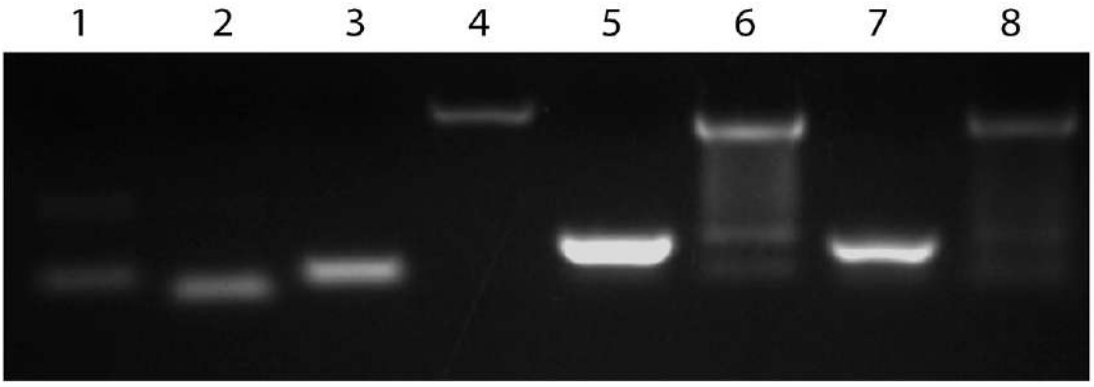

Figure A4- 1 miRNA annealing gel. Agarose gel electrophoresis showing miR155 (lane 1), 23 nt miR155 DNA homolog (lane 2), ssBio23 (lane 3), ssBio23+MS (lane 4), annealed ssBio23-miR155 heteroduplex (lane 5), annealed ssBio23-miR155+MS (lane 6), ssBio23 annealed to miR155 DNA homolog (lane 7), and ssBio23 annealed to miR155 DNA homolog +MS (lane 8).

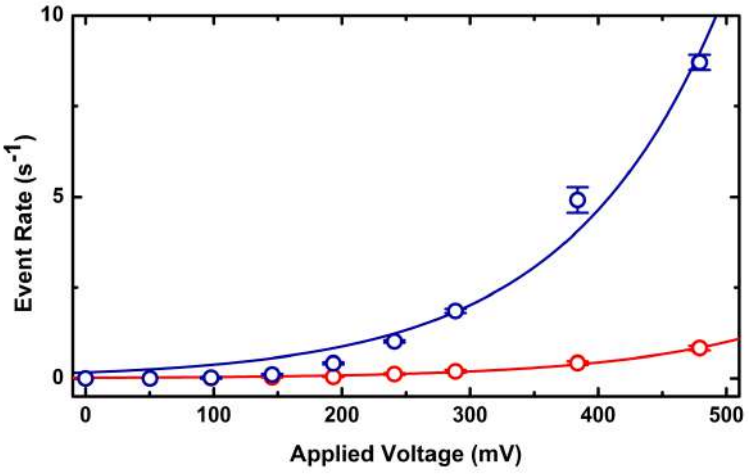

Figure A4- 2 Additional DNA data set. Event rate vs. applied voltage for ssBio34 (red) and dsBio34 (blue) with MS bound (1 μM) on a second SS-nanopore. All solid lines are exponential fits to the data.

A.5. Quantifying mammalian genomic DNA hydroxymethylcytosine content using solid-state nanopores

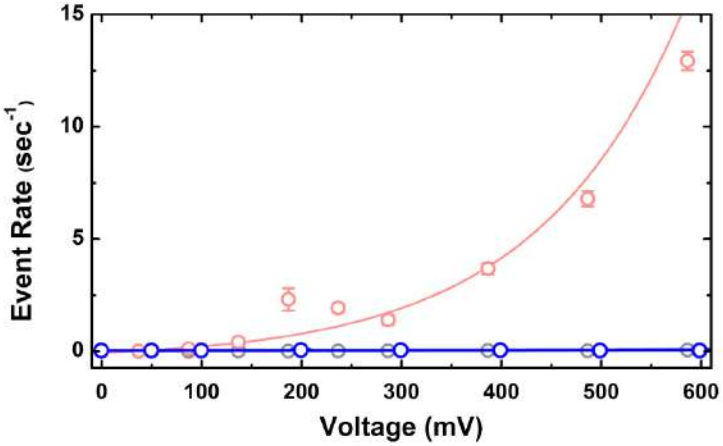

Figure A5- 1 MS Event rate. SS-nanopore event rate for monovalent streptavidin alone (blue, 2.5 μM) measured with a 8.5 nm diameter pore. The background displays rates for biotin-labeled 156 bp mono-5hmC dsDNA (550 nM) both with (red) and without (gray) bound MS for comparison, identical to Fig. 1c from the main text. Solid lines are exponential fits to the data.

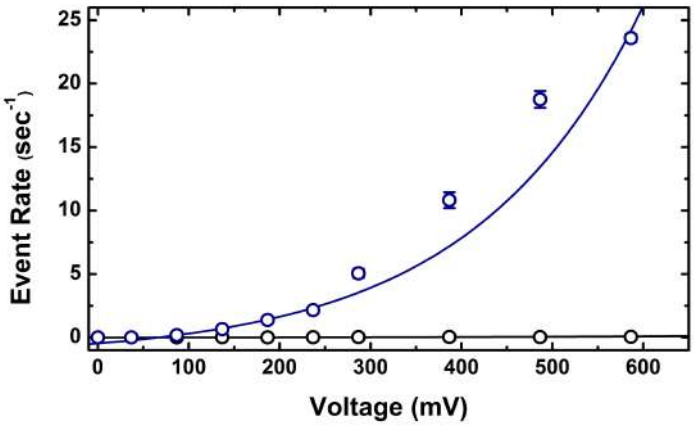

Figure A5- 2 Synthetic biotinylated dsDNA event rate. SS-nanopore event rates for synthetic 156 bp monobiotinylated dsDNA both with (blue) and without (black) bound

MS. DNA concentration for both data sets are 385 nM, similar to the biotin-labeled 5hmC oligonucleotide in Fig. 1c from the main text, considering labeling yield.

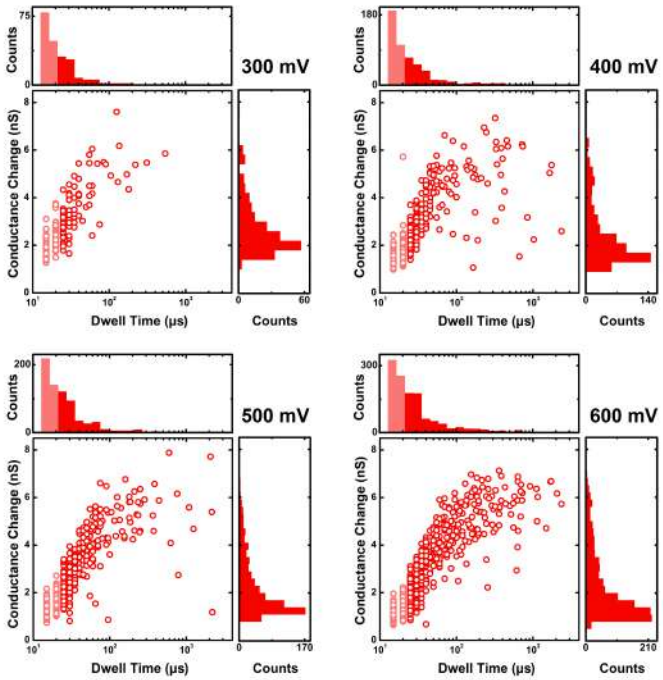

Figure A5- 3 Biotin-labeled mono-5hmC scatter plots. Mean dwell time vs. mean conductance change scatter plots for 156 bp mono-5hmC dsDNA translocation events (see Fig. 1c from main text) following biotin labeling and incubation with MS. Total number of events considered from 300-600 mV are 217, 529, 717, and 1219, respectively. Faded regions represent events below the resolution limit.

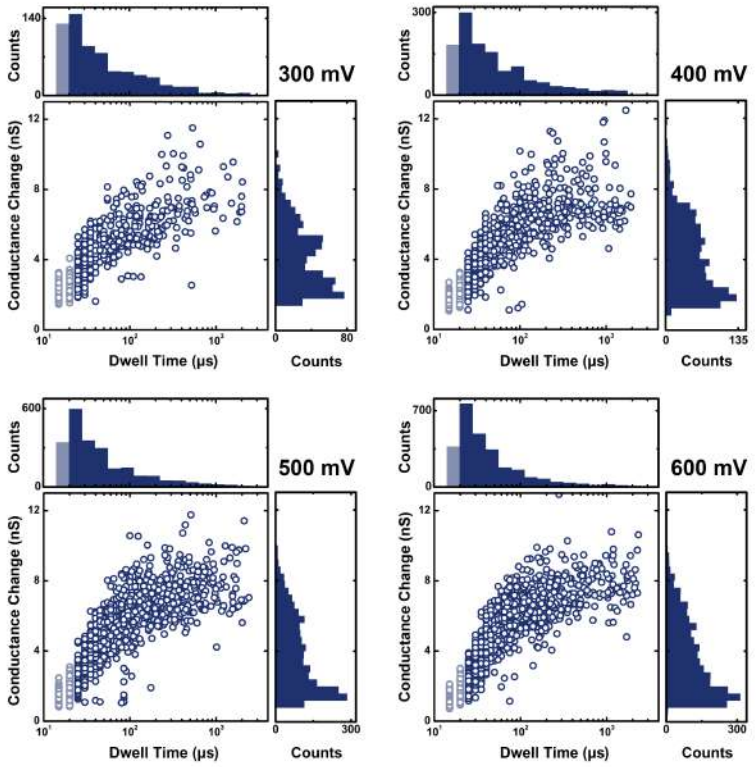

Figure A5- 4 Monobiotinylated dsDNA scatter plots. Mean dwell time vs. mean conductance change scatter plots for 156 bp synthetic monobiotinylated dsDNA translocation events (see Fig. S1) following incubation with MS. Total number of events considered from 300-600 mV are 673, 1250, 2250, and 2650, respectively. Faded regions represent events below the resolution limit.

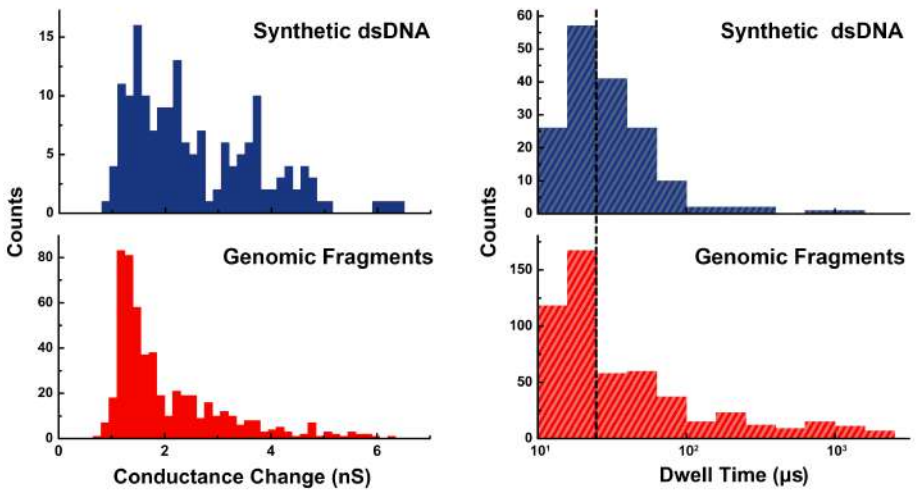

Figure A5- 5 Event characteristics of genomic fragments and synthetic dsDNA. Mean conductance change (left) and dwell time (right) of translocation events collected at 400 mV for 75 bp synthetic monobiotinylated dsDNA (blue, 50 nM, n= 168) and mouse genomic DNA fragmented to an average length of 75 bp and biotin labeled at 5hmC sites (red, c_o= 475 nM, n= 492). A qualitatively similar distribution is observed for both materials. Dashed line on dwell time plots represents the resolution limit.

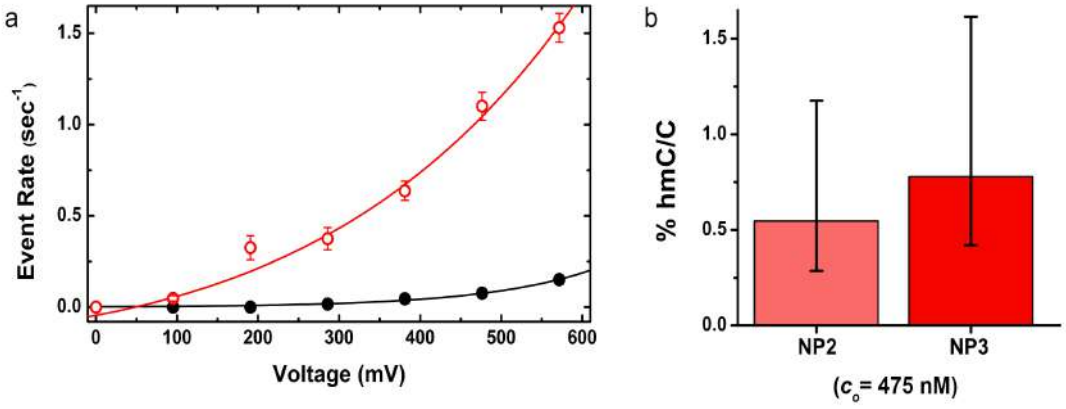

Figure A5- 6 Additional 5hmC/C quantification. (a) Event rate vs. applied voltage (400 mV) for fragmented genomic DNA (c_o= 475 nM) biotin-labeled at 5hmC sites measured on a third SS-nanopore (NP3, see Table S2). Data shows measurement with

(red) and without (black) bound MS and solid lines are exponential fits to the data. (b) Comparison of resulting 5hmC/C ratio R with that of NP2 from the main text (Fig. 3b), also measured at c_o= 475 nM. Quantitative agreement is observed.

Curriculum Vitae

Osama Kazi Zahid
School of Biomedical Engineering Sciences
Wake Forest University School of Medicine
575 N Patterson Ave., Winston-Salem, NC, USA
Mobile: +1 919 841 6753 *E-mail:* osama.zahid87@gmail.com

EDUCATION

2011-2017 **PhD - Molecular & Cellular Biosciences**
Wake Forest University, Winston-Salem, NC, USA
Thesis advisor: Dr. Adam R. Hall
Assistant Professor of Biomedical Engineering

- Developed a novel single-molecule nanopore detection assay and demonstrated quantification of 5-hydroxymethylcytosine (5hmC) content in mouse brain tissue
- Detection of sequence specific biomarkers including microRNAs

2007-2009 **Master of Science in Biotechnology**
Jamia Millia Islamia (JMI), New Delhi-25, India

2004-2007 **Bachelor of Science in Biotechnology**
Manipal Academy of Higher Education (MAHE), Dubai, U.A.E

EXPERIENCE

July 2016 – present **Technology Transfer Intern at Wake Forest Innovations**
Evaluate new disclosures for technical fitness & commercial viability, including intellectual property. Develop primary evaluation summaries and slide decks for marketing technologies. Contributed to a successful grant application ($75,000 from NCBC's TEG Grant). Review and edit non-exclusive licensing agreements and sponsored research agreements.

2009 – 2011 **Research Assistant**,
UAE University, **(Alain, UAE)**
Dr. Mohamed Meetani, Assistant professor of Analytical Chemistry

Project title: *Analysis of peptides, low molecular weight proteins and Antimicrobial peptides from Frog Species using Pyrolysis Liquid Chromatography Mass Spectrometry*

January-July 2009 MSc thesis research **Jawaharlal Nehru University**
Project title: *Chemo-preventive effects of food supplement (Fennel)*

Peer Review Journal Publications

1. "Solid-state nanopore analysis of diverse DNA base modifications using a modular enzymatic labeling process", F. Wang, **O. K. Zahid**, B. Swain, D. Parsonage, T. Hollis, S. Harvey, F. Perrino, R. M. Kohli, E. W. Taylor, A. R. Hall, Nano Letters, DOI: 10.1021/acs.nanolett.7b03911, 2017
2. "Quantifying mammalian genomic DNA hydroxymethylcytosine content using solid-state nanopores", **O. K. Zahid**, B. S. Zhao, C. He, A. R. Hall, Scientific Reports, DOI:10.1038/srep29565, 2016
3. "Sequence-specific recognition of microRNAs and other short nucleic acids with solid-state nanopores", **O. K. Zahid**, F. Wang, J. A. Ruzicka, E. W. Taylor, A. R. Hall, Nano Letters, DOI: 10.1021/acs.nanolett.6b00001, 2016
4. "Nanopore Analysis of Single-Stranded Binding Protein Interactions with DNA", M. M. Marshall, J. Ruzicka, **O. K. Zahid**, V. C. Henrich, E. W. Taylor, and A. R. Hall, Langmuir, DOI: 10.1021/acs.langmuir.5b00457, 2015
5. "Selective detection and quantification of modified DNA with solid-state nanopores", A.T. Carlsen, **O.K. Zahid**, J.A. Ruzicka, E.W. Taylor, A. R. Hall, Nano Letters, DOI:10.1021/nl501340d, 2014
6. "Interpreting the conductance blockades of DNA translocations through solid-state nanopores", A.T. Carlsen, **O.K. Zahid**, J.A. Ruzicka, E.W. Taylor, A.R. Hall, ACS Nano, DOI:10.1021/nn501694n, 2014
7. "Fluorescent Studies of Salicylaldehyde and Other Related Carbonyl Compounds for the Selective and Sensitive Detection of Zinc(II) Ions in Aqueous Solution", S. Hisaindee, **O.K. Zahid**, M.A. Meetani, J. Graham, J. Fluor, DOI: 10.1007/s10895-011-1003-3, 2011
8. "Caerulein-and xenopsin-related peptides with insulin-releasing activities from skin secretions of the clawed frogs, Xenopus borealis and Xenopus amieti (Pipidae)", **O.K. Zahid**, M. Mechkarskab, O.O. Ojo, Y.H.A. Abdel-Wahab, P. R. Flatt, M.A. Meetani, J. M. Conlon, Gen. Comp. Endocrinol., DOI:10.1016/j.ygcen.2011.03.022, 2011
9. "Investigation of the pyrolysis products of methionine-enkephalin-Arg-Gly-Leu using liquid chromatography–tandem mass spectrometry", M.A. Meetani, **O.K. Zahid**, J.M. Conlon, J. Mass Spectrom., DOI: 10.1002/jms.1845,2010

Book Chapters

1. "Helium Ion Microscope Fabrication of Solid-State Nanopore Devices for Biomolecule Analysis", **O.K. Zahid** and A.R. Hall, in Helium Ion Microscopy, (Ed. G. Hlawacek, A. Gölzhäuser), Springer Press, 2016
2. "Characterization of Biological and Condensed Matter at the Nanoscale", A.R. Hall, **O.K. Zahid**, F. Sawafta, A.T. Carlsen, in Advances in Nanoscience and Nanoengineering, (Ed. A.D. Kelkar, D. Herr, J.G. Ryan), CRC Press, 2013

Awards

- **Sigma-Xi Grants-in-Aid of Research (2014)** – Applied and won a $1000 research grant

Student Training and oversight

Laura Miron - Wake Forest University undergraduate intern 2017
Zoe Hurtado - Wake Forest University undergraduate intern 2017
Lauren Huff – High school student, summer lab intern 2017
Austin Bauer – Wake Forest University undergraduate intern 2016-2017
Ellen Oliver – Wake Forest University undergraduate intern 2016-2017
Eran Brown – VT- Wake Forest SBES undergraduate summer intern 2014
Weili Qu – VT- Wake Forest SBES undergraduate summer intern 2014
Nick Rosato – High school student, summer lab intern 2013
Daniel Partridge – undergraduate summer intern UNCG 2012

Training and assisting incoming Graduate Students and Visiting Scholars since Summer 2014

Outreach

Gateway to Science – University of North Carolina Greensboro 2013
NanoDays - North Carolina State University 2012

Languages

Fluent in English, Hindi & Urdu

References

Adam R. Hall, Phd
Assistant Professor of Biomedical Engineering
Virginia Tech-Wake Forest SBES
Wake Forest University School of Medicine
Winston-Salem, NC 27101, USA
Tel +1 336 716 5384
arhall@wakehealth.edu

Jan A. Ruzicka, PhD
Assistant Professor
Department of Basic Pharmaceutical Sciences, High Point University
High Point, NC 27268, USA
Tel +1 706 338 7823
jruzicka@highpoint.edu

Janel Suburu Ezzell, PhD
Patent Specialist
Smith Moore Leatherwood LLP
101 N. Tryon Street,
Charlotte, NC 28246, USA
Tel +1 704 999 5474
Janel.ezzell@gmail.com

Ethan W. Taylor, PhD
Professor
Department of Chemistry
UNCG
Greensboro, NC 27412
Tel +1 336 285 2890
ewtaylor@uncg.edu