2020

Volatile Organic Compound Detection And Disease Diagnostics Using Dna-Functionalized Carbon Nanotube Sensor Arrays

Christopher Erkki Kehayias
University of Pennsylvania

Volatile Organic Compound Detection And Disease Diagnostics Using Dna-Functionalized Carbon Nanotube Sensor Arrays

Abstract

There is a strong desire for novel chemical sensors that can detect low concentrations of volatile organic compounds (VOCs) for early-stage disease diagnostics as well as various environmental monitoring applications. The aim of this thesis work was to address these challenges by developing an “electronic nose” (e-nose) platform based on chemical sensor arrays capable of detecting and differentiating between various VOCs of interest. Sensor arrays were fabricated in a field-effect transistor (FET) configuration with exquisitely sensitive carbon nanotubes (CNTs) as the channel material. The nanotubes were functionalized with a variety of single-stranded DNA oligomers, forming DNA-NT hybrid structures with affinity to a wide variety of VOC targets. Interactions between DNA-NTs and VOCs yielded changes in sensor conductivity that depended strongly on the base sequence of DNA. Arrays of CNT devices were functionalized with up to ten different DNA oligomers to enable electronic signature readouts of VOC binding events. DNA-NT responses were processed with pattern recognition algorithms in order to classify different VOC targets according to their chemical “fingerprints.” This technology was used to measure VOC biomarkers associated with ovarian cancer and COVID-19 from human fluid media. DNA-NT arrays measured headspaces VOCs from 58 blood plasma samples from individual people, including 15 with a late-stage malignant form of ovarian cancer, 6 with early-stage malignant cancer, 16 with a benign form of cancer, and 21 healthy age-matched controls. Statistical techniques based on machine learning were used to discriminate between the malignant, benign, and healthy groups with 90 – 95% classification accuracy. Furthermore, all six early-stage samples were correctly identified with the malignant group, indicating significant progress towards an effective screening method for ovarian cancer. Similar investigations were conducted on sweat samples procured from patients who had tested positive for COVID-19 (CoV+) and those who had tested negative (CoV-). Statistical analysis of the DNA-NT responses to the sweat headspace VOCs revealed highly differentiated clusters associated with the CoV+ and CoV- groups. A binary classifier was constructed using the response data and was estimated to have a 99% classification success rate, suggesting strong potential for utilizing DNA-NTs for COVID screening. Finally, DNA-NT arrays were assessed based on various performance characteristics desired for remote environmental monitoring applications such as pollution monitoring and explosives detection in a warzone. A series of experiments was conducted to evaluate DNA-NT sensitivity, specificity, and longevity using mixtures of 2,6-dinitrotoluene (DNT) and dimethyl methylphosphonate (DMMP) to simulate complex VOC environments. The sensors demonstrated sensitivity to parts-per-billion concentrations of DNT in a highly concentrated background of DMMP. Moreover, the shelf life of these sensors was projected on the order of months, making DNA-NTs promising candidates for a wide range of applications.

Degree Type

Dissertation

Degree Name

Doctor of Philosophy (PhD)

Graduate Group

Physics & Astronomy

First Advisor

Alan T. Johnson

Keywords

Carbon nanotubes, Chemical sensing, Disease diagnostics, DNA, Electronic nose, Volatile organic compounds

Subject Categories

Condensed Matter Physics

VOLATILE ORGANIC COMPOUND DETECTION AND DISEASE DIAGNOSTICS USING DNA-FUNCTIONALIZED CARBON NANOTUBE SENSOR ARRAYS

Christopher E. Kehayias

A DISSERTATION

in

Physics and Astronomy

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2021

Supervisor of Dissertation

_____

A. T. Charlie Johnson, Rebecca W. Bushnell Professor, Department of Physics and Astronomy

Graduate Group Chairperson

_____

Joshua R. Klein, Edmund J. and Louise W. Kahn Professor, Department of Physics and Astronomy

Dissertation Committee

Marija Drndić, Fay R. and Eugene L. Langberg Professor, Department of Physics and Astronomy

Mark Goulian, Charles and William L. Day Distinguished Professor, Department of Physics and Astronomy

Bo Zhen, Professor, Department of Physics and Astronomy

Gary Bernstein, Reese W. Flower Professor, Department of Physics and Astronomy

This work is dedicated to my mother and father whose undying dedication to my well-being and happiness aided me in all my struggles.

ACKNOWLEDGMENTS

I would like to thank the following individuals whose contributions, assistance, and support have made the completion of my thesis possible:

First and foremost, I would like to thank my thesis advisor, Professor Charlie Johnson, for all his wonderful guidance and support. Charlie provided me with ample guidance, direction, and focus for my research while allowing me the freedom to conduct my own investigations. Because of this, I learned through my own experiences about what works and what doesn't in experimental research. Through Charlie's mentorship, I also sharpened my science communication skills and learned proper etiquette for research collaboration. The invaluable lessons I learned in each regard will undoubtedly assist me throughout my career. Perhaps what distinguish Charlie most as an excellent advisor are his limitless patience and compassion. Throughout my graduate academic pursuits, Charlie was always humanely understanding of my own personal hardships and challenges and he repeatedly went out of his way to offer assistance and support when I needed it the most.

I also owe a great deal of gratitude to the Johnson group members who preceded me and kindly showed me the ropes. My foremost mentor was Nicholas Kybert who taught me the ins and outs of DNA-NT fabrication, vapor sensing, and data analysis. I was very fortunate to have worked with such an energetic, intelligent, and talented individual. Carl Naylor helped me tremendously with various aspects of my laboratory work as well as my graduate experience at Penn, as did Pedro Ducos, Ramya Vishnubhotla, Jeremy Yodh, William Zhou, Camilla Schneier, and postdocs Jinglei Ping, Zhaoli Gao, and Madeline Diaz-Serrano.

I want to give special thanks to all members of the group who aided me with my thesis work in various ways. I am indebted to Qicheng Zhang, our brilliant postdoc, who helped improve various aspects of data acquisition and procedures for my experiments. Many thanks go to Emilie Benson, who helped restructure the layout of our gas-delivery system to increase the throughput of our data collection procedure. Emilie spent countless hours with me setting up experiments, taking measurements, fabricating sensor arrays, analyzing and interpreting response data, and so on. I also received wonderful assistance from Ali Ghorashi, a talented Penn undergraduate at the time, as well as from Olivia Dickens, Inayat Bajwa, Jisoo Kang, Yeonjoon Suh, and Sunny Zhang.

I also owe thanks to our collaborators. Dr. George Preti of the Monell Institute was a revered analytical chemist and a wonderful scientist, collaborator, and mentor. His tragic passing earlier this year was heart-wrenching and marked a significant loss to the group. I am thankful for Dr. Preti's group members as well Dr. Young Lee and Jason Eades. Huge thanks go to Dr. Cynthia Otto from the Penn Vet Working Dog Center as well as Dr. Janos Tanyi from Penn Hospital and Dr. Jody Piltz-Seymour for their crucial contributions. I am also grateful for our brief yet fruitful interactions with Rohinton Mehta and Neha Kondeka from X Development.

Finally, I'd like to acknowledge the remaining group members who each played a crucial role in making the Johnson lab a wonderful research environment: postdoc Mengqiang Zhao, graduate students Ram Gona, Chengyu Wen and Rhimjim Chaudhary, and undergraduates Srinivas Mandyam and Adithya Sriram.

ABSTRACT

VOLATILE ORGANIC COMPOUND DETECTION AND DISEASE DIAGNOSTICS USING DNA-FUNCTIONALIZED CARBON NANOTUBE SENSOR ARRAYS

Christopher E. Kehayias

A. T. Charlie Johnson

There is a strong desire for novel chemical sensors that can detect low concentrations of volatile organic compounds (VOCs) for early-stage disease diagnostics as well as various environmental monitoring applications. The aim of this thesis work was to address these challenges by developing an "electronic nose" (e-nose) platform based on chemical sensor arrays capable of detecting and differentiating between various VOCs of interest. Sensor arrays were fabricated in a field-effect transistor (FET) configuration with exquisitely sensitive carbon nanotubes (CNTs) as the channel material. The nanotubes were functionalized with a variety of single-stranded DNA oligomers, forming DNA-NT hybrid structures with affinity to a wide variety of VOC targets. Interactions between DNA-NTs and VOCs yielded changes in sensor conductivity that depended strongly on the base sequence of DNA. Arrays of CNT devices were functionalized with up to ten different DNA oligomers to enable electronic signature readouts of VOC binding events. DNA-NT responses were processed with pattern recognition algorithms in order to classify different VOC targets according to their chemical "fingerprints." This technology was used to measure VOC biomarkers associated with ovarian cancer and COVID-19 from human fluid media. DNA-NT arrays measured headspaces VOCs from 58 blood plasma samples from individual people, including 15 with a late-stage malignant form of ovarian cancer, 6 with

early-stage malignant cancer, 16 with a benign form of cancer, and 21 healthy age-matched controls. Statistical techniques based on machine learning were used to discriminate between the malignant, benign, and healthy groups with 90 – 95% classification accuracy. Furthermore, all six early-stage samples were correctly identified with the malignant group, indicating significant progress towards an effective screening method for ovarian cancer. Similar investigations were conducted on sweat samples procured from patients who had tested positive for COVID-19 (CoV+) and those who had tested negative (CoV-). Statistical analysis of the DNA-NT responses to the sweat headspace VOCs revealed highly differentiated clusters associated with the CoV+ and CoV- groups. A binary classifier was constructed using the response data and was estimated to have a 99% classification success rate, suggesting strong potential for utilizing DNA-NTs for COVID screening. Finally, DNA-NT arrays were assessed based on various performance characteristics desired for remote environmental monitoring applications such as pollution monitoring and explosives detection in a warzone. A series of experiments was conducted to evaluate DNA-NT sensitivity, specificity, and longevity using mixtures of 2,6-dinitrotoluene (DNT) and dimethyl methylphosphonate (DMMP) to simulate complex VOC environments. The sensors demonstrated sensitivity to parts-per-billion concentrations of DNT in a highly concentrated background of DMMP. Moreover, the shelf life of these sensors was projected on the order of months, making DNA-NTs promising candidates for a wide range of applications.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER 1: Overview of the Thesis

There is an ever-increasing need for vapor-phase chemical sensors in biomedical applications including early-stage disease diagnostics or environmental monitoring applications such as pollution detection. Until recently, few technological advancements have been able to effectively respond to these needs because such applications demand high sensitivity to a wide variety of volatile organic compounds (VOCs) as well as fast response times and long sensor lifetimes. The emergence of nano-bio hybrid materials has stirred much excitement within the chemical sensing community due to the combined sensitivity of novel nanomaterials such as carbon nanotubes and graphene with various biomolecules with tunable affinities towards specific or a wide range of target vapor analytes. Electronic systems based on nano-bio materials are uniquely well-suited for chemical sensing applications due to their exquisitely high surface-to-volume ratios, ensuring that electronic transport activity occurs in close proximity to chemical binding events that take place on the material surface. This work focuses on the development of an electronic nose system based on DNA-functionalized carbon nanotube field-effect transistors and the evaluation of this system for ovarian cancer diagnosis of human blood plasma samples, COVID-19 screening based on human sweat, and various performance characteristics desirable in environmental monitoring applications.

Chapter 2 of this thesis discusses the importance of detection systems based on olfaction and motivates the development of vapor sensing technology due that can operate

at the same proficiency as mammalian olfactory systems. Electronic nose systems based on DNA-functionalized carbon nanotube (DNA-NT) vapor sensors are introduced. The electronic properties of carbon nanotubes are discussed, motivating their use in chemical sensing applications due to their remarkable sensitivity. The field-effect transistor (FET) architecture of DNA-NT sensor arrays is presented and an overview of their operability is provided. Chemical functionalization of the nanotubes (NTs) via single-stranded DNA is covered, as is the interaction between target VOCs and the DNA-NT hybrid structures. The Hill-Langmuir theory of binding dynamics is proposed as a model for VOC binding events and is used to explain the relationship between variations in target concentrations and DNA-NT conductivity. Finally, statistical analysis techniques based on two machine learning algorithms, principal component analysis (PCA) and linear discriminant analysis (LDA), are summarized with more vigorous derivations found in the Appendix of this thesis. Binary classification based on receiver operating characteristics (ROCs) are also introduced.

Chapter 3 describes electrode fabrication processes for FET arrays based on photolithography as well as procedures for deposition of the carbon nanotubes across the FET channels and chemical functionalization via single-stranded DNA. Electrical characterization of DNA-NT arrays is presented. General experimental procedures and data acquisition protocols are also introduced, as are data processing techniques. An overview of the equipment used in each experiment is also included.

Chapter 4 discusses the challenge of early-stage diagnosis of ovarian cancer and motivates the development of a reliable screening technique to improve the survivability

of patients who develop this disease. A collaboration between Penn Hospital, the Johnson Group, the Penn Vet Working Dog Center and the Monell Chemical Senses Center is presented with the aim of analyzing VOCs from human blood plasma associated with ovarian cancer patients as well as healthy controls using three different approaches: (1) analytical chemistry using gas chromatography / mass spectrometry to identify differences in VOC composition between diseased and healthy blood; (2) training of sniffer dogs to identify blood samples as diseased or healthy based on their odors; and (3) using DNA-NT arrays to differentiate between VOC signatures associated with cancerous and non-cancerous individuals. This chapter focuses on the third approach. DNA-NT measurements of headspace VOCs from pooled plasma and plasma from individuals are described. Data processing and classification of the sensor response data using LDA is presented. Cross-validation techniques are introduced to test the statistical robustness of the LDA-based classifiers. Lastly, modifications to the experimental procedure in order to accommodate challenges measure plasma VOCs are discussed.

Chapter 5 introduces a collaboration between the Johnson Group and Rohinton Mehta's group from X Development with the aim of evaluating DNA-NT sensitivity, ability to detect low-concentration targets in a complex background, response times, and longevity. Experiments involving a newly fabricated DNA-NT array are described in which array is exposed to multiple different concentrations of 2,6-dinitrotoluene (DNT), a compound known to have a relatively low vapor pressure. Additional experiments involving parts-per-billion concentrations of DNT diluted in a highly concentrated background of dimethyl methylphosphonate (DMMP) are discussed. Similar experiments

are presented in which the fresh DNA-NT array is replaced with one that was prepared four months prior. Special data processing techniques using boxcar smoothing to assist in resolving low-concentration signals are introduced. Conclusions are made is assessment of the DNA-NT performance in all experiments.

Chapter 6 discusses the deadly implications of the new COVID-19 disease and drawbacks of current screening procedures. An investigation involving DNA-NT discrimination of VOCs from human sweat procured from COVID-positive (CoV+) and COVID-negative (CoV-) volunteers is described. LDA processing of the ensuing response data is presented, revealing promising separation between the CoV+ and CoV- groups. Sensitivity and specificity estimates are presented for a binary classifier based on the LDA results, as are estimates of the overall classification accuracy using a ROC curve.

Chapter 7 includes a summary of the thesis, along with a discussion of future prospects and developments needed to convert our apparatus into a reliable technology for clinical or commercial use.

CHAPTER 2: Principles of Electronic Nose Chemical Sensors

In this chapter, we discuss the motivation and advantages of developing an electronic nose (e-nose) technology to explore volatile signatures appropriate for a diverse set of applications. This technology can probe the relatively untapped wealth of information found in chemical signatures based on volatile organic compounds (VOCs) produced from various biological and industrial sources, potentially leading to new and compelling applications.

Section 2.1 describes the e-nose concept and its advantages over other chemical sensing techniques. We introduce an e-nose based on a large array of diversified chemical sensors that produce a unique olfaction response to a given target analyte. The sensing medium is a combination of highly sensitive carbon nanotubes and single-stranded DNA that induces an affinity to a wide variety of VOCs.

In Section 2.2, we discuss the properties of carbon nanotubes. These are nanostructures that are exquisitely sensitive to changes in their local environment, making them ideal candidates for chemical sensing applications. Section 2.3 outlines how we combine carbon nanotubes with field effect transistors to form chemical sensors with an electronic readout. A “chemical gating” mechanism is proposed as the interaction between target analytes and the carbon nanotubes that produces a measurable electrical response. Chemical functionalization with single-stranded DNA is also described.

Section 2.4 introduces the Hill-Langmuir model that describes binding dynamics between chemical receptors and target species. Finally, Section 2.5 outlines basic principles of machine learning algorithms used to perform pattern recognition and statistical analysis on our e-nose response data to identify the chemical signatures of various VOCs of interest.

2.1 Introduction to Electronic Nose Systems

The important role olfaction plays in the survival of living creatures can hardly be overemphasized. One of nature's oldest senses, olfaction enables a lifeform to detect and characterize traces of molecules in the air that indicate the presence of food, danger, or potential mates.[1] In nature, many such molecules are volatile organic compounds (VOCs) produced metabolically by plants and animals. VOCs are compounds that have relatively high vapor pressure and low water solubility for standard temperatures.[2] Generations of evolution have shaped olfaction mechanisms in a variety of organisms into highly successful chemical sensing systems able to discriminate between VOCs with remarkable precision.[3-4]

Compared to countless advancements inspired by our visual and auditory systems, progress towards a technological equivalent of a nose has been relatively slow. Currently, our most reliable chemical vapor sensor is the canine. Dogs' noses have about 300 million olfactory receptors represented by 1100 unique olfactory receptor proteins (ORPs) compared to six million sensors and 350 ORPs for humans, and thus dogs are able to

identify a much wider variety of VOC targets.[5] Dogs' noses are also 10,000 – 100,000 more sensitive than our own, able to detect VOCs at sub-parts per trillion levels.[6-7] Sniffer dogs are highly successful in applications ranging from tracking odor signatures of missing personnel to detecting explosives or narcotics. However, using sniffer dogs in analytical applications is not as ideal. Unlike scientific instruments, a sniffing dog requires long-term professional training and its chemical "readout" is an anecdotal interpretation of its behavior. Often, the dog's response is ineffectively communicated to its trainer and can be influenced by fatigue or short attention span.[8] Moreover, it is practically impossible to determine which chemical compounds in particular are triggering the dogs' responses.[8]

A more scientifically robust approach is found in today's electronic nose (e-nose) sensors. An e-nose is a device that incorporates electronic sensor arrays with pattern recognition systems to detect and discriminate between various odors.[9] The architecture of e-nose systems is inspired by mammalian olfaction, making use many different types of cross-selective olfactory receptors that produce diversified electronic outputs upon exposure to a broad range of VOC analytes.[9-10] These outputs are then recorded in a computer database and processed using pattern recognition and statistical analysis algorithms. The first e-nose was developed in 1982 based on three varieties of semiconductor transducers with broad, overlapping chemical selectivity.[11] The sensing material was composed of conductive carbon black coatings coupled with insulating polymers. Adsorption of VOCs would cause the material to swell, causing a measurable decrease in its conductivity. Unfortunately, these early e-nose sensors could only differentiate between certain hydrophilic and hydrophobic molecules – nevertheless, their

emergence sparked the development of increasingly sophisticated e-nose designs with recent progress towards explosives detection and environmental monitoring applications.[12-14]

In more recent years, e-nose gas sensors have also become popular candidates as detectors for VOC disease biomarkers.[15-16] Disease monitoring based on VOC biomarkers poses the additional challenge that the target molecules are very complex chemically.[17] To differentiate between these targets requires sensing materials that can accommodate such complexity. Fortunately, many biomolecular compounds are available that have natural affinity to biomarkers. These biomolecules can either be borrowed from nature, such as DNA or protein antibodies produced by the body, or engineered to serve a particular function.[18-19] In addition, biomolecules can be chemically linked to electronic materials such as novel nanomaterials that possess exquisitely sensitive electronic properties.[18, 20] Nanomaterials are a natural choice for chemical sensing applications because they exhibit large surface-to-volume ratios and are atomically thin, thus confining charge carrier transport to the material surface where interactions with the environment take place.[21]

This thesis work makes use of an e-nose system based on arrays of single-walled carbon nanotube field-effect transistors (NTFETs). A field-effect transistor (FET) is a three-terminal device in which the conduction between a source and a drain electrode pair is modulated by a voltage applied to a third gate electrode. A special electrode layout for a NTFET sensor array was designed to accommodate multiple sensor types that would respond simultaneously to target analytes (Figure 2.1a). A given array comprises 100 NTFETs divided into ten groups of ten sensors each (Figure 2.1b). Each group is

chemically functionalized with an oligomer of single-stranded DNA with a different base sequence to induce differential affinity to a wide range of VOC targets.[22-23] The sensor arrays are exposed to target biomarkers of interest using a home-built gas-delivery system. VOC analytes that adsorb onto the DNA-nanotube (DNA-NT) hybrid structures elicit electrical responses from the CNTs that are recorded and stored in a computer. The data is later processed using pattern-recognition algorithms to identify informative features useful for discriminating between different VOCs.

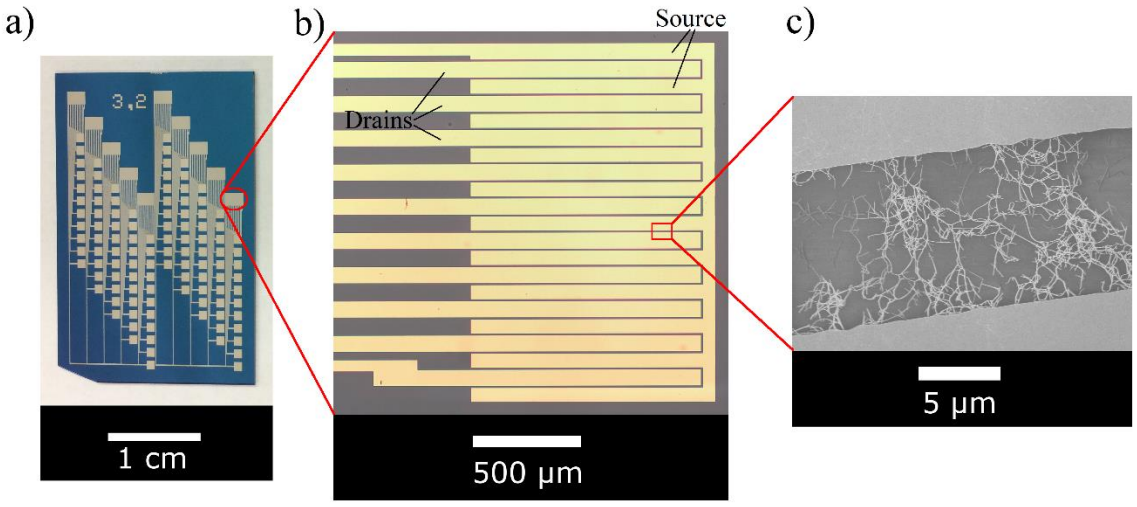

Figure 2.1. (a) Electrode layout for an array of 100 NTFETs partitioned into ten groups of ten devices each. The ten groups are arranged in staggered columns, facilitating individual functionalization with ten different DNA oligomers without cross-contamination between groups. Source and drain electrodes are positioned at the top of each column (one region circled in red). Directly below are square contact pads for external measurement. (b) Microscope image of ten DNA-NT devices composed of ten drain electrodes and one common source electrode arranged in an interdigitated finger configuration. (c) Scanning electron microscopy image of a channel reveals a network of interspersed CNTs between a source-drain electrode pair.

2.2 Principles of Carbon Nanotubes

2.2.1 Overview of Single-Walled Carbon Nanotubes

Single-walled carbon nanotubes (SWCNTs) are carbon allotropes with tubular structures that have diameter length scales on the order of nanometers. Each SWCNT is composed of a single-atomic layer of sp2-hybridized carbon atoms arranged in a hexagonal "honeycomb" lattice. We can immediately draw a similarity to the structure of graphene which is a planar lattice of carbon atoms also arranged in a honeycomb lattice.[24] The geometry of a SWCNT can be imagined by rolling a graphene sheet along a certain axis into a cylinder with seamlessly-bonded carbon atoms forming the surface. In order to preserve the lattice periodicity on the SWCNT, only certain choices for the rolling direction are allowed. This direction is called the chiral vector and the discrete set of geometric conformations for SWCNTs are known as chiralities.

It is possible to obtain multi-walled carbon nanotubes (MWCNTs) in which several coaxial SWCNTs of varying diameters are merged into a single structure.[25] However, I only used SWCNTs throughout my PhD research – hence, I will refer to SWCNTs simply as CNTs for the remainder of this dissertation.

2.2.2 Semiconducting Carbon Nanotubes

The dimensions and chirality of a CNT play crucial roles in determining its electronic properties. Given two linearly independent translation vectors on the graphene lattice, $\vec{\boldsymbol{a}}_{\boldsymbol{1}}$ and $\vec{\boldsymbol{a}}_{\boldsymbol{2}}$, the chiral vector of a CNT, $\vec{\boldsymbol{R}}$, is represented by

$$\vec{\boldsymbol{R}} = n\vec{\boldsymbol{a}}_{\boldsymbol{1}} + m\vec{\boldsymbol{a}}_{\boldsymbol{2}} \tag{2.2.1}$$

where *n* and *m* are integers. The chirality of a given CNT is commonly represented using the shorthand (n, m). Conventionally, *n* is positive with $n \geq m$. Two special cases include the "zigzag" chirality, where $m = 0$, and "armchair" CNTs in which $n = m$ (Figure 2.2). The diameter, *d*, of a (n, m) CNT is given by the following formula:

$$d = a\sqrt{3(n^2 + nm + m^2)}/\pi \tag{2.2.2}$$

where $a = 1.42$ Å is the interatomic distance between adjacent carbon atoms.[25]

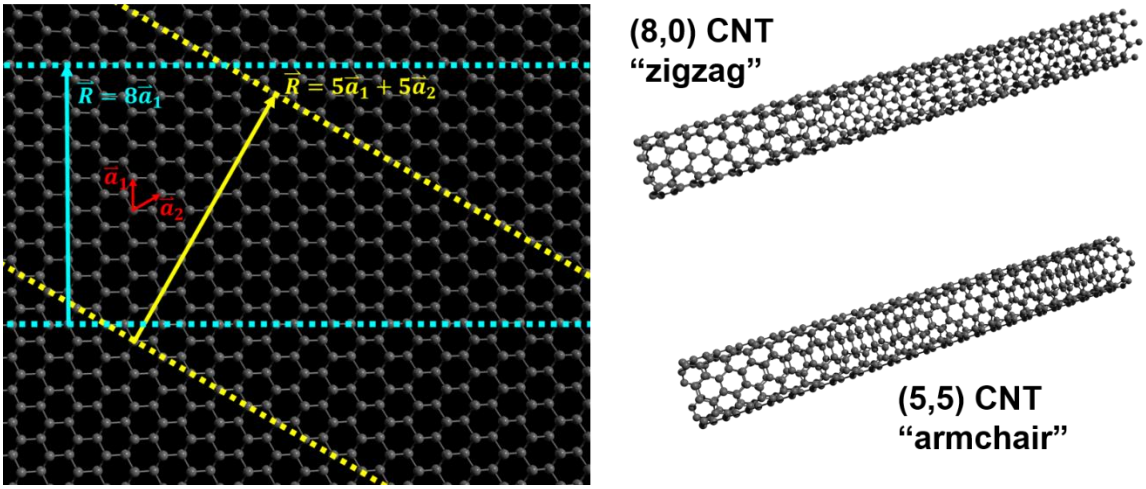

Figure 2.2. The hexagonal "honeycomb" lattice of the graphene structure (left) can be mathematically represented by the translation vectors $\vec{a}_1$ and $\vec{a}_2$. Single-walled carbon nanotubes (SWCNTs) can be constructed by rolling ribbon-like subsets of this lattice (indicated by the dashed edges) into cylinders with circumference $|\vec{R}| = |n\vec{a}_1 + m\vec{a}_2|$ where *n* and *m* are integers. Chiral vectors are shown for a (8, 0) "zigzag" nanotube (blue) and a (5, 5) "armchair" nanotube (yellow).

The chirality of a CNT can indicate whether it behaves as a metal or semiconductor. Semiconducting CNTs are far more desirable for use in our e-nose because their conductances vary drastically when exposed to nearby VOC analytes compared to only negligible changes for metallic CNTs. A (n, m) CNT is metallic if $n = m$ or quasi-metallic with a very small bandgap if $n - m$ is a multiple of 3 with $n \neq m$ and $nm \neq 0$; a semiconducting CNT results in all other cases.[25-26]

The vapor sensors used in this thesis work were fabricated using a solution of 98% semiconducting CNTs suspended in an aqueous solution purchased from NanoIntegris Inc. (Quebec, Canada). From an electronic transport point of view, these CNTs are essentially one-dimensional structures since they are quantum-confined spatially in two dimensions.[27] An elementary result from solid state physics reveals that the electronic density of states, $D(E)$, for a one-dimensional potential takes the following form:

$$D(E) = \frac{1}{2\pi}\sqrt{\frac{2m_e}{\hbar^2}}(E - Ec)^{-1/2} \qquad (2.2.3)$$

where m_e is the electron mass, E is its energy and Ec is the minimum energy of a given subband.[25] The $1/\sqrt{E}$ relationship yields a series of discontinuities in $\frac{dD(E)}{dE}$ called Van Hove singularities that indicate energies within a subband that have a large number of available states (Figure 2.3). The first Van Hove singularity below the Fermi energy coincides with the edge of the valence band which consequently corresponds to a large number of available hole states (the CNTs used in this work are p-type so holes are the majority charge carrier). The number of holes occupying these states can be increased with an external electric field until a generated hole current is well above the noise threshold of

roughly 1 pA. This phenomenon is enabled by the field effect which will be described in the following section.

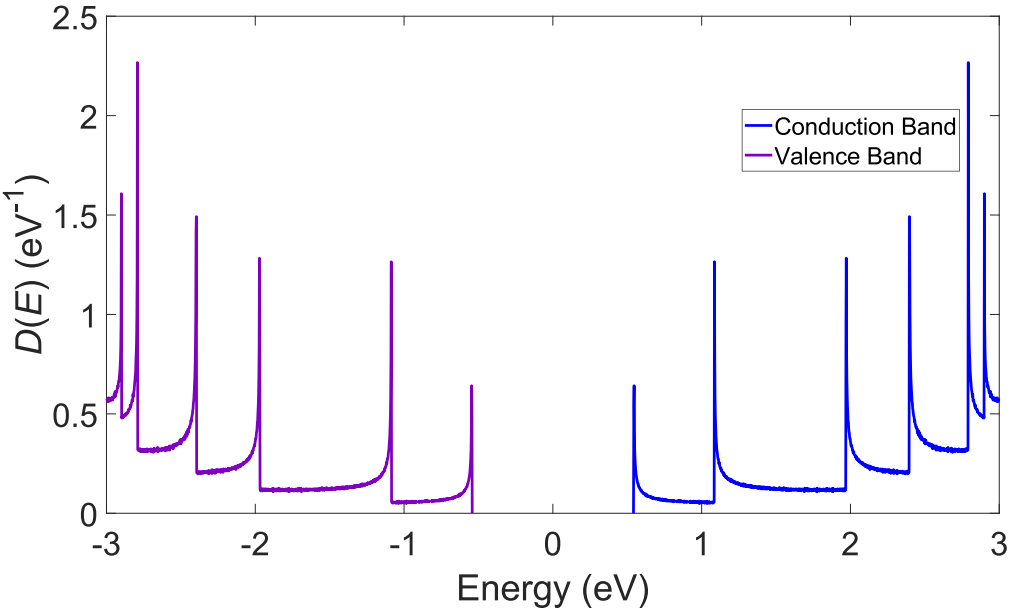

Figure 2.3. The density of electronic states $D(E)$ for a semiconducting (6,5) CNT with a diameter of approximately 1 nm. As a consequence of this nanoscale confinement, the conduction and valence bands are divided into subbands that take the form $D(E) \propto 1/\sqrt{E}$. A series of Van Hove singularities occur at energies where $\frac{dD(E)}{dE}$ diverges. These are energies for which a large number of electronic states are available per unit energy. This figure was adapted from http://www.photon.t.u-tokyo.ac.jp/~maruyama/kataura/kataura.html.

2.3 Carbon Nanotube-Based Sensor Arrays

2.3.1 Field and Chemical Gating of Carbon Nanotube Field-Effect Transistors

The e-nose sensor arrays utilize semiconducting CNTs in a field-effect transistor (FET) configuration. CNTs are deposited between source and drain electrodes patterned

onto Si/SiO_2 substrates, forming a NTFET (Figure 2.4a). A fixed source-drain bias (V_{SD}) is applied to generate a hole current (I_{SD}) through the NTFET channel which is measured in real time with an ammeter. A gate potential (V_G) is simultaneously applied to the underside of the substrate. The Si layer is heavily p-doped which effectively allows charge carriers to transport through the material as they would in a conductor. Hence, the holes near the CNT layer are capacitively coupled to the gate potential with the oxide layer in between serving as a dielectric. A negative gate potential will drive electrons to occupy the Si layer near the oxide which draws holes from the source electrode to match the change in opposing charge. The additional holes increase the local hole concentration, ultimately increasing the conductivity of the p-type CNTs. Similarly, a positive V_G will deprive the Si layer of electrons, thus depleting the local hole concentration near the CNTs and ultimately decreasing the overall device conductivity. This is demonstrated in an $I - V_G$ characteristic where I_{SD} for an NTFET is measured as a function of a varying V_G (Figure 2.4b).

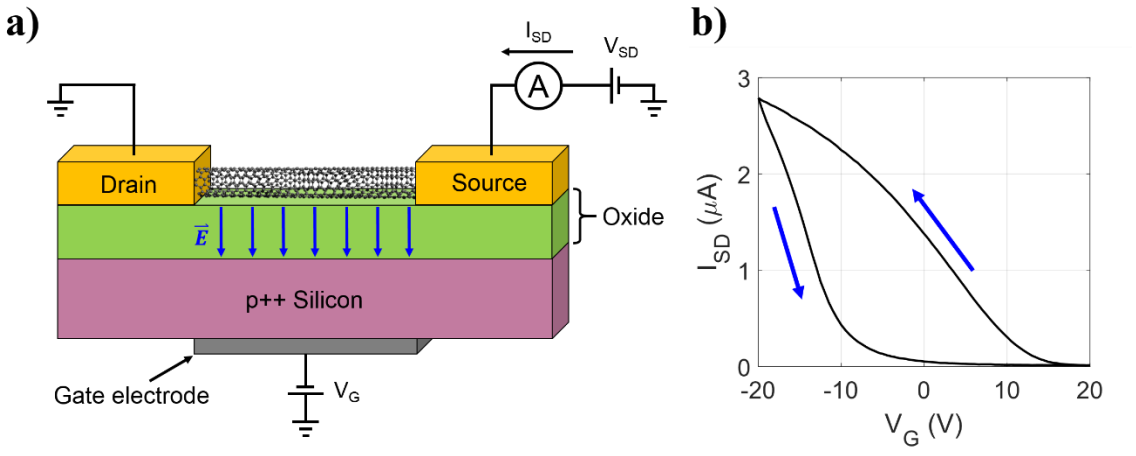

Figure 2.4. (a) A semiconducting CNT is contacted by a source and a drain electrode on a Si/SiO2 substrate. A source-drain bias, V_{SD}, generates a current, I_{SD}, through the NTFET channel. This

current is modulated by a gate potential V_G, applied to the underside of the p-doped Si layer. (b) $I - V_G$ characteristic of a single NTFET depicting the source-drain current as V_G is swept from $-20V$ to $+20V$ and back to $-20V$ as indicated by the blue arrows. The discrepancy between the forward and reverse sweeps is due to a hysteresis effect.[28]

2.3.2 DNA Functionalization of NTFETs

NTFET conductance is also affected by VOC analytes in close proximity to the CNTs via "chemical gating." Though the mechanism is not fully understood, it is believed that certain VOCs will either donate or accept electrons to or from the CNTs, varying the local charge carrier concentration and thus causing a shift in I_{SD} from its baseline value.[29] Other VOCs are thought to dissociate in the presence of an interfacial water layer, leaving behind charged molecules that vary the hole density due to electrostatics.[29]

The sp2-hybridized CNT structure offers a lacking diversity of interactions with VOCs, limiting the range of detectable VOCs.[30] Hence, the CNTs are functionalized with single-stranded DNA oligomers of a particular base sequence that introduce binding sites for chemical sensing.[30] DNA oligomers adhere onto CNT sidewalls via $\pi - \pi$ stacking interactions, forming DNA-NT hybrid structures. This is a relatively strong chemical attraction that does not interfere with the covalent carbon-carbon bonds of the CNT, thus preserving the desirable electronic sensitivity of the overall structure.[31] Furthermore, the DNA oligomers assume sequence-specific conformations with loop and hairpin structures that enable diversified interactions between VOC analytes and the DNA-NT structures.[30, 32-34] The sensor arrays used in this research were functionalized with ten DNA oligomers 21 – 24 bases long (see Chapter 3, Table 3.1), offering a vast wealth of geometrical and

chemical complexity that is similarly enjoyed by olfactory receptor proteins found in nature. The base sequences of these oligomers are random, though in principle they can also be engineered to have targeted affinities towards certain types of VOC analytes of interest.[22-23] Moreover, customized oligomers of single-stranded DNA are easily accessible and can be purchased inexpensively from biotechnology vendors such as ThermoFisher (Waltham, MA) for roughly one US dollar per base.

For a typical experiment, a DNA-NT array is exposed to a mixture of VOC analytes and water vapor diluted in nitrogen gas. Based on thermodynamics studies of similar sensing mechanisms, it is hypothesized that the VOCs are solubilized by water molecules that form an interfacial layer with the DNA structures, enabling an interaction between the analytes and the DNA (Figure 2.5a).[35-36] The VOCs can acquire a charge via dissociation or ionization, or possibly by exchanging electrons with the DNA-NT hybrid structure. The DNA-NT fabrication process is designed to limit the DNA functionalization layer to nanoscale thickness, ensuring that target binding events are in close proximity to the CNT. The result is an amplified electrostatic interaction between the charged VOCs and the CNTs, causing measurable shifts in device currents from baseline. Earlier investigations demonstrated that propionic acid, a known proton donor, evoked positive shifts in DNA-NT currents from baseline,[30] suggesting that a negatively-charged layer of deprotonated propionic acid ions evoked an increase in CNT hole concentration (Figure 2.5b).[30] The same study also demonstrated that dimethyl methylphosphonate (DMMP), a strong electron donor,[37] elicited negative DNA-NT current responses (Figure 2.5c).

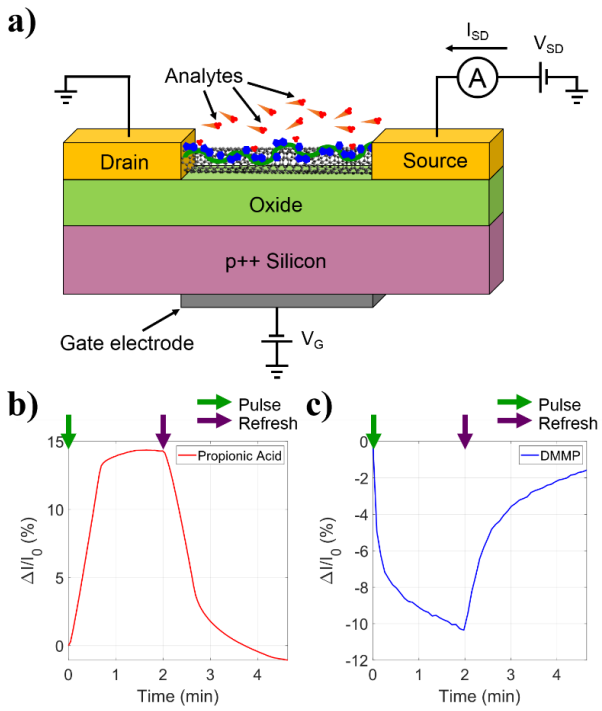

Figure 2.5. (a) A NTFET, depicted by a single CNT for simplicity. The CNT is functionalized with a single-stranded DNA oligomer of a particular base sequence (blue and green structure) forming a DNA-NT hybrid material. Target VOCs (red) interact with DNA binding sites and are adsorbed onto the DNA-NT. (b) A single DNA-NT sensor is exposed to a two-minute pulse of hydrated propionic acid. A positive fractional current response ensues, presented as $\Delta I/I_0$ where I_0 is the baseline current. The DNA-NT is subsequently refreshed with nitrogen gas at the same relative humidity to remove desorbing VOCs, essentially recovering the device current to baseline. (c) The same DNA-NT was exposed to dimethyl methylphosphonate (DMMP) in the same manner, which elicits a negative shift in current relative to baseline.

2.4 Hill-Langmuir Binding Dynamics

The binding interaction between single-stranded DNA and target VOCs is analogous to the binding dynamics of certain protein complexes in which ligand molecules

link to one or more receptors on a biomolecule to activate a particular biological function such as various types of molecular machinery within a living cell.[38-42] The physical mechanism of such systems can be modeled using Hill-Langmuir analyte binding dynamics inspired by a simplified system of ligands occupying substrate-dwelling receptors or binding sites (Figure 2.6). In our case, the ligands are VOCs occurring at a fixed concentration, *c*, which is large enough such that VOC binding events cause a negligible decrease in *c*.

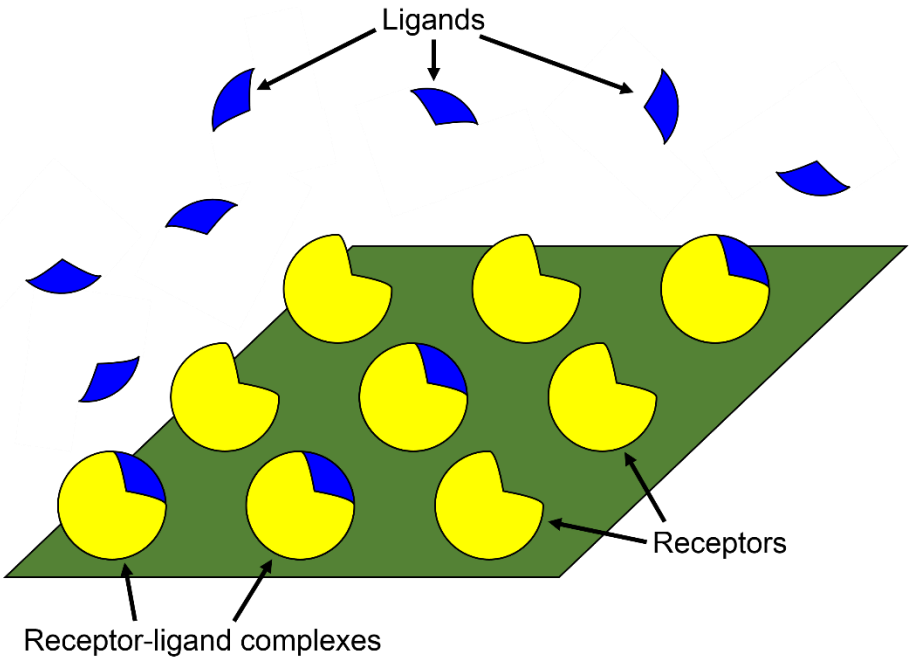

Figure 2.6. An array of receptors (yellow) is exposed to VOC ligands (blue) at a fixed concentration. Some of the ligands bind with receptors to form receptor-ligand protein complexes.

Once a ligand is bound to a receptor, the substrate's affinity for binding another VOC depends on the cooperativity of binding, *n*. Positively cooperative binding occurs

when $n > 0$ in which one adsorbed ligand increases the substrate's affinity while negatively cooperative binding occurs for $n < 0$ in which the affinity decreases. The binding is said to be noncooperative for $n = 0$ if the affinity is independent of already-bound ligands. The fractional occupancy, θ, is defined as the fraction of receptors with bound ligands. When the system is in thermal equilibrium, θ will reach a steady-state value given by the Hill equation:

$$\theta = \frac{(c/K_a)^n}{1 + (c/K_a)^n} \tag{2.4.1}$$

where K_a is the VOC concentration for which half the binding sites are occupied and n is the Hill coefficient describing the cooperativity of binding. K_a and n are generally obtained experimentally by fitting chemical response data to (2.4.1) for multiple values of c (Figure 2.7).

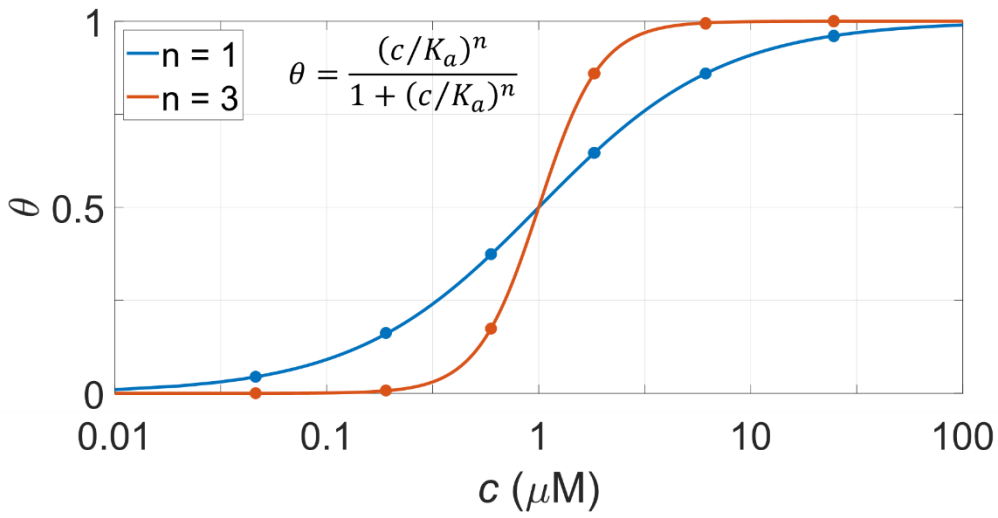

Figure 2.7. Hypothetical Hill-Langmuir curves outline the relationship between the fractional occupancy, θ, and the ligand concentration, c, for two different values of the Hill coefficient, n. K_a = 1 μM for both curves, the concentration at which θ is equal to one half. The colored dots represent experimentally obtained values from which K_a and n are extrapolated.

DNA-NT responses in equilibrium (i.e. responses that reach a stable value) are proportional to θ since $\Delta I/I_0$ depends on the number of adsorbed VOCs. Thus, a given DNA-NT's sensitivity to changes in VOC concentration is strongly related to the mutual affinity between the DNA and the VOCs which is reflected by K_a and n.

2.5 Introduction to Machine Learning Algorithms

As was discussed in Section 2.1, an electronic nose consists of an array of chemically diverse sensors that can discriminate between a vast set of VOCs. Ideally, each DNA oligomer contributes some orthogonal information about a target odor signature that enhances the discrimination power, or "chemical resolution," of the e-nose. Thus, there is an apparent advantage for e-nose systems that incorporate a large number of distinct sensor types.[43-44] A consequence of this, however, is that the ensuing high-parameter sensor output is necessarily complex, making it challenging to extract useful information out of it using conventional data analysis techniques. First-principles investigations of the system's behavior are only possible for drastically simplified models. Meanwhile, graphical visualization is generally fruitless because DNA oligomers have largely overlapping selectivity, thus any informative features are generally obscured by high parameter correlation. A specialized statistical approach is needed to analyze the NTFET response outputs.

Fortunately, there are numerous computational tools available for constructing predictive models in exploratory data analysis. Principal component analysis (PCA) and

linear discriminant analysis (LDA) are two such algorithms that we have used to analyze NTFET array outputs. Both PCA and LDA are linear algebra-based techniques used to identify patterns in high-dimensional data sets. In doing so, these algorithms can represent the original data with reduced dimensionality, keeping the most important discriminatory features. The following sections will review the mathematical foundations of PCA and LDA.

2.5.1 Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is an unsupervised statistical procedure that performs an orthogonal linear transformation on a set of observations of possibly correlated variables to convert them into a set of values represented by linearly uncorrelated variables called principal components.[45-46] This is essentially a change of basis onto a new coordinate system such that the first principal component represents the direction with the greatest variance when scalar projections of the data lie on this axis. Subsequent principal components are associated with lower variances in descending order. Any correlation between variables in the original data representation indicates redundancy, and thus the data can be simplified by replacing the correlated variables with a single variable. In this way, significant trends and features can be revealed using only a few components.

Figure 2.8 shows a simulated set of 1,000 DNA-NT responses to a given type of VOC. The points are drawn randomly from a bivariate gaussian distribution to represent variations in signal due to differences in DNA-NT sensitivities as well as sources of random noise. For simplicity, only Seq1 and Seq2 responses are shown. PCA is an

unsupervised algorithm in that no prior knowledge of class distinction between data points is assumed. The first principal component, PCA1, maximizes the variance of the overall data and is associated with the measurement signal while the orthogonal axis, PCA2, represents the weakest correlation and is associated with signal noise. Effectively, the first principal component is the axis which maximizes the signal-to-noise ratio for the measurements. A more detailed mathematical explanation of PCA can be found in Appendix A.1.

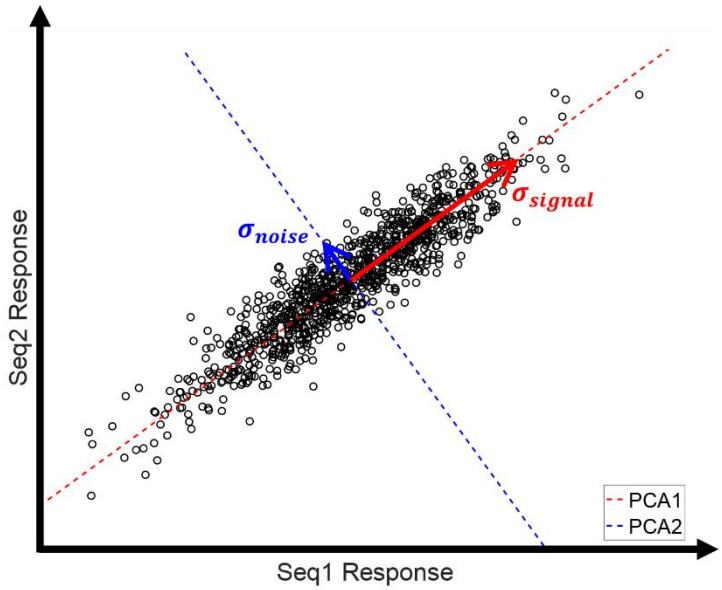

Figure 2.8. A simulated set of 1,000 DNA-NT responses are drawn randomly from a bivariate Gaussian distribution. Only Seq1 and Seq2 responses are shown for simplicity. The first principal component, PCA1, maximizes the variance of the overall data set and is associated with the signal while the orthogonal axis, PCA2, represents the weakest correlation and is associated with signal noise.

2.5.2 Linear Discriminant Analysis (LDA)

Like PCA, LDA performs a linear transformation on a dataset to reduce its dimensionality and facilitate pattern recognition. LDA is a supervised algorithm that finds directions (called linear discriminants) that maximize separation between two or more data classes. This is achieved when the separation between the centroids of known class distributions is greatest when compared to the measure spread for the two clusters. Figure 2.9 illustrates two hypothetical distributions of Seq1 and Seq2 responses to two different types of VOCs. The measurements would clearly overlap significantly if projected onto either the horizontal or the vertical axis which would result in poor separation between the clusters. The optimal separation is expressed by projections onto the axis designated by the first linear discriminant (see Appendix A.2 for the mathematical foundations for LDA).

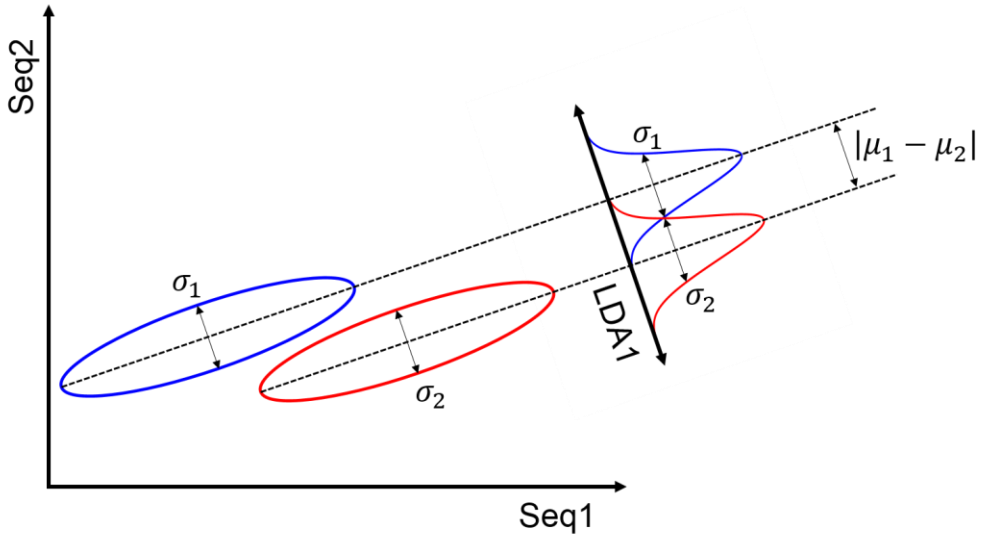

Figure 2.9. A simulated two-parameter data set with two classes of Gaussian-distributed data (red and blue). The distributions overlap significantly along the horizontal and vertical axes. Using LDA, one can discover a rotated representation of these axes (shown as the diagonal

lines). The distributions on the left show histograms of the data projected onto the corresponding axis, resulting in optimized class separation.

A predictive classifier can be constructed by fitting LDA parameters to an existing set of class-labeled data points. A multivariate normal distribution is used to model each class distribution which assigns a probability to any new data point. The estimated classification of the new point is selected based on the class for which this probability is the highest.

2.5.3 Receiver Operating Characteristics

The accuracy of a predictive classifier model can be validated using a receiver operating characteristic (ROC) curve. Given a collection of measurements of two distinct observables, a ROC curve provides an estimate of the classification accuracy of the model by comparing the false positive rate (FPR) and true positive rate (TPR) for all possible measurement values. This is represented visually in Figure 2.10, where two overlapping distributions of measurements for a positive and negative signal (Figure 2.10a) are distinguished by sweeping a measurement parameter, ζ, and computing the relationship between the FPR and TPR (Figure 2.10b). The area under the curve (AUC) formed by this relationship determines the probability of correctly classifying a measurement picked at random as negative or positive. This AUC is an estimate of the overall accuracy of the classifier model.

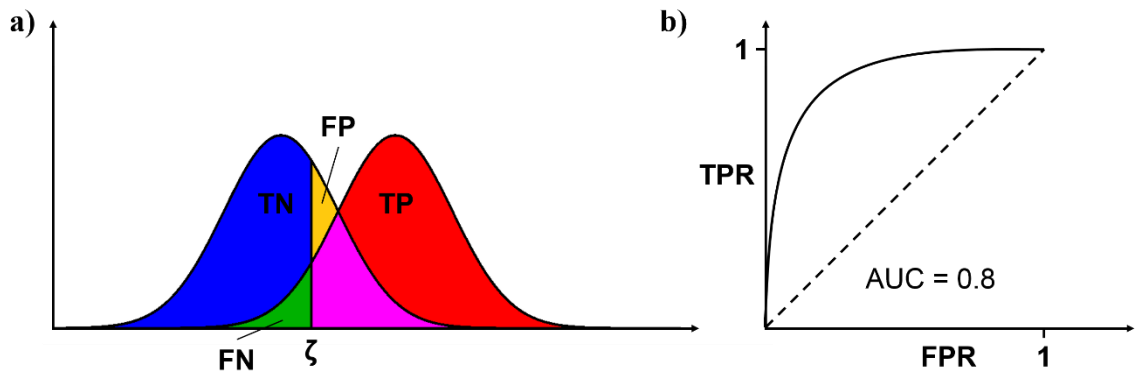

Figure 2.10. Two overlapping distributions of measurements for an imagined two-class system are distinguished for every possible value of the measurement parameter, ζ. The colored areas represent the true negative (TN), true positive (TP), false negative (FN), and false positive (FP) regions for a given ζ. (b) An imagined ROC curve shows the relationship between the false positive rate (FPR) and the true positive rate (TPR). The area under this curve (AUC) determines the probability of correctly classifying a measurement picked at random. The dashed line represents the ROC when the two distributions are perfectly indistinguishable.

References

1. Sarafoleanu, C.; Mella, C.; Georgescu, M.; Perederco, C., The importance of the olfactory sense in the human behavior and evolution. *J Med Life* **2009,** *2* (2), 196-8.
2. *What are volatile organic compounds (VOCs)?*, U. S. Environmental Protection Agency, August 2019, https://www.epa.gov/indoor-air-quality-iaq/what-are-volatile-organic-compounds-vocs.
3. Hoover, K. C., Evolution of olfactory receptors. *Methods Mol Biol* **2013,** *1003*, 241-9.
4. Niimura, Y., Olfactory receptor multigene family in vertebrates: from the viewpoint of evolutionary genomics. *Curr Genomics* **2012,** *13* (2), 103-14.
5. Padodara, R., Olfactory Sense in Different Animals. *The Indian Journal off Veterinary Science* **2014,** *2*, 1-14.
6. Walker, D. B.; Walker, J. C.; Cavnar, P. J.; Taylor, J. L.; Pickel, D. H.; Hall, S. B.; Suarez, J. C., Naturalistic quantification of canine olfactory sensitivity. *Applied Animal Behaviour Science* **2006,** *97* (2-4), 241-254.
7. Krestel, D.; Passe, D.; Smith, J. C.; Jonsson, L., Behavioral Determination of Olfactory Thresholds to Amyl Acetate in Dogs. *Neuroscience and Biobehavioral Reviews* **1984,** *8* (2), 169-174.

8. Russell, R. A., Tracking chemical plumes in constrained environments. *Robotica* **2001,** *19*, 451-458.
9. Wilson, A. D.; Baietto, M., Applications and advances in electronic-nose technologies. *Sensors (Basel)* **2009,** *9* (7), 5099-148.
10. Wilson, A. D.; Baietto, M., Advances in electronic-nose technologies developed for biomedical applications. *Sensors (Basel)* **2011,** *11* (1), 1105-76.
11. Persaud, K.; Dodd, G., Analysis of discrimination mechanisms in the mammalian olfactory system using a model nose. *Nature* **1982,** *299* (5881), 352-5.
12. Wiederoder, M. S.; Nallon, E. C.; Weiss, M.; McGraw, S. K.; Schnee, V. P.; Bright, C. J.; Polcha, M. P.; Paffenroth, R.; Uzarski, J. R., Graphene Nanoplatelet-Polymer Chemiresistive Sensor Arrays for the Detection and Discrimination of Chemical Warfare Agent Simulants. *Acs Sensors* **2017,** *2* (11), 1669-1678.
13. Zhong, Y.; He, Y.; Ge, Y.; Song, G., beta-Cyclodextrin protected Cu nanoclusters as a novel fluorescence sensor for graphene oxide in environmental water samples. *Luminescence* **2017,** *32* (4), 596-601.
14. Benitez-Martinez, S.; Lopez-Lorente, A. I.; Valcarcel, M., Graphene quantum dots sensor for the determination of graphene oxide in environmental water samples. *Anal Chem* **2014,** *86* (24), 12279-84.
15. Kodogiannis, V.; Wadge, E., The use of gas-sensor arrays to diagnose urinary tract infections. *Int J Neural Syst* **2005,** *15* (5), 363-76.
16. McCulloch, M.; Jezierski, T.; Broffman, M.; Hubbard, A.; Turner, K.; Janecki, T., Diagnostic accuracy of canine scent detection in early- and late-stage lung and breast cancers. *Integr Cancer Ther* **2006,** *5* (1), 30-9.
17. Rusling, J. F.; Kumar, C. V.; Gutkind, J. S.; Patel, V., Measurement of biomarker proteins for point-of-care early detection and monitoring of cancer. *Analyst* **2010,** *135* (10), 2496-511.
18. Goldsmith, B. R.; Mitala, J. J.; Josue, J.; Castro, A.; Lerner, M. B.; Bayburt, T. H.; Khamis, S. M.; Jones, R. A.; Brand, J. G.; Sligar, S. G.; Luetje, C. W.; Gelperin, A.; Rhodes, P. A.; Discher, B. M.; Johnson, A. T. C., Biomimetic Chemical Sensors Using Nanoelectronic Readout of Olfactory Receptor Proteins. *Acs Nano* **2011,** *5* (7), 5408-5416.
19. Naylor, C. H.; Kybert, N. J.; Schneier, C.; Xi, J.; Romero, G.; Saven, J. G.; Liu, R. Y.; Johnson, A. T. C., Scalable Production of Molybdenum Disulfide Based Biosensors. *Acs Nano* **2016,** *10* (6), 6173-6179.
20. Lerner, M. B.; Matsunaga, F.; Han, G. H.; Hong, S. J.; Xi, J.; Crook, A.; Perez-Aguilar, J. M.; Park, Y. W.; Saven, J. G.; Liu, R. Y.; Johnson, A. T. C., Scalable Production of Highly Sensitive Nanosensors Based on Graphene Functionalized with a Designed G Protein-Coupled Receptor. *Nano Letters* **2014,** *14* (5), 2709-2714.
21. Lerner, M. B.; Resczenski, J. M.; Amin, A.; Johnson, R. R.; Goldsmith, J. I.; Johnson, A. T. C., Toward Quantifying the Electrostatic Transduction Mechanism in Carbon Nanotube Molecular Sensors. *Journal of the American Chemical Society* **2012,** *134* (35), 14318-14321.

22. Patel, D. J.; Suri, A. K.; Jiang, F.; Jiang, L.; Fan, P.; Kumar, R. A.; Nonin, S., Structure, recognition and adaptive binding in RNA aptamer complexes. *J Mol Biol* **1997,** *272* (5), 645-64.
23. Breaker, R. R., Natural and engineered nucleic acids as tools to explore biology. *Nature* **2004,** *432* (7019), 838-45.
24. Kwak, J.; Gallagher, M.; Ozdener, M. H.; Wysocki, C. J.; Goldsmith, B. R.; Isamah, A.; Faranda, A.; Fakharzadeh, S. S.; Herlyn, M.; Johnson, A. T.; Preti, G., Volatile biomarkers from human melanoma cells. *J Chromatogr B Analyt Technol Biomed Life Sci* **2013,** *931*, 90-6.
25. Natelson, D., Nanostructures and nanotechnology. Cambridge University Press: 2015; p 182-191.
26. Laird, E. A. K., F.; Steele, G. A.; Grove-Rasmussen, K.; Nygård, J.; Flensberg, K.; Kouwenhoven, L. P., Quantum Transport in Carbon Nanotubes. *Reviews of Modern Physics* **2015,** *87* (3), 703-764.
27. Aqel, A.; Abou El-Nour, K. M. M.; Ammar, R. A. A.; Al-Warthan, A., Carbon nanotubes, science and technology part (I) structure, synthesis and characterisation. *Arabian Journal of Chemistry* **2012,** *5* (1), 1-23.
28. Radosavljević, M., Freitag, M., Thadani, K. V., & Johnson, A. T., Nonvolatile Molecular Memory Elements Based on Ambipolar Nanotube Field Effect Transistors. *Nano Letters* **2002,** *2* (7), 761–764.
29. Kong, J.; Dai, H. J., Full and modulated chemical gating of individual carbon nanotubes by organic amine compounds. *Journal of Physical Chemistry B* **2001,** *105* (15), 2890-2893.
30. Staii, C.; Johnson, A. T., Jr.; Chen, M.; Gelperin, A., DNA-decorated carbon nanotubes for chemical sensing. *Nano Lett* **2005,** *5* (9), 1774-8.
31. Zheng, M.; Jagota, A.; Semke, E. D.; Diner, B. A.; McLean, R. S.; Lustig, S. R.; Richardson, R. E.; Tassi, N. G., DNA-assisted dispersion and separation of carbon nanotubes. *Nat Mater* **2003,** *2* (5), 338-42.
32. Rutkauskas, D.; Zhan, H.; Matthews, K. S.; Pavone, F. S.; Vanzi, F., Tetramer opening in LacI-mediated DNA looping. *Proc Natl Acad Sci U S A* **2009,** *106* (39), 16627-32.
33. Johnson, R. R.; Johnson, A. T. C.; Klein, M. L., Probing the structure of DNA-carbon nanotube hybrids with molecular dynamics. *Nano Letters* **2008,** *8* (1), 69-75.
34. Johnson, R. R.; Kohlmeyer, A.; Johnson, A. T. C.; Klein, M. L., Free Energy Landscape of a DNA-Carbon Nanotube Hybrid Using Replica Exchange Molecular Dynamics. *Nano Letters* **2009,** *9* (2), 537-541.
35. Hierlemann, A.; Zellers, E. T.; Ricco, A. J., Use of linear solvation energy relationships for modeling responses from polymer-coated acoustic-wave vapor sensors. *Analytical Chemistry* **2001,** *73* (14), 3458-3466.
36. Johnson, K. J.; Rose-Pehrsson, S. L., Sensor Array Design for Complex Sensing Tasks. *Annual Review of Analytical Chemistry, Vol 8* **2015,** *8*, 287-310.
37. Novak, J. P.; Snow, E. S.; Houser, E. J.; Park, D.; Stepnowski, J. L.; McGill, R. A., Nerve agent detection using networks of single-walled carbon nanotubes. *Applied Physics Letters* **2003,** *83* (19), 4026-4028.

38. Ding, S.; Sachs, F., Single channel properties of P2X2 purinoceptors. *J Gen Physiol* **1999,** *113* (5), 695-720.
39. Chu, D.; Zabet, N. R.; Mitavskiy, B., Models of transcription factor binding: sensitivity of activation functions to model assumptions. *J Theor Biol* **2009,** *257* (3), 419-29.
40. Alon, U., *An introduction to systems biology : design principles of biological circuits*. Chapman & Hall/CRC: Boca Raton, FL, 2007; p xvi, 301 p., 4 p. of plates.
41. Hartwell, L. H.; Hopfield, J. J.; Leibler, S.; Murray, A. W., From molecular to modular cell biology. *Nature* **1999,** *402* (6761 Suppl), C47-52.
42. Teif, V. B., Ligand-induced DNA condensation: Choosing the model. *Biophysical Journal* **2005,** *89* (4), 2574-2587.
43. Hopfield, J. J., Odor space and olfactory processing: collective algorithms and neural implementation. *Proc Natl Acad Sci U S A* **1999,** *96* (22), 12506-11.
44. Gelperin, A. and Hopfield, J.J., Electronic and Computational Olfaction, in Chemistry of Taste: Mechanisms, Behaviors and Mimics. (eds. P. Given & D. Paredes) 289-317 (American Chemical Society, Washington DC; 2002).
45. Joliffe, I. T. (n.d.). *Principal Component Analysis*. Springer.
46. Bishop, C. M., *Pattern recognition and machine learning*. Springer: New York, 2006; p 561-565.

CHAPTER 3: Device Fabrication, Experimental Methods, and Data Processing

In Section 3.1, we discuss the fabrication process of DNA-NT arrays and how we utilize them in experiments. Our processes are designed to yield high-yield, ultrasensitive chemical sensors using standard and easily reproducible techniques. We fabricate source-drain electrode pairs on silicon-based substrates using photolithography. A CNT network is deposited across the FET channels, each chemically functionalized with single-stranded DNA of a particular base sequence. DNA-NT arrays are characterized for quality and sensitivity using I-V_G sweeps.

Section 3.2 outlines the vapor delivery system used to expose DNA-NT arrays to target VOCs. A home-built experimental setup was is used to house multiple VOC sources at a time. A series of computer-automated valves and mass flow controllers deliver the volatile headspace of these samples to a specially constructed sensor chamber according to a computer-automated protocol. The sensor chamber is fitted with contacts that enable rapid, high-precision measurements of 100 DNA-NT currents in real time.

Section 3.3 describes the typical procedure for collecting DNA-NT chemical signature data as well as data analysis methods. DNA-NT arrays are generally exposed to a series of pulses of target VOCs from a given sample. A refresh period follows each pulse to remove analytes from the system, returning DNA-NTs to a quiescent state. The current responses to each exposure are computed as the fractional current shift relative to a baseline curve that is fitted to the response data using linear interpolation. Baseline-fitted responses

from all DNA-NTs functionalized with the same sequence of DNA are averaged together to produce a representative response for that sensor type.

3.1 Fabrication and Characterization of NTFET Sensor Arrays

3.1.1. Photolithographic Patterning of Electrode Arrays

A standard photolithographic procedure is implemented to fabricate FET electrodes using the Quattrone Nanofabrication Facility at the Krishna P. Singh Center (see Appendix C for a full detailed description of the procedure). The substrates are standard boron-doped <100> silicon wafers, approximately 10 cm (4 inches) in diameter and 500 – 550 μm in thickness. Each Si wafer has 285 nm layer of polished dry thermal oxide (SiO_2). Prior to electrode patterning, 15 nm of Al_2O_3 is deposited via atomic layer deposition (ALD) on a Si/SiO_2 wafer (Figure 3.1). This is done because CNTs have improved adhesion to Al_2O_3 over the bare SiO_2 surface. The wafer is then coated with a positive photoresist, an organic material that degrades when exposed to certain wavelengths of light. The wafer is placed in a mask aligner where ultraviolet light is passed through a photomask, exposing the photoresist only in selected regions. Any exposed resist is dissolved away with a developer, leaving behind bare substrate. The entire wafer is then coated with 5 nm Cr followed by 40 nm of Au using an electron beam physical vapor deposition (EBPVD) system. The Cr layer is necessary to increase the adhesion between the oxide and the Au layer. All remaining photoresist is dissolved by submerging the wafer into a solvent containing 1-methyl-2-

pyrrolidinone. Whatever metal not in direct contact with the wafer is lifted off and removed, thus leaving behind the desired electrode pattern.

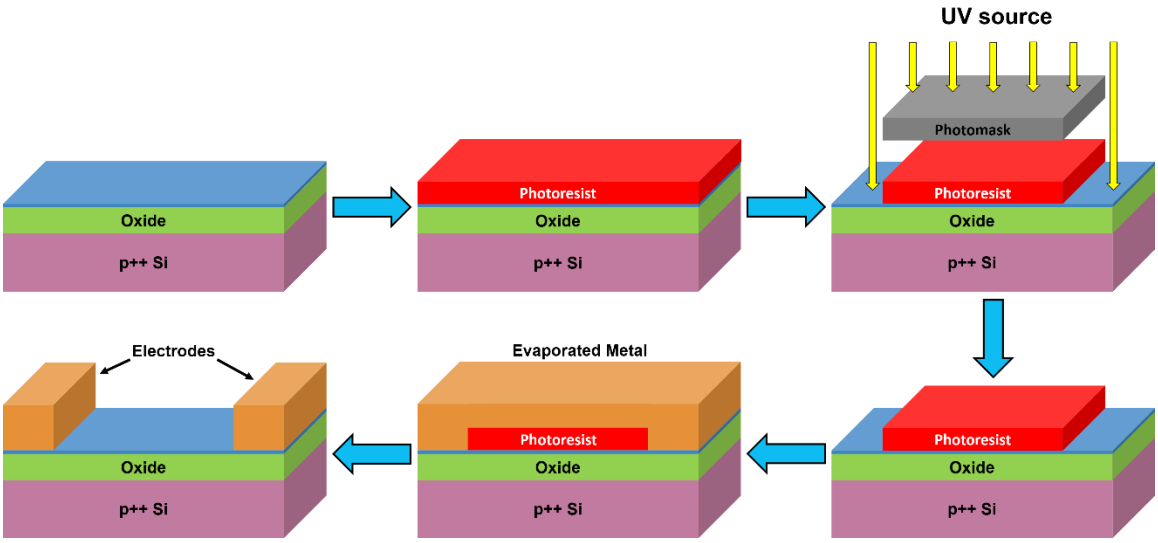

Figure 3.1. A standard photolithography process is used to define electrode features on a Si/SiO_2 substrate. Starting at the upper-left diagram, a layer of Al_2O_3 (shown in blue) is deposited onto the substrate to increase its affinity to CNTs. Positive photoresist is then spin-coated uniformly onto the substrate. A photomask is placed between the substrate and a source of ultraviolet light, blocking selected regions from being exposed. Any photoresist exposed to the light is degraded and removed. A layer of Au is evaporated onto the entire wafer. The remaining photoresist is lifted off, leaving behind metallized regions in the desired pattern.

Following a successful lithography process, the wafer is diced into individual chips, each containing a full set of electrodes for a single DNA-NT array.

3.1.2 Deposition and Functionalization of CNTs

The vapor sensors used in this thesis work were fabricated using a solution of 98% semiconducting CNTs suspended in an aqueous solution purchased from NanoIntegris Inc. (Quebec, Canada). CNTs dissolved in pure water will tend to aggregate together into bulk carbon precipitate that exhibits metallic rather than semiconducting properties. To prevent this, the solution contains a water-soluble surfactant that keeps the CNTs separated. The CNTs themselves are grown using chemical vapor deposition and placed in solution. Using a process known as density gradient ultracentrifugation, the solution is spun in a centrifuge at 200,000 rpm over the course of several days to separate CNTs according to diameter (Figure 3.2).[1] Extracting the proper layer of CNTs from the density gradient yields high-purity semiconducting CNTs with a relatively uniform distribution of CNT diameters.[1]

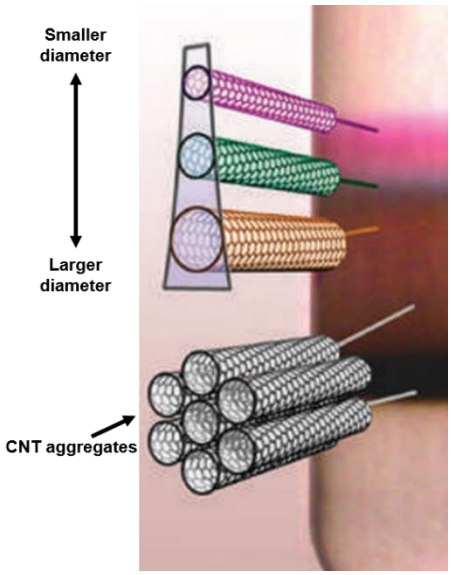

Figure 3.2. Density gradient ultracentrifugation is used to separate CNTs according to diameter and size with larger diameter CNTs settling toward the bottom. Image adapted from [1].

A special procedure is used to deposit CNTs across the source-drain channels of each sensor array (see Appendix D for a full description of the process). Approximately 140 µL of 98% semiconducting CNT solution from NanoIntegris is pipetted onto the source-drain regions on each chip and placed in a humid environment for 30 minutes (Figure 3.3). A fraction CNTs gradually self-assemble onto the substrate while the humidity prevents the solution from evaporating too quickly. This procedure is designed to yield a relatively uniform CNT network of an appropriate density (Figure 3.4). The residual CNT solution is washed away with isopropanol to remove residual surfactant and then again with deionized water to rinse away the isopropanol. The substrate is then annealed at 165°C on a hotplate for one hour to improve the contact between the CNTs and the electrodes, effectively lowering the electrode contact resistance.

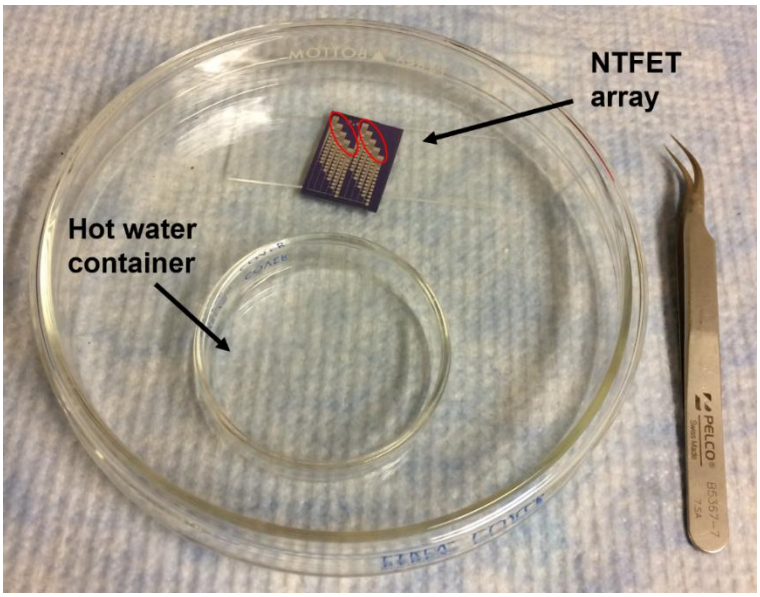

Figure 3.3. A sensor array is placed inside an enclosed volume together with a hot water container to humidify the environment. CNT solution is deposited across the FET channel

regions (outlined in red). The humidity prevents the solution from evaporating while the CNTs gradually accumulate onto the substrate.

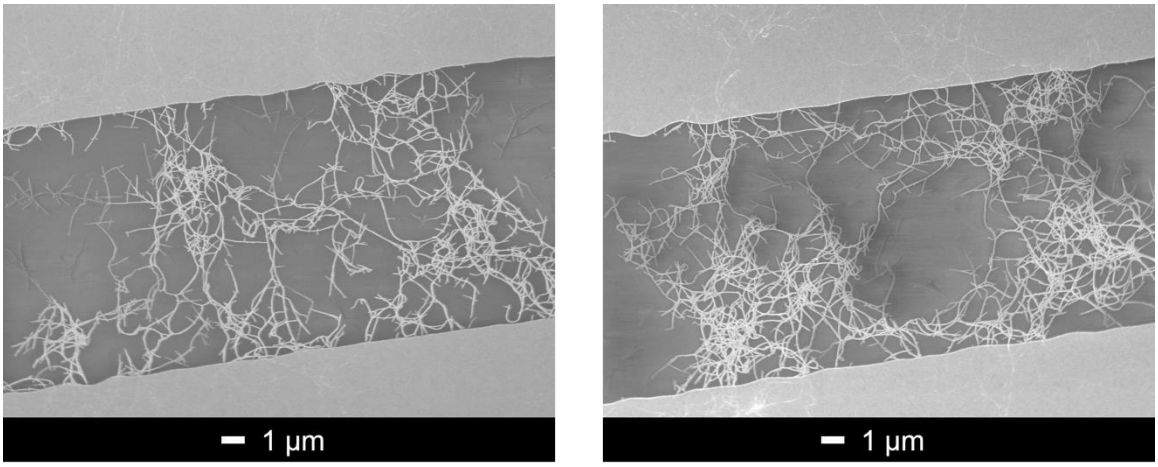

Figure 3.4. Scanning electron microscope images of two representative NTFET channels reveal disordered CNT networks forming conductive pathways from source to drain electrodes positioned above and below.

A similar procedure is used to functionalize the NTFETs with single-stranded DNA of a particular base sequence (see Appendix E for a full description of the process). Ten customized DNA oligomers are purchased from Invitrogen (Carlsbad, CA), each a 21- or 24-base oligomer of randomized sequences (Table 3.1).

Table 3.1. Ten single-stranded DNA oligomers used to chemically functionalize DNA-NT arrays. Each oligomer is a randomized sequence of 21 or 24 nucleotides.

Oligomer Name	Base Sequence
Seq1	GAGTCTGTGGAGGAGGTAGTC
Seq2	CTTCTGTCTTGATGTTTGTCAAAC
Seq3	GTACGGACTGTGAATGCGCGTTAG
Seq4	CCCGTTGGTATGGGAGTTGAGTGC
Seq5	GCGCATTGGGTATCTCGCCCGGCT
Seq6	GTATCTAGAGCGGGCGGGTACTCC
Seq7	AAACAAATCTAATAATACTTCCCA
Seq8	AGTTCGGCATGTGGAAACTCCTTC
Seq9	CGCCTAGAGGTCAAGCGTGGTTGC
Seq10	TGAAAGTGGGAAGCGACACGATGG

The DNA is diluted to 100 µM in deionized water, aliquoted, and stored in a -20°C freezer until use. One aliquot of each sequence is thawed and a 2 µL droplet is carefully pipetted onto each of ten source-drain regions on a DNA-NT array (Figure 3.5). As before, the array is placed in an enclosed humid environment for 30 minutes, allowing the DNA oligomers to settle out of solution and adsorb onto the CNT surfaces. The droplets are then removed from the substrate using compressed nitrogen.

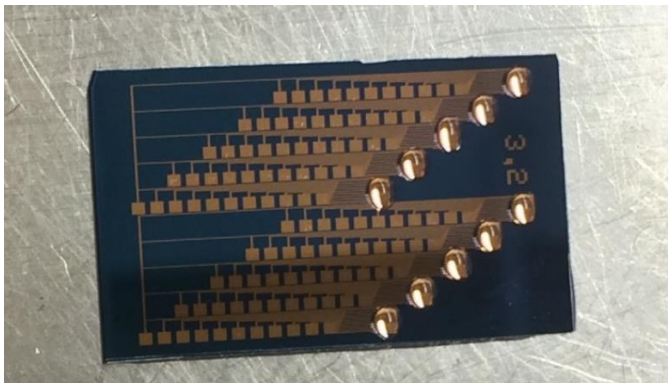

Figure 3.5. Ten groups of source-drain regions are functionalized with one of ten oligomers of single-stranded DNA. 2 µL droplets are carefully pipetted onto each region, allowing the DNA to gradually settle out of solution onto the CNT network on the substrate. This yields a DNA-NT hybrid structure only several nanometers in thickness.

3.1.3 Electrical Characterization of NTFET Arrays

DNA-NT arrays are characterized by measuring current-gate voltage (I-V_G) characteristics using a FormFactor MPS150 probing station. Conductive silver paint is applied to the underside of the array to serve as a back-gate contact. The array is placed on a metal stage to which is a gate voltage is applied. A customized probe card from Amprobe (Everett, WA) is used to interrogate each contact electrode, routing source-drain currents to a Keithley 6485 picoammeter as V_G is swept from -20 V to +20 V with a fixed source-drain bias of 0.1 V. Normally, currents converge towards a particular on-state current for increasingly negative V_G and tend towards a near-zero off-state current for positive V_G (Figure 3.6a). The overall quality of each array is assessed based on distributions of DNA-NT on-state currents and on-off ratios. Ideally, on-state currents are significantly larger than the noise threshold and generally do not exceed 100 μA.

A measure of a DNA-NT's sensitivity to changes in its electrostatic environment is the transconductance, $g_m = \frac{dI}{dV_G}$. Chemical gating due to adsorbed VOCs effectively shifts V_G from its nominal value which changes the device current by $\Delta I \cong g_m \Delta V_G$. V_G is fixed such that $|g_m|$ is maximized for the greatest number of DNA-NTs (Figure 3.6b).

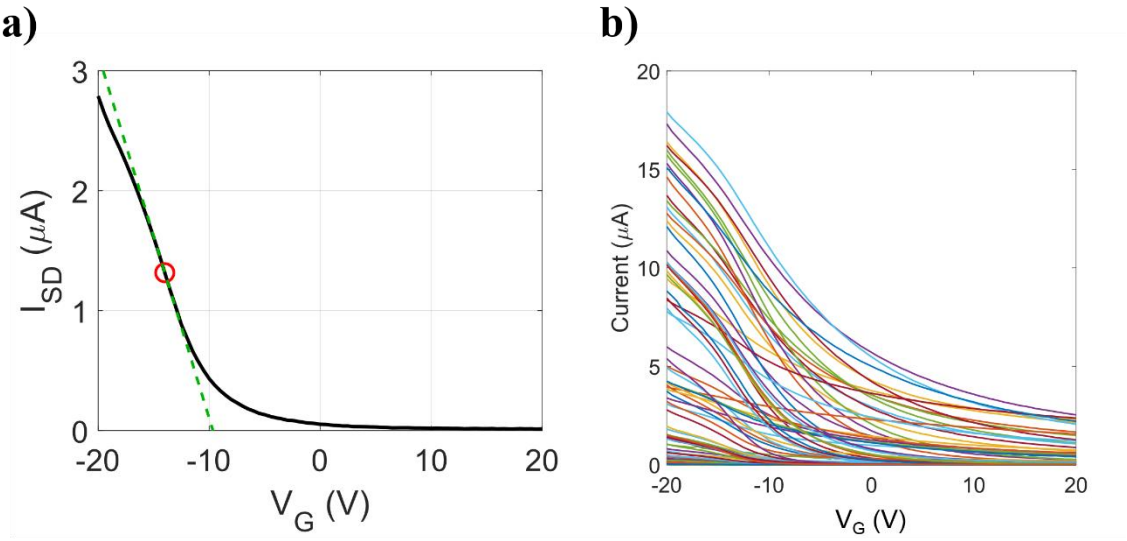

Figure 3.6. (a) I-V_G curve for a single DNA-NT shows the relationship between the device current and the applied gate voltage. The diagonal green line indicates the steepest slopes in the I-V_G curve, indicating the gate potential ($\mathbf{V_G = -14V}$) at which the corresponding device current is maximally sensitive to variations in local charge environment. (b) I-V_G characteristics for all 100 devices from the same sensor array. All DNA-NT transconductances are maximized for gate potentials between $\mathbf{V_G = -14V}$ and $\mathbf{V_G = -10V}$.

3.2 Experimental Procedure

A home-built vapor delivery system is utilized to expose sensor arrays to headspace VOCs from up to five samples at a time while simultaneously measuring device currents in real time (Figure 3.7a). Low viscosity fluids, such as propionic acid or dimethyl methylphosphonate, can be placed inside gas bubblers in which headspace VOCs are pushed out a side arm by bubbling nitrogen from beneath the liquid level (Figure 3.7b). Blood plasma samples have much higher viscosity and are available in low volume (1 mL

or less), so they are instead loaded into two-neck round-bottom flasks (Figure 3.7c). Due to the relatively low vapor pressures of plasma VOCs at room temperature, the flasks are placed in a 50 °C hot water bath to increase the VOC headspace concentration. Micro stir bars are included with fluid samples to further promote the release of dissolved VOCs. High purity compressed nitrogen gas is delivered to the sample flasks or bubblers to carry the headspace VOCs to a stainless steel sensor chamber housing a DNA-NT array. The chamber is simultaneously humidified at a fixed relative humidity with water vapor generated from a bubbler filled with deionized water to enable VOC adsorption onto the DNA binding sites. The chamber is fitted with an intake and exhaust outlet and is open from atop to allow placement of a DNA-NT array inside (Figure 3.7d). An acrylic lid fits over the top of the chamber and is compressed against an O-ring with screws to prevent gas leaks. The lid is equipped with feed-through pogo pins that make electrical contact with each DNA-NT, routing current outputs to a National Instruments PXI-2535 switching matrix used to cycle through each output to be individually measured by a Keithley 6485 picoammeter. The base of the chamber is charged to the desired gate voltage using a Keithley 617 electrometer. The associated flow rates are regulated using six MKS 1179 computer-controlled mass flow controllers (MFCs), one for each of the five samples and the sixth for the water bubbler. The various components are connected together with ¼" Teflon tubing. Portions of the tubing are heated using a heating tape to prevent VOC adsorption inside the lines. Three-way solenoid valves (ASCO Valve 8320G202) are used to select the VOC headspace of one sample to be measured at a time. These valves are actuated electronically via 120VAC solid state relays that are controlled with a computer.

A LabVIEW program is used to automate the proceedings of the experiment once it has been set up. The program adjusts MFC flow rates and switches valve states according to a flow recipe, a precompiled set of instructions that synchronize with a timetable. The software also controls bias and gate voltages applied to the sensor array and records source-drain currents from the picoammeter. This level of automation allows the system to run for many hours without user input.

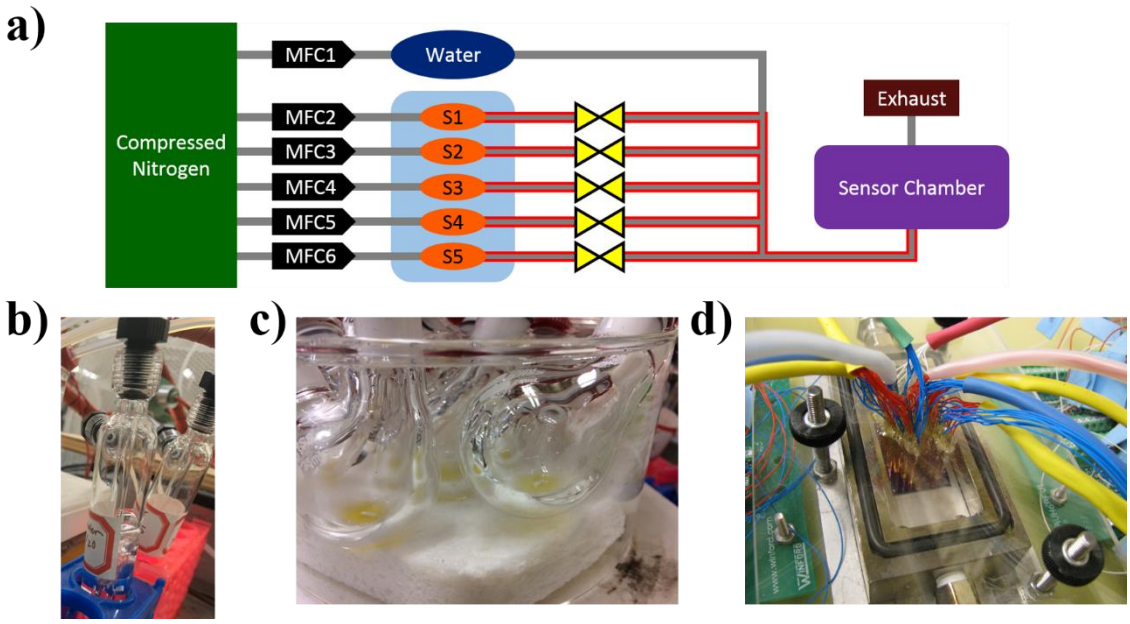

Figure 3.7. (a) Schematic overview of the vapor delivery system. Compressed nitrogen is used as a carrier gas to deliver VOCs in the headspaces of samples (S1 – S5) to a sensor chamber housing a DNA-NT array. The chamber is humidified with water vapor to enable VOC interaction with the DNA. Teflon tubing is used to connect the various components together. The gas flow rates are regulated using mass flow controllers (MFCs) controlled automatically using a computer program. The samples are heated in a hot water bath to promote the generation of VOCs. The tubing between the samples and the sensor chamber is also heated (red outline) to prevent VOC condensation inside the lines. Computer-controlled valves (yellow structures) are

used to select headspace VOCs from one sample to be measured at a time. (b) Two gas bubblers, one filled with deionized water for hydration (front). (c) Round-bottom flasks containing blood plasma (pale yellow regions) are heated in a hot water bath. Magnetic stirring rods (circled in red) are included to agitate the plasma. (d) Sensor chamber fitted with contact pins to read out currents from the array of 100 DNA-NTs inside. Three screws compress the top acrylic surface against an O-ring, preventing VOCs from leaking outside.

3.3 Data Acquisition Protocol and Data Processing

The protocol for taking measurements includes a pulse phase, in which a DNA-NT array is exposed to a mixture of target VOCs and water vapor diluted in nitrogen gas, and a refresh phase, where the mixture is replaced with a stream of nitrogen at the same relative humidity (RH) and total flow rate (Figure 3.8a). The purpose of the refresh is to carry away residual analytes from the sensor chamber, allowing the DNA-NT currents to recover to their baseline value. DNA-NT conductances vary significantly from device to device, hence the distribution of baseline currents can span several orders of magnitude (Figure 3.8b). However, fractional current responses to target VOCs are found to be comparable – hence, responses are generally reported as fractional current shifts from baseline, $\Delta I/I_0$. DNA-NT baseline currents are not constant with time, but rather drift in response to changing ambient conditions. We account for this drift by fitting a baseline to each pulse measurement using linear interpolation (Figure 3.8c). The fractional current response for each DNA-NT is then computed as:

$$\frac{\Delta I}{I_0} = \frac{I_{Measured} - I_{Baseline}}{I_{Baseline}} \times 100\% \qquad (3.3.1)$$

Baseline-fitted responses from all DNA-NTs functionalized with the same sequence of DNA are averaged together to produce a representative response for that sequence (Figure 3.8d). Variations on the mean value of $\Delta I/I_0$ due to sources of random error are nominally reduced by a factor of $1/\sqrt{n}$ when averaging responses from n DNA-NTs, thus revealing the advantage of devoting many DNA-NTs to each DNA sequence.

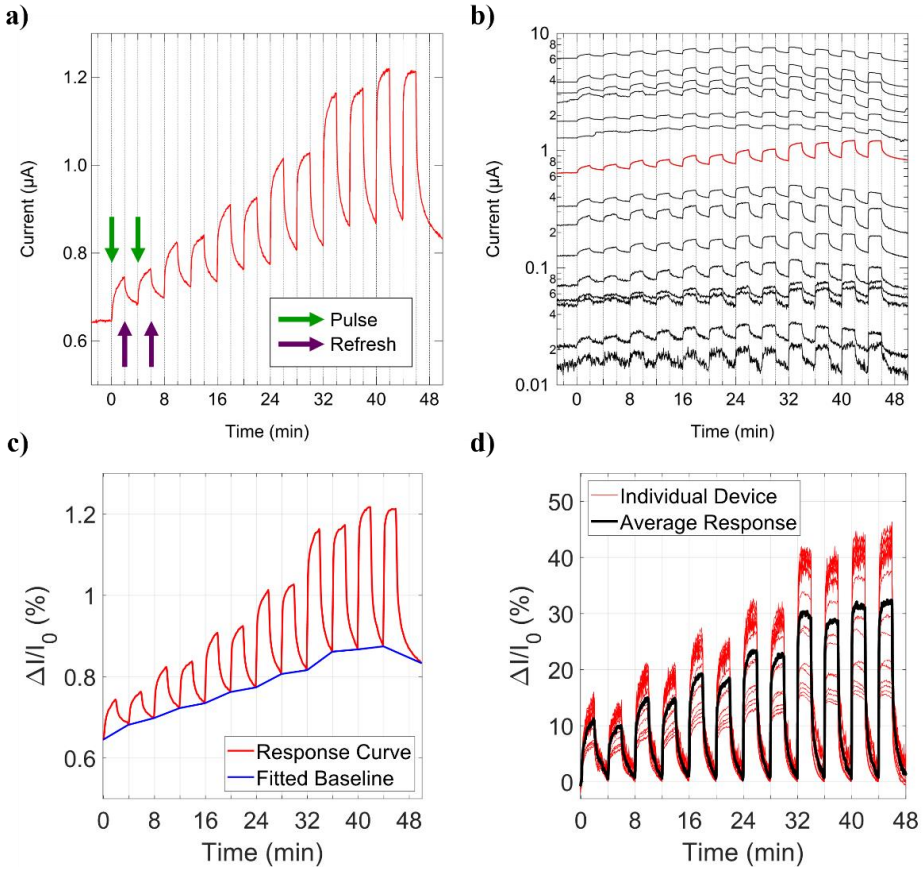

Figure 3.8. (a) A single DNA-NT is exposed to a series of two-minute pulses of propionic acid mixed with nitrogen at 33% relative humidity (RH). After each pulse, the DNA-NT is flushed with nitrogen at the same RH for two minutes, allowing the current to recover. The concentration of propionic acid vapor is incrementally increased after every other pulse from 62 ppm to 3100 ppm with each increase yielding larger associated responses. (b) Responses from the device in (a) (shown in red) are compared to those of 16 other DNA-NTs from the same array (black). Currents

are plotted on a logarithmic scale to show that the fractional responses across all devices are comparable, even though the magnitudes of baseline currents span several orders of magnitude. (c) The current response curve from (a) is baseline-fitted using linear interpolation to account for baseline drift. (d) $\Delta I/I_0$ for all 17 devices (red traces) with their average superimposed (black). The average fractional response exhibits high signal-to-noise ratio.

References

1. Arnold, M. S.; Green, A. A.; Hulvat, J. F.; Stupp, S. I.; Hersam, M. C., Sorting carbon nanotubes by electronic structure using density differentiation. *Nat Nanotechnol* **2006,** *1* (1), 60-5.

CHAPTER 4: Early-Stage Detection of Ovarian Cancer

This chapter presents the use of our e-nose technology to screen blood plasma samples for volatile indicators of ovarian cancer. Samples were collected from patients with a malignant form of ovarian cancer, those suffering from benign ovarian lesions, and healthy age-matched women. DNA-NT sensor arrays were exposed to headspace VOCs of each sample. The ensuing sensor responses were processed using statistical analysis tools which revealed the sensor platform's capacity to differentiate between metabolic VOCs characteristic of healthy individuals and of those suffering from ovarian cancer.

Section 4.1 proposes a motivation for developing a reliable screening test for ovarian cancer, particularly for early-stage cancer when treatments are the most effective. Currently, there are no accepted screening tests for ovarian cancer for women of average risk. Clinical procedures that reveal ovarian tumors usually do so in the late stages of cancer when it is often too late to save the life of the patient. Furthermore, these tests exhibit high false positive rates which puts many patients through unnecessary treatment that causes significant risks for the patient. Remarkably, several studies have shown that sniffer dogs can be trained to differentiate between VOCs from blood plasma and other bodily fluids collected from cancer-suffering patients and healthy individuals. E-nose sensor platforms based on highly sensitive nanomaterials also show promise in this application, particularly for detection of early-stage cases where VOC biomarker concentrations are extremely low.

Section 4.2 describes experiments involving blood plasma samples procured from (1) patients suffering from malignant ovarian lesions, (2) patients with benign tumors, and

(3) age-matched healthy controls. Initial experiments involved three pools of plasma samples representing the malignant, benign, and healthy groups. A DNA-NT was functionalized with four different DNA oligomers and then exposed to headspace VOCs from the pools. The sensor groups produced diversified responses to the VOCs with some groups demonstrating promising differentiation between VOCs associated with malignant, benign, and healthy groups. Additional measurements were conducted using a DNA-NT functionalized with all ten DNA oligomers to probe headspace VOCs from plasma collected from 58 individual people. Of these, 21 had a malignant form of ovarian cancer, 16 suffered from a benign from of cancer, and 21 were healthy controls. A special experimental protocol was developed specifically for measuring plasma VOCs to ensure reproducibility and signal quality of DNA-NT responses. Once all 58 samples were measured, the 10-channel DNA-NT responses were dimensionally reduced using linear discriminant analysis (LDA) yielding well-separated response distributions for the healthy, benign, and malignant groups.

Section 4.3 discusses predictive modeling based on machine learning algorithms to enable association of DNA-NT response data with healthy or diseased individuals. Classifier models based on four machine learning algorithms – LDA, support vector machine (SVM), *k*-nearest neighbors, and random forest – were constructed using the plasma response data as inputs. Each classifier was evaluated based on predictive accuracy and robustness using two different cross-validation techniques. All four classifiers demonstrated high validation accuracies and relatively low false-positive rates, signifying good progress towards a reliable and practical screening technology for ovarian cancer.

4.1 An Interdisciplinary Effort Towards Reliable Ovarian Cancer Screening

Ovarian cancer is the fifth leading cause of cancer deaths in United States women. In 2020 alone, an estimated 13,940 women will have died from ovarian cancer while the number of new diagnoses is projected at 21,750.[1] This staggering death toll is largely due to a complete lack of screening methods for accurate and early detection of the disease. The symptoms of ovarian cancer are diverse, non-specific, and commonly mistaken for those of other diseases such as various gastrointestinal illnesses.[2] More than 80% of diagnoses are associated with late-stage ovarian cancer at which point treatments are no longer effective – as a result, roughly 70% of patients with late-stage ovarian cancer die within 5 years.[2] On the other hand, if discovered early, ovarian cancer can be treated effectively with a 90% survival rate.[3] Any practical test that can more successfully detect ovarian cancer in its early stage could significantly increase patient survivability.

Modern medical diagnostics are largely based on some form of visualization, auscultation, or palpation. Currently, discoveries of ovarian tumors are largely achieved using some form of medical imaging and clinical laboratory tests. Ultrasound and cancer antigen 125 (CA-125) blood tests were once proposed as screening tests for ovarian cancer.[4] Ultrasound imaging attempts to image tumor masses around a patient's ovaries while CA-125 blood tests can identify individuals that have either healthy or elevated levels of the CA-125 protein which could indicate the presence of cancer. Unfortunately, both techniques suffer from high false positive rates. Ultrasound cannot distinguish tumors from

noncancerous masses and there are many noncancerous conditions that also increase levels of CA-125 in blood such as endometriosis and pelvic inflammatory disease. As such, neither technique is considered reliable screening techniques for ovarian cancer.[5]

In previous centuries, olfaction also played a vital role as doctors would often make diagnoses based on their own personal senses of sight, touch, and smell. From an evolutionary point of view, olfaction serves as a mechanism for social animals that provides a means of communicating information about an individual to others of the same species. A diseased individual will sometimes give off a foul odor, causing alarm to its surrounding company. Unfortunately, a human's sense of smell is largely inferior to other animals', and with the emergence of positron emission tomography, ultrasound, liquid biopsy and other powerful medical aids, olfaction-based medical practices have been falling by the wayside. Only recently have researchers begun revisiting screening strategies based on olfaction.[6-7] There have been reports of pet dogs behaving differently around their owners who were later diagnosed with some form of cancer.[6, 8] It is believed that this reaction was triggered by the dogs' ability to detect minute abnormalities in the VOC content of bodily odors given off by their owners.[8] Though seemingly miraculous, there exist scientific theories that offer explanations for remarkable behavior. One such theory is that the metabolic byproducts due to cell division differ significantly between tumor cells and their healthy counterparts since tumor cells multiply more rapidly and have a higher metabolic rate. Some of these byproducts are VOCs that are absorbed into the blood stream and distributed to other parts of the body where they can appear in other bodily fluids such as sweat, saliva, urine, or tears. Another explanation for this apparent change in VOC

content from bodily fluids is due to an immune response to the presence of cancer rather than the cancer itself. Numerous studies have validated canines' ability to detect VOC biomarkers associated with a variety of cancers including breast[6], bladder[9], and ovarian[8] from breath, urine, and blood samples, respectively.

A collaboration between the University of Pennsylvania and the Monell Chemical Senses Center in Philadelphia is using three approaches to research the differences in odor signature in bodily fluids between women diagnosed with malignant ovarian cancer, those identified as having benign ovarian tumors, and age-matched healthy individuals. The first approach, directed by Dr. Cynthia Otto at the Penn Working Dog Center, involves training dogs to identify blood samples as healthy or non-healthy (i.e. cancerous) by sniffing the VOCs emitted by blood samples collected from individual people. The second approach focuses on using gas chromatography / mass spectrometry (GC/MS) to analyze the chemical makeup of characteristic VOCs of cancerous blood samples versus healthy samples. The third approach is to develop a laboratory platform based on DNA-NT arrays for ovarian cancer screening which is the scope of this dissertation.

Samples for this work were procured by our collaborator Professor Janos Tanyi through the Ovarian Cancer Center at the University of Pennsylvania Hospital. To ensure VOCs were retained for analysis, all samples were spun at 3000 rpm for ten minutes at 4°C and then stored frozen at -80 °C. The samples were later defrosted on ice and then divided into aliquots of 500 µL. The aliquots were distributed among all collaborators so that all samples could be analyzed using the sniffing dogs, GC/MS, and DNA-NT e-nose methods.

The aliquots were refrozen at -30 °C and stored until needed. For experiments, samples were thawed and then immediately analyzed by the DNA-NTs.

4.2 DNA-NT Measurements of Headspace VOCs from Blood Plasma

4.2.1 Initial Investigations Using Pooled Plasma

The ovarian cancer project was initiated by Dr. Nicholas Kybert, a former Johnson PhD student. A preliminary experiment involved three pools of blood plasma procured from women with a malignant form of ovarian cancer, women suffering from benign tumors, and healthy age-matched controls. Each pool contained 500 µL of plasma from ten individual women. The pools were loaded into three separate 25 mL two-neck round-bottom flasks that were each placed in a 50 °C hot water bath to promote VOC generation. Each pool was also agitated with a micro stir bar to further encourage the release of VOCs. The flasks were connected to the other e-nose components with Teflon tubing wrapped with a heating element to inhibit VOC condensation. An early-generation array of 80 DNA-NTs was functionalized with four distinct sequences of DNA (Seq1, Seq2, Seq4, and Seq5) to probe the VOC headspaces of the three pools.

Once the samples were brought to temperature, a 30-minute waiting period was issued to allow VOCs to accumulate within the headspaces of the flasks. The sensor array was then exposed to VOCs from the malignant pool diluted in nitrogen at a 1:4 volume ratio. Because plasma is mostly composed of water, the nitrogen component of the mixture was humidified to 100% relative humidity (RH). The malignant pulse was sustained for two minutes, after which the DNA-NTs were refreshed with 100% RH nitrogen for two

additional minutes. VOCs from the benign and healthy pools were measured in the same manner. Three measurements were completed for each pool.

Exposure to plasma VOCs caused a negative shift in DNA-NT currents (Figure 4.1). Repeated measurements of each sample were closely reproduced and the signal-to-noise ratios of averaged responses were well above the noise threshold. Seq1 sensors demonstrated a remarkable ability to discriminate between the plasma headspaces of each pool (Figure 4.1a). Notably, responses to healthy VOCs were consistently the largest in magnitude and those associated with malignant pool were always the smallest with the benign responses in between. Unlike Seq1, Seq4 responses to malignant, benign, and healthy VOCs were far less diverse and hardly demonstrated any ability to distinguish any of the pools. The remaining sensors demonstrated some differentiation capacity, though neither were as pronounced as Seq1 (Figure 4.1b). Overall, the diversity of responses from only four sensor types suggested a strong potential of DNA-NT arrays to discriminate between ovarian cancer biomarkers in a complex plasma headspace.

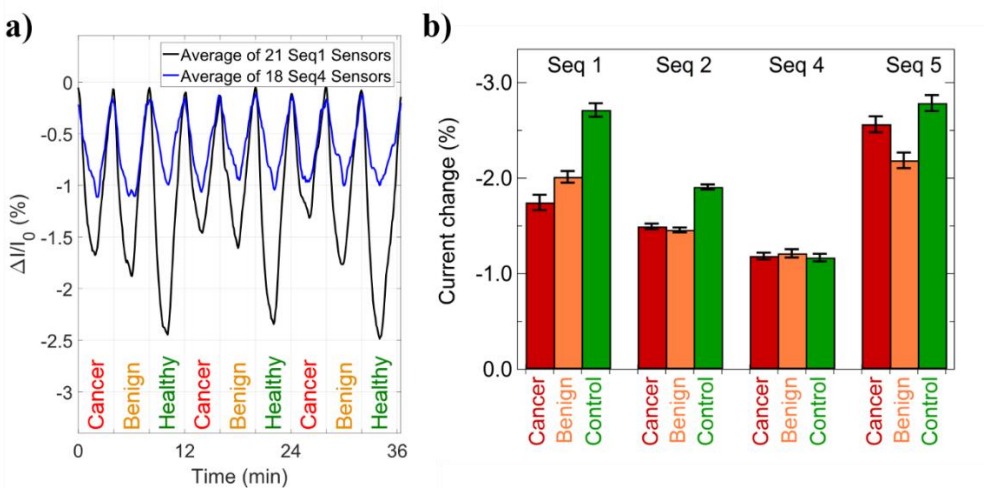

Figure 4.1. (a) Averaged sensor responses ($\Delta I/I_0$) of 21 Seq1 and 18 Seq4 DNA-NT devices exposed to VOCs from the cancerous, benign, and control pools. Three cycles of responses for

each pool are shown to show strong reproducibility between repeated measurements. Seq1 responses demonstrate a strong ability to differentiate between the pools with the healthy pool showing the largest responses, malignant showing the smallest, and the benign in between. In contrast, Seq4 responses show almost no ability to distinguish them. (b) Responses to all four DNA sequences exhibit a diversified olfaction signature of the pooled samples.

4.2.2 Measurement of Blood Plasma from Individuals

Follow-up measurements were conducted on additional plasma samples collected from 58 different people. Of these, 21 had a malignant form of ovarian cancer, 16 suffered from a benign from of cancer, and 21 were healthy controls. Six of the malignant samples were from women suffering from Stage I or Stage II cancer while the remaining were from women with Stage III or IV cancer. As before, each sample was loaded into 25 mL round-bottom flask and heated in a 50 °C water bath. DNA-NT arrays were coated with all ten DNA oligomers listed in Table 3.1 (including the four used in the pooled experiment) and then exposed to VOCs according to the same two-minute pulse and two-minute refresh protocol used previously (Figure 4.2a). Furthermore, responses from all sensor types formed three clusters corresponding to the malignant, benign, and healthy groups (Figure 4.2b). These clusters overlapped significantly, indicating that effective discrimination of VOCs would require the combined efforts of multiple DNA-NT sensor types. This variation of responses within each cluster is attributed to biological differences between individual people.

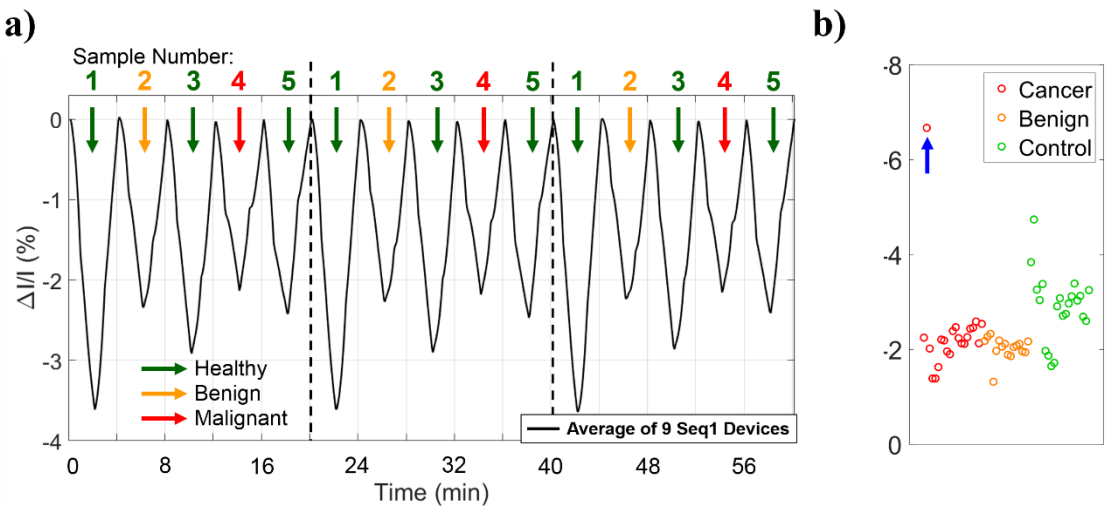

Figure 4.2. (a) Average of 9 Seq1 DNA-NT responses exhibit differentiation of headspace VOCs from plasma samples derived from three healthy individuals, a patient with benign ovarian cancer, and a patient with a malignant form of cancer. Three cycles of measurements are shown to demonstrate the reproducibility of responses between repeated measurements. (b) Seq1 responses for all 58 samples reveal clustering of data within malignant, benign, and healthy groups save for one malignant outlier (indicated by the blue arrow). Data points are displayed on an arbitrary horizontal axis to aid in visualizing overlapping points. The clusters overlap significantly, indicating that further data analysis techniques are needed to improve discrimination of each group.

Improved separation between the clusters was achieved using linear discriminant analysis (LDA), a predecessor to many advanced machine learning techniques used today (Figure 4.3). LDA is a supervised learning algorithm that requires prior knowledge of data classification. Hence, the data was supplemented with healthy, benign, and malignant class labels which were provided as additional inputs for LDA. No distinction was made to treat early-stage and late-stage cancer samples at this time. Projections of data onto the first linear discriminant (LDA1) revealed three well-separated distributions corresponding to

each class. Remarkably, all early-stage malignant samples were well-situated within the malignant distribution, suggesting a strong potential for reliable early-stage detection using DNA-NT sensors.

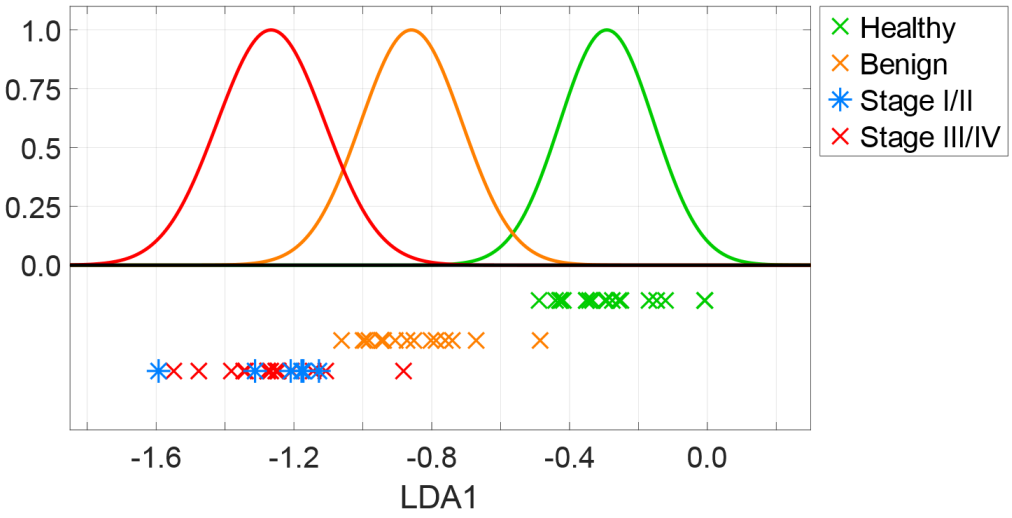

Figure 4.3. DNA-NT responses to headspace VOCs from plasma collected from 58 individuals are projected onto the first linear discriminant (LDA1) computed by linear discriminant analysis (LDA).

4.2.3 Headspace Regulation of Plasma Samples

Repeated measurements of plasma headspace VOCs caused a significant systematic depletion of DNA-NT baseline currents over the course of each experiment (Figure 4.4a). It is not fully understood why this occurs with VOCs associated with blood plasma as this phenomenon has not been observed for VOCs from other media. The current hypothesis is that this is caused by gradual accumulation of one or more "sticky" molecular components of the plasma headspace that are reluctant to desorb from the DNA-NTs,

resulting in only partial recovery of device currents after each refresh and thus attenuating DNA-NT responses over time (Figure 4.4b).

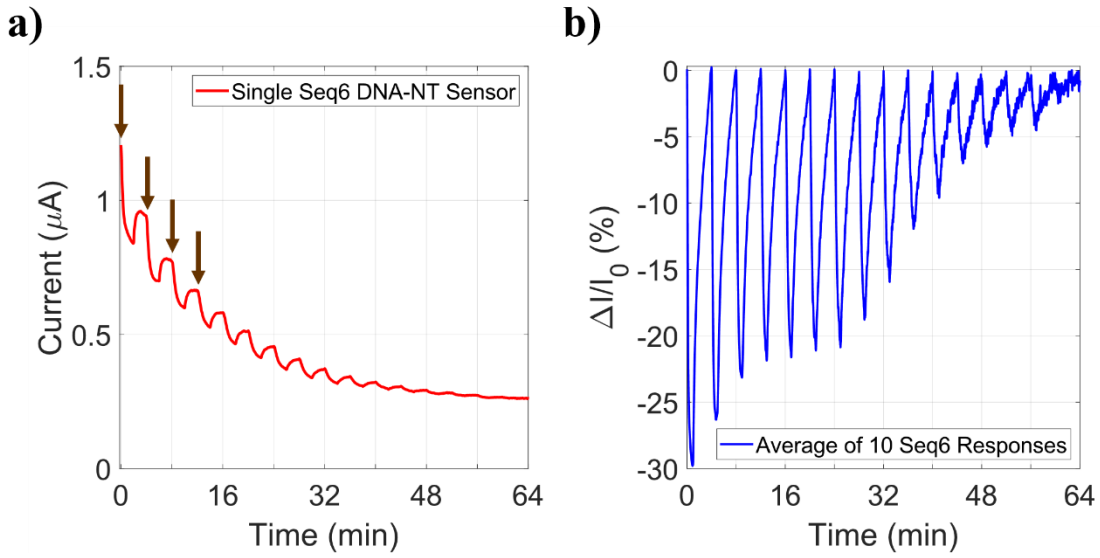

Figure 4.4. (a) A single DNA-NT sensor is exposed to a series of sixteen exposures to headspace VOCs from one blood plasma sample to demonstrate attenuation of both the baseline current and the response levels. The start times of the first four exposures are marked by the brown arrows. (b) The average current responses of all ten Seq6 devices also demonstrates signal attenuation and deteriorating signal-to-noise ratios over time.

A special experimental protocol was developed to reduce the attenuation of DNA-NT currents due to plasma VOC exposure. Typically, sample VOCs are allowed to accumulate within the headspace of the corresponding container between measurements (i.e. whenever the system is busy measuring a different sample). This was altered specifically for measuring plasma by constantly delivering nitrogen to all sample-containing flasks throughout all stages of the experiment, effectively reducing the associated headspace VOC concentrations. Three-way solenoid valves were used to direct

the flow of sample VOCs either to the sensor chamber to be measured or out a separate exhaust path (Figure 4.5).

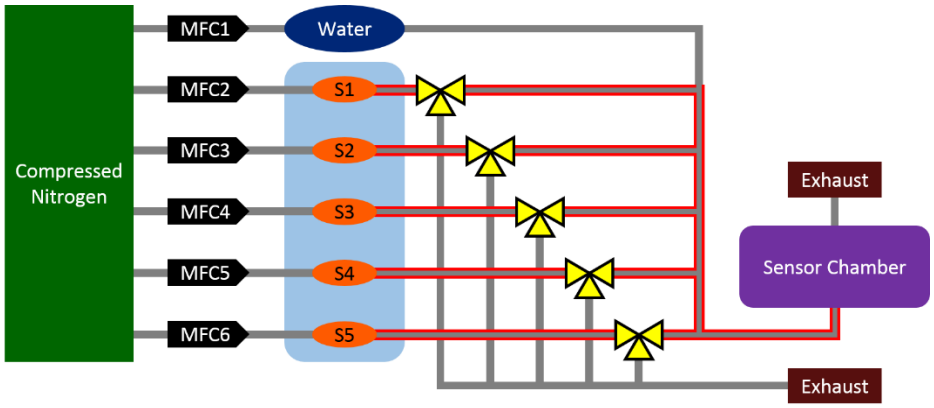

Figure 4.5. Modified gas-delivery setup for measuring headspace VOCs from blood plasma samples. Nitrogen gas is continuously delivered to all samples throughout the experiment while three-way solenoid valves (yellow structures) direct each flow either to the sensor chamber to be measured or out a separate exhaust. This decreases the headspace concentrations of plasma VOCs which slows the rate of DNA-NT response attenuation.

This technique markedly improved the reproducibility of DNA-NT responses to plasma VOCs, nearly eliminating the systematic signal drift observed earlier (Figure 4.6). In addition, signal-to-noise ratios remained high above the noise threshold even though the overall concentration of VOCs was reduced.

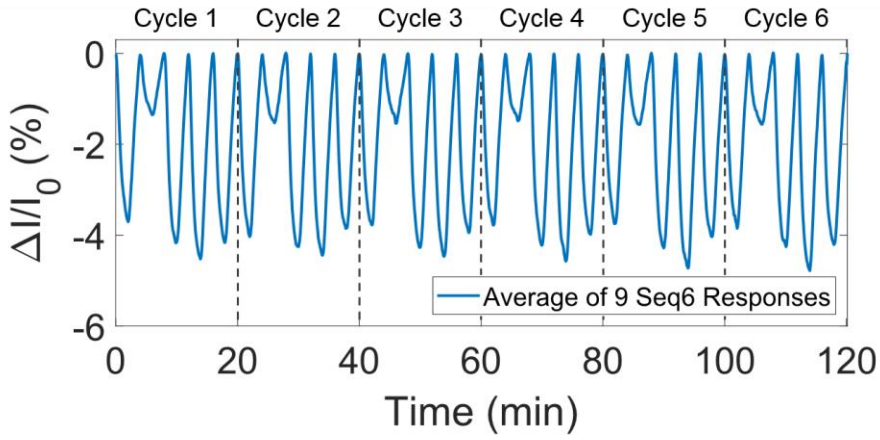

Figure 4.6. Nine Seq6 DNA-NTs were exposed to six cycles of measurements of five plasma samples in sequence. The samples were measured in the same order for each cycle. The responses demonstrate significantly improved reproducibility and a high signal-to-noise ratio.

4.3 Sample Classification Using Machine Learning

4.3.1 Sample Classification Based on Cross-Validation

Predictive modeling based on pattern recognition algorithms was performed on the chemical sensor data to enable association of DNA-NT responses with healthy or diseased individuals. Initially, an LDA-based classifier was constructed with the ovarian response data using the Scikit-learn package for the Python programming language.[10] The robustness of the classifier was verified using two cross-validation techniques: Leave One Out Cross-Validation (LOOCV) and stratified *k*-fold cross-validation. Cross-validation is often used to test for biased training of a dataset in machine-learning applications.

LOOCV is designed to test whether or not a single new data point would be correctly categorized by the classifier. This is simulated by building the classifier using all

but one of the samples in the database. The classifier is then used to predict the identity of the left-out sample. This process is repeated until each individual sample in the database has been validated. For the ovarian data set, 55 of the 58 samples were correctly identified, yielding a validation accuracy of 95% (Table 4.1). Both misclassified samples were false negatives with one malignant sample identified as benign and one benign sample identified as healthy. All early-stage samples were correctly classified as malignant.

Table 4.1. Results of LOOCV on the entire ovarian dataset. 55 of the 58 samples were correctly classified, yielding a validation accuracy of 95%. The three misclassified samples were all false negatives with two malignant samples identified as benign and one benign identified as healthy. The misclassified malignant sample was from a late-stage cancer patient.

		CLASSIFIED (LOOCV)		
		Malignant	Benign	Healthy
ACTUAL	Malignant	19	2	0
	Benign	0	15	1
	Healthy	0	0	21

Stratified *k*-fold cross-validation is a derivative of *k*-fold cross-validation in which a dataset is randomly partitioned into *k* equal-sized subsets (Figure 4.7). One partition is chosen as the validation data while the remaining $k - 1$ subsets are used as the training data. This process is repeated until all partitions have been validated once. Due to the element of randomness, it is common practice to repeat the entire *k*-fold cross-validation process multiple times so that the validation results converge towards a representative set of values. *k*-fold cross-validation does not take the classification of each data point into account which can be problematic for relatively small data sets since there is a significant chance of obtain

training sets that are underrepresented by one or two of the sample groups. A solution is to stratify the *k* partitions such that each subset comprises an equal fraction of data points from each class, ensuring that the mean response for each class is approximately equal in all partitions.

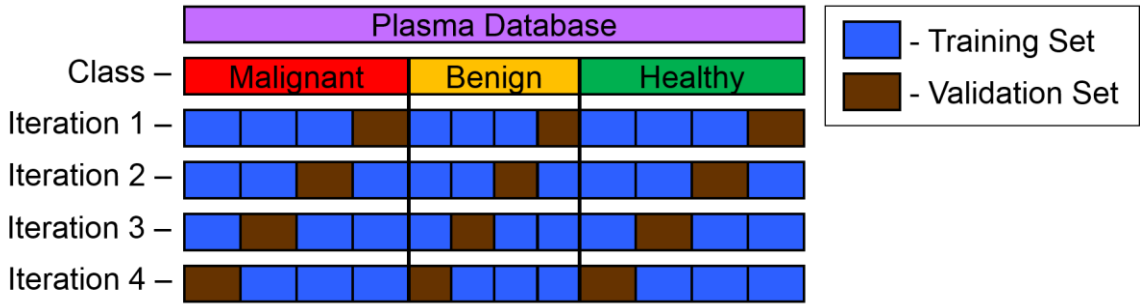

Figure 4.7. Using the stratified *k*-fold cross-validation technique, the malignant, benign, and healthy data are each randomly shuffled and then partitioned into four equal-sized subsets. Three subsets from each class are combined to form the training set while the remaining form the validation set. The process is repeated until all data points have been validated exactly once.

Stratified *k*-fold cross-validation was used to partition the ovarian plasma database into four subsets, using 75% of the data as the training set with the remaining 25% as the validation set. This process was performed 10,000 times to obtain convergent validation estimates. On average, 3.76 samples out of 58 were misclassified, corresponding to a validation accuracy of 94%. The overall false-positives rate remained low at less than 0.1%. A detailed summary of the results is presented in Table 4.2.

Table 4.2. Stratified *k*-fold cross-validation was performed 10,000 times on the ovarian database. The percentages indicate the number of correctly or incorrectly classified samples for each corresponding class. The overall validation accuracy was 94% while the overall false positives rate was less than 0.1%.

		CLASSIFIED (stratified *k*-fold)		
		Malignant	Benign	Healthy
ACTUAL	Malignant	86.0%	13.6%	0.4%
	Benign	0%	93.8%	6.2%
	Healthy	0%	0.5%	99.5%

The classification performances of three additional learning algorithms, each more sophisticated than LDA, were also investigated. Support vector machine (SVM), *k*-nearest neighbors (KNN), and random forest were each used to construct new classifiers based on the combined dataset. Each algorithm was trained and tested according to LOOCV (Table 4.3) and repeated stratified *k*-fold (Table 4.4) cross-validation methods. For the latter, 90% of the samples were used for the training set and the remaining 10% for the test set. As before, this was repeated 10 times so that each sample was in the testing set exactly once. The entire process was repeated 1000 times, shuffling the data each time to form independent trials. Again, each independent trial was stratified such that approximately equal proportions of malignant, benign, and healthy data were used for the training and testing sets for all repeats. In all cases, high validation accuracies were achieved. Despite being the least sophisticated algorithm, LDA performed the best in terms of overall accuracy, scoring around 95% for both LOOCV and *k*-fold. The other classifiers achieved accuracies of approximately 90%. This is perhaps due to the fact that SVM, KNN, and random forest are designed for larger datasets, in which case they may have been

overfitting the 58 ovarian responses and compromising the validation accuracies as a result. Remarkably, the number of false positives were consistently low in all cases with all algorithms successfully classifying 100% of all six early-stage cancer samples with the exception of random forest which misclassified one as benign.

Table 4.3. Validation performances for Leave-One-Out Cross-Validation (LOOCV) for LDA, SVM, KNN, and random forest classification algorithms. LDA classified all samples with the highest accuracy (95%) with zero false positives.

Classifier	Overall	Healthy	Benign	Malignant	False Positives	False Negatives
LDA	55/58 (**95%**)	21/21	15/16	19/21	0	3
SVM	52/58 (**90%**)	20/21	15/16	17/21	1	5
KNN	53/58 (**91%**)	21/21	15/16	17/21	0	5
Random Forest	51/58 (**88%**)	19/21	14/16	18/21	3	4

Table 4.4. Repeated stratified *k*-fold evaluation of the same four classification algorithms using 10 bins and 1000 repeats.

Classifier	Overall	Healthy	Benign	Malignant	False Positives	False Negatives
LDA	54.5/58 (**94%**)	20.9/21.0	15.0/16.0	18.6/21.0	0.1	3.4
SVM	52.1/58 (**90%**)	20.0/21.0	15.0/16.0	17.1/21.0	1.0	4.9
KNN	52.8/58 (**91%**)	21.0/21.0	15.0/16.0	16.8/21.0	0.0	5.2
Random Forest	51.7/58 (**89%**)	19.5/21.0	14.0/16.0	18.2/21.0	2.4	3.8

4.3.2 Classification Based on Receiver Operating Characteristics

The classification performance of LDA was also verified with receiver operating characteristics (ROCs). Normally, ROCs are used to characterize the predictive accuracy of a two-class system. Because there are three classes for our database (i.e. healthy, benign,

and malignant), three ROC curves were generated using one class as the "positive" signal and the other two as the "negative" signal. ROC outputs were based on stratified *k*-fold cross-validation using training / validation splits of 90% / 10% (Figure 4.8a) and 75% / 10% (Figure 4.8b). In each case, the validation accuracies for the healthy and malignant groups were nearly 100% while the benign group was validated with nearly 90% accuracy. Though the benign accuracy fell by 3% relative to the prior LDA-based *k*-fold results, the malignant accuracy rose by about 10%, marking a significant improvement for malignant classification. Furthermore, the nearly identical results reported by both sets of train / validation splits confirm the stability of the overall dataset.

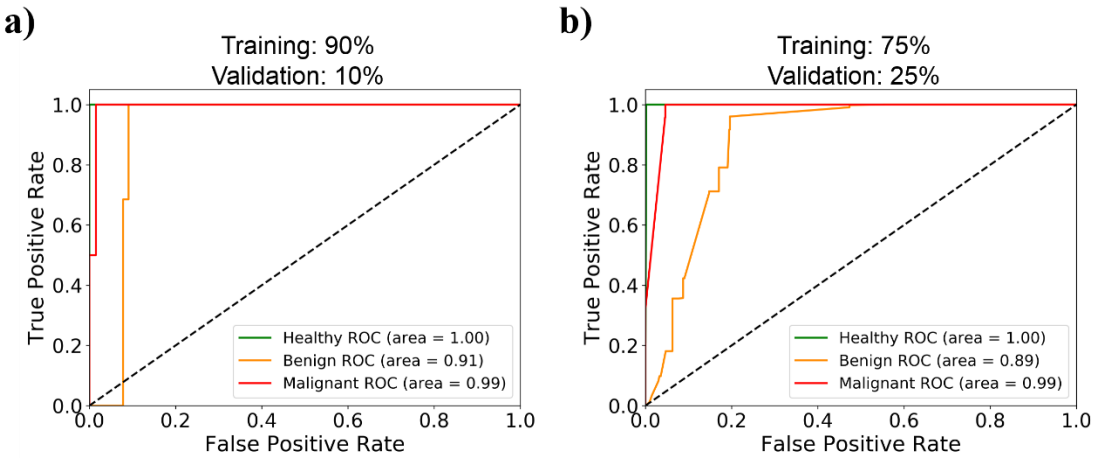

Figure 4.8. (a) ROC curves generated from stratified *k*-fold cross-validation using 10,000 repeats and a training / validation split ratio of 90% / 10%. (b) The procedure is repeated using a split ratio of 75% / 25%. The accuracies reported in both (a) and (b) are nearly identical.

In conclusion, the successful differentiation of VOCs associated with healthy and cancerous individuals indicates the strong potential of DNA-NT sensor arrays as a screening method for ovarian cancer. Particularly promising is the consistently accurate

classification of early-stage samples which would place DNA-NTs above other clinical trials. Such results support the proposed advantage of early-stage cancer screening based on VOC biomarker detection over other techniques that lack the necessary sensitivity, possibly due to lack of signal. In the case of liquid biopsy, which attempts to collect circulating tumor DNA (ctDNA), there is necessarily a lack of ctDNA within early-stage cancer patients since the tumors at this point would be relatively small. On the other hand, it is plausible that even early-stage cancer can cause a response from the host's immune system that is reflected in the VOC contents found within the patient's bodily fluids or body odor. Another clinical advantage of DNA-NT screening is the apparently low rate of false positives boasted by the pattern-recognition algorithms. In a real clinical setting, avoiding a false positive diagnosis would mean preventing unnecessary treatments like chemotherapy that are often accompanied by troublesome side-effects. Overall, the results signify the promise for this technology in comparison to currently established techniques and motivate further development to manifest an effective screening technique for ovarian cancer screening.

The performance of this technique could be enhanced by improving the various components of the e-nose system. To date, we have only explored chemical responses in the scope of ten distinct DNA sequences. By investigating new DNA oligomers, we could discover other sequences that offer orthogonal chemical sensing information that is not accessed by the original sequences. Unlike the current set of DNA oligomers, which were chosen randomly, new oligomers could be engineered to target particular VOC biomarkers for ovarian cancer discovered by GC/MS. Furthermore, testing new DNA oligomers allows

us to pick and choose the most successful sequences for this application. We intend to redesign our sensor array to increase the number of distinct sensor types from 10 to 100. This will increase the "chemical resolution" with which we probe the volatile headspaces of each plasma sample. It will also allow us to test a greater number of DNA oligomers at once in order to efficiently discover optimal sequences. We also aim to increase our sample size to better represent the chemical vapor signatures from various populations. This will enable us to utilize more specialized machine-learning methodologies to further improve the overall classification accuracy.

References

1. SEER Cancer Stat Facts: Ovarian Cancer. National Cancer Institute. Bethesda, MD, https://seer.cancer.gov/statfacts/html/ovary.html.
2. Petricoin, E. F.; Ardekani, A. M.; Hitt, B. A.; Levine, P. J.; Fusaro, V. A.; Steinberg, S. M.; Mills, G. B.; Simone, C.; Fishman, D. A.; Kohn, E. C.; Liotta, L. A., Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* **2002,** *359* (9306), 572-577.
3. *American Cancer Society, Cancer Facts and Figures 2019*. 2019.
4. SEER Cancer Stat Facts: Ovarian Cancer. National Cancer Institute. Bethesda, MD, https://seer.cancer.gov/statfacts/html/ovary.html.
5. Buys, S. S.; Partridge, E.; Black, A.; Johnson, C. C.; Lamerato, L.; Isaacs, C.; Reding, D. J.; Greenlee, R. T.; Yokochi, L. A.; Kessel, B.; Crawford, E. D.; Church, T. R.; Andriole, G. L.; Weissfeld, J. L.; Fouad, M. N.; Chia, D.; O'Brien, B.; Ragard, L. R.; Clapp, J. D.; Rathmell, J. M.; Riley, T. L.; Hartge, P.; Pinsky, P. F.; Zhu, C. S.; Izmirlian, G.; Kramer, B. S.; Miller, A. B.; Xu, J. L.; Prorok, P. C.; Gohagan, J. K.; Berg, C. D., Effect of screening on ovarian cancer mortality: the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Randomized Controlled Trial. *JAMA* **2011,** *305* (22), 2295-303.
6. McCulloch, M.; Jezierski, T.; Broffman, M.; Hubbard, A.; Turner, K.; Janecki, T., Diagnostic accuracy of canine scent detection in early- and late-stage lung and breast cancers. *Integr Cancer Ther* **2006,** *5* (1), 30-9.
7. McCulloch, M.; Turner, K.; Broffman, M., Lung cancer detection by canine scent: will there be a lab in the lab? *Eur Respir J* **2012,** *39* (3), 511-2.

8. Horvath, G.; Andersson, H.; Nemes, S., Cancer odor in the blood of ovarian cancer patients: a retrospective study of detection by dogs during treatment, 3 and 6 months afterward. *BMC Cancer* **2013,** *13*, 396.
9. Willis, C. M.; Church, S. M.; Guest, C. M.; Cook, W. A.; McCarthy, N.; Bransbury, A. J.; Church, M. R.; Church, J. C., Olfactory detection of human bladder cancer by dogs: proof of principle study. *BMJ* **2004,** *329* (7468), 712.
10. *Scikit-Learn: Machine Learning in Python*, Scikit-Learn, July 2017, https://scikit-learn.org/stable/.

CHAPTER 5: Assessment of DNA-NT Performance for Remote Monitoring Applications

This chapter presents potential uses of DNA-NT arrays in remote monitoring applications such as pollution detection and monitoring of agricultural produce. Various performance characteristics of DNA-NTs were investigated with a series of experiments designed to simulate certain complex environmental conditions related to such applications.

Section 5.1 motivates several desired qualities of ideal environmental chemical sensors, including low detection limits, detection of dilute targets in the presence of a complex background, fast chemical response readouts, and long lifetime.

Section 5.2 discusses experiments involving a freshly made DNA-NT array exposed to various concentrations of 2,6-dinitrotoluene (DNT). The array demonstrated sensitivity to parts-per-billion (ppb) concentrations of DNT. The experiment was repeated with exposure times reduced from two minutes to one minute. DNA-NT responses to the DNT were diminished but still visibly apparent after some signal processing, demonstrating high sensitivity and fast response times of our sensors.

Section 5.3 discusses follow-up experiments involving the same array, this time exposed to DNT mixed with dimethyl methylphosphonate (DMMP) to simulate a complex background. DNA-NTs successfully distinguished 152 parts-per-million (ppm) DMMP from 152 ppm DMMP mixed with 149 ppb DNT, indicating sensitivity to a dilute target even in the presence of a 1000-times concentrated background. The fresh DNA-NT array

was replaced with another identical array that was prepared four months prior and the complex background experiment was repeated. Responses to DNT and DMMP mixtures were expectedly diminished, possibly due to partial denaturing of the sensing material. Nevertheless, the older array was also able to discriminate between mixtures with and without DNT. Thus, the DNA-NTs performed desirably in terms of sensitivity, selectivity, and longevity, making them promising candidates for a wide range of environmental applications.

5.1 Introduction

There is a strong desire for chemical vapor sensors that can operate in real-world settings with numerous potential uses in agricultural[1-2], military,[3-4] and environmental monitoring applications.[5-10] Examples include assessment of ripeness or vitality of crops and produce, remote detection of mines in a warzone, and detection of pollutants in the atmosphere. In each case, chemical sensors must be able to identify target molecules in a complex chemical background. Furthermore, unlike laboratory settings where relative concentrations of VOC mixtures can be controlled precisely, the volatile backgrounds in real-world settings are constantly fluctuating which makes successful classification of VOCs more challenging. In practice, commercial sensors are often expected to collect vast amounts of data for extended periods of time, sometimes on the order of weeks or months. Because of this, sensors with low-power consumption and long lifetimes are preferred. Fast

measurement readouts are also coveted so as to capture as many VOC binding events as possible.

A series of experiments was initiated in collaboration with Rohinton Mehta's group at X Development (Mountain View, CA) to assess the performance of our DNA-NTs in real-world environments. Five separate experiments were conducted within the laboratory using the e-nose system to simulate various environmental conditions. DNA-NT performance was evaluated based on the following desired characteristics: (1) detection of low concentrations of VOC analytes down to parts-per-billion levels; (2) detection of fractional changes of analyte concentration in the presence of a concentrated complex background; (3) fast responses to analytes; and (4) long-lasting devices that can remain functional for weeks or months.

5.2 Parts-per-billion Detection of 2,6-Dinitrotoluene

The first experiment sought to demonstrate parts-per-billion (ppb) sensitivity of the DNA-NTs. 2,6-Dinitrotoluene (DNT) was chosen as the target analyte due to its relatively low vapor pressure ($5.67 \cdot 10^{-4}$ Torr).[11] Roughly 10 mL of solid DNT powder was placed in a gas bubbler and allowed ample time to generate headspace volatiles. A fresh DNA-NT array with 80% device yield was prepared the day before the experiment. The array was initially exposed to two-minute pulses of DNT mixed with nitrogen at three different concentrations: 14.9 ppb, 149 ppb, and a fully saturated concentration of 746 ppb (Figure 5.1a). Each pulse was followed by a two-minute refresh with pure nitrogen. Water vapor

was not included in the mixture to remove any doubt that the ensuing responses were caused by DNT alone. Exposure to DNT caused positive current shifts in DNA-NT currents, which was expected due to prior investigation.[12] Responses to 149 ppb and 746 ppb DNT were easily resolved. Responses to 14.9 ppb DNT were less profound and not easily resolved without additional data processing. Hence, all DNA-NT currents were baseline-fitted and then smoothed using 7-point boxcar averaging to improve signal-to-noise. This revealed visibly apparent responses to 14.9 ppb DNT.

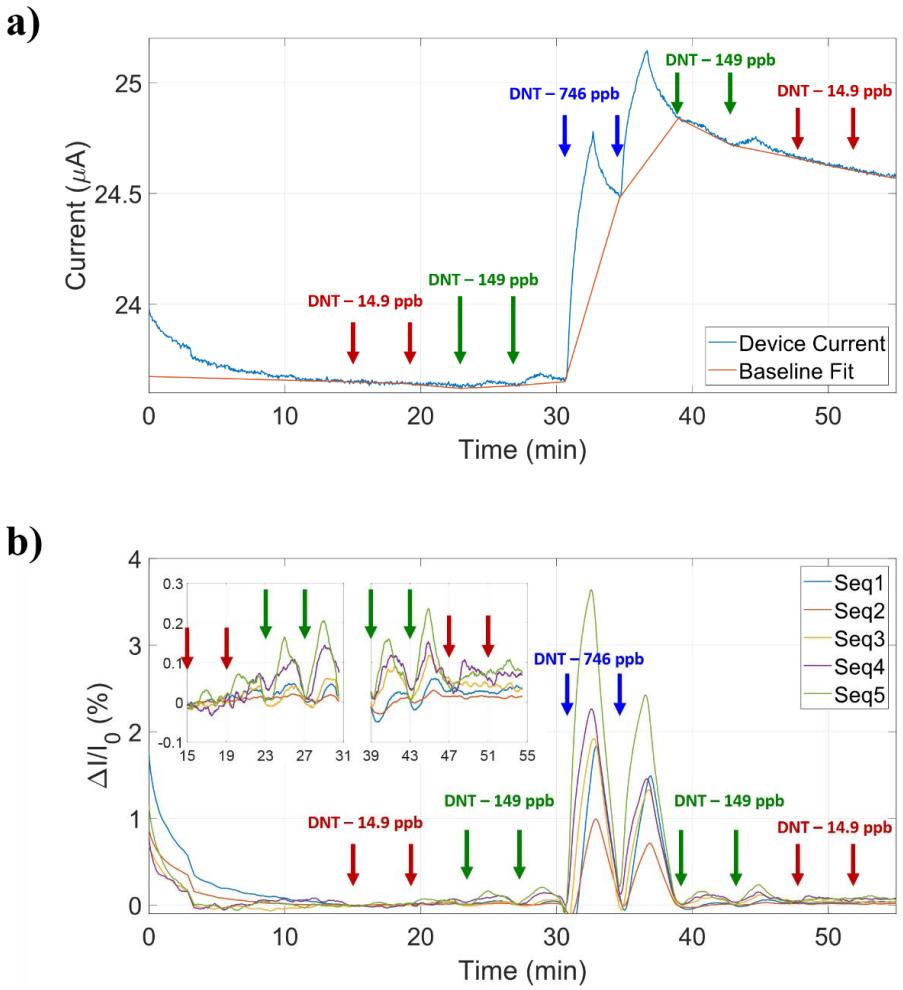

Figure 5.1. (a) A day-old DNA-NT array was exposed to 2,6-dinitrotoluene (DNT) at various concentrations. Current responses are shown for a single Seq4 DNA-NT (blue). Responses are

clear for 149 ppb and 746 ppb concentrations of DNT, but less so for 14.9 ppb. A fitted current baseline (red) was subtracted from the raw current trace to reveal responses to 14.9 ppb DNT. (b) Average current responses for five DNA-NT sensor types (Seq1 through Seq5). The other five sensor types (Seq6 through Seq10) produced visually similar results and are not shown. Responses to 14.9 ppb DNT are small, yet visibly apparent. 149 ppb and 746 ppb responses are easily resolved. All averaged current responses were smoothed using 7-point boxcar averaging.

The experiment was repeated with exposure and periods reduced from two minutes to one minute. DNA-NT currents were again baseline-fitted and smoothed using boxcar averaging. The response magnitudes were predictably reduced due to the lower exposure time. Nevertheless, Seq4 and Seq5 current responses were clearly resolved for all target concentrations while the other sensor groups also evidenced some detection capacity for 14.9 ppb and 149 ppb DNT (Figure 5.2). These results demonstrate the desirably high sensitivity and fast response capabilities of DNA-NTs.

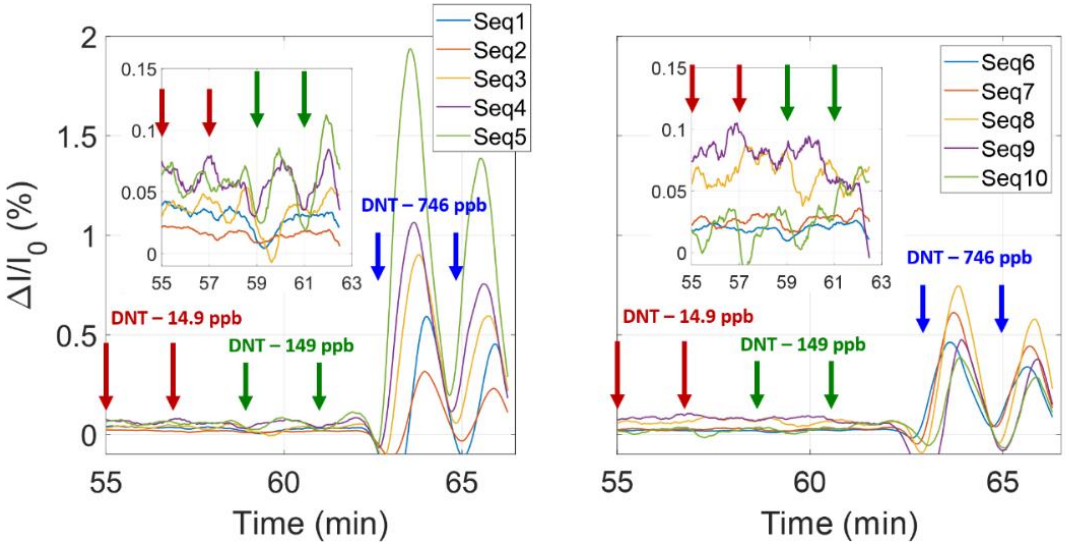

Figure 5.2. DNT exposure times are reduced from two minutes to one minute. The responses are smaller as expected. Only Seq4 and Seq5 responses demonstrate apparent detection of 14.9 ppb and 149 ppb DNT.

5.3 Detection of DNT Target in Complex Background

The second experiment simulated detection of a target analyte in a complex background. The same day-old DNA-NT array was initially exposed to a mixture of DNT with dimethyl methylphosphonate (DMMP), a compound known to have a significantly larger vapor pressure (0.579 Torr) than DNT.[13] The relative concentrations of DNT and DMMP were fixed at 149 ppb and 152 parts per million (ppm), respectively, thus creating a 1000-times diluted target-to-background mixture. After a one-minute exposure, the DNA-NTs were refreshed using 0% relatively humid (dry) nitrogen. This was followed by an exposure to 152 ppm DMMP devoid of DNT in the mixture. DMMP is known to cause negative shifts in DNA-NT currents,[12] and since the DMMP was more concentrated, the responses to both types of pulses were expectedly negative. The DNA-NT responses were baseline-fitted and smoothed as before. This revealed that the current responses of the DNT-containing pulses were shifted slightly positive relative to responses to the DMMP background alone (Figure 5.3). Given that DNT exposures elicit positive responses, this result suggested that the presence of highly diluted DNT was indeed detected in the 1000-times concentrated complex background.

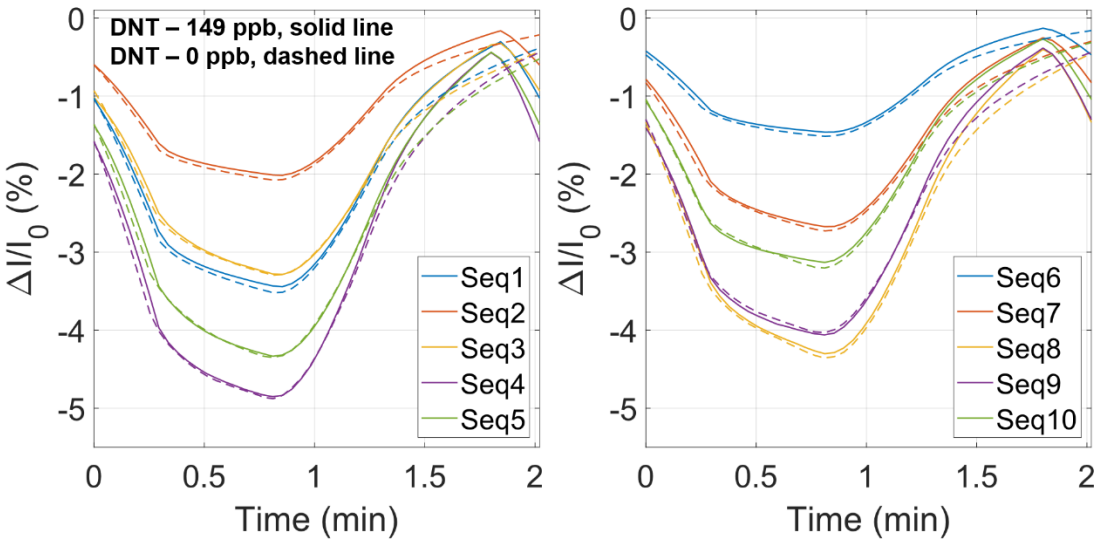

Figure 5.3. DNA-NT responses to the 152 ppm DMMP + 149 ppb DNT mixture are superimposed with responses to only DMMP for comparison. All responses were smoothed using 7-point boxcar averaging. Solid lines represent the mixture and dashed lines are for the suppressed mixture. A slight relative positive shift in response is observed for the mixture containing DNT, suggesting that the DNA-NTs detected the lack of DNT molecules which are known to evoke positive current responses.

The DNT detection in complex background experiment was repeated using a four-month-old DNA-NT array in place of the fresh array. The old array had not been used in any prior experiments and was merely kept in a plastic petri dish stored in air after it was fabricated. Up to this point, we had never observed responses of DNA-NT arrays more than a few weeks old and thus, it was not known whether the months-old array would reproduce the results from the previous experiment or if denaturing of the DNA had compromised its sensing capabilities. Despite its age, the array boasted a device yield of 78% which was near that of the fresh array. As before, a one-minute exposure to 149 ppb

DNT with 152 ppm DMMP background was issued, followed by a one-minute exposure to 0 ppb DNT and 152 DMMP. The DNA-NTs were again refreshed with dry nitrogen in between each exposure. This cycle of measurements was repeated two more times and the average responses to each mixture were compared (Figure 5.4). The overall response magnitudes for the older array were diminished by roughly a factor of ten, suggesting that some denaturing of the sensing material had occurred. Nevertheless, a relative positive shift was observed for responses to mixtures containing DNT, thus reproducing the results of the fresh array. This result suggests a long-term shelf life of DNA-NTs projected on the order of months.

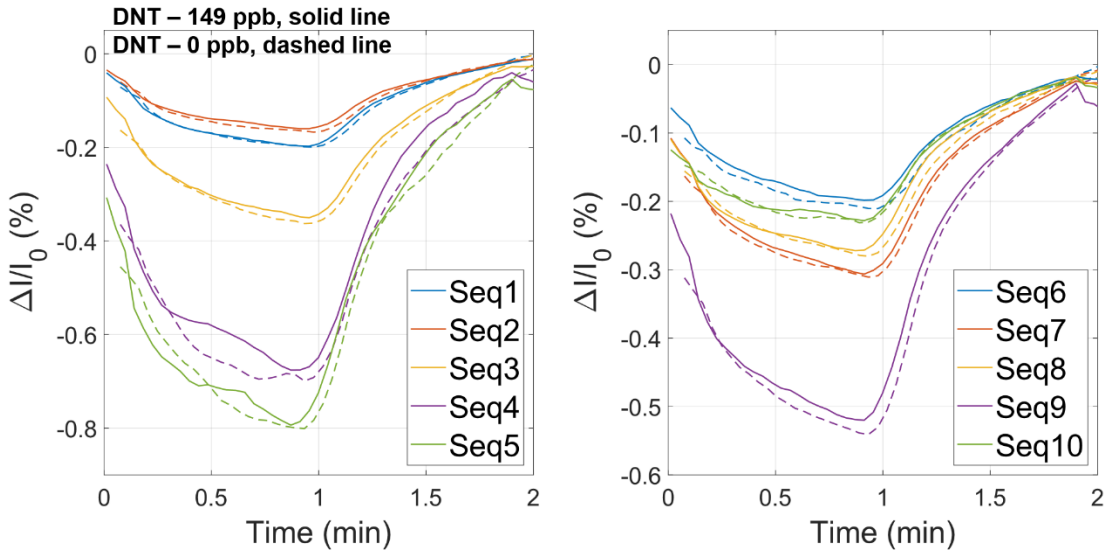

Figure 5.4. The complex mixtures experiment was repeated using an DNA-NT array stored in air for four months. All responses were smoothed using 7-point boxcar averaging. The response profiles for 149 ppb DNT (solid line) and 0 ppb DNT (dashed line) represent averages of the three responses to each mixture. The overall response magnitudes for the older array were diminished by roughly a factor of ten, suggesting that some denaturing of the sensing material had occurred. However, the same relative positive shift was observed for responses to mixtures containing

DNT, confirming that the months-old DNA-NT array was able to reproduce the results of the fresh array.

For certain applications, it may not be feasible to refresh the DNA-NTs with nitrogen after every exposure. To simulate the constant presence of a complex background, the latter experiment was repeated once more using a 152 ppm DMMP mixture as the refresh instead of dry nitrogen. The months-old array was exposed to alternating mixtures of 149 ppb DNT in DMMP background and plain DMMP background (Figure 5.5). Slight negative current shifts were observed in the absence of DNT in the mixture, validating the trend observed in the previous experiments and thus confirming the capability of DNA-NTs to detect a dilute target despite the lack of a clean refresh.

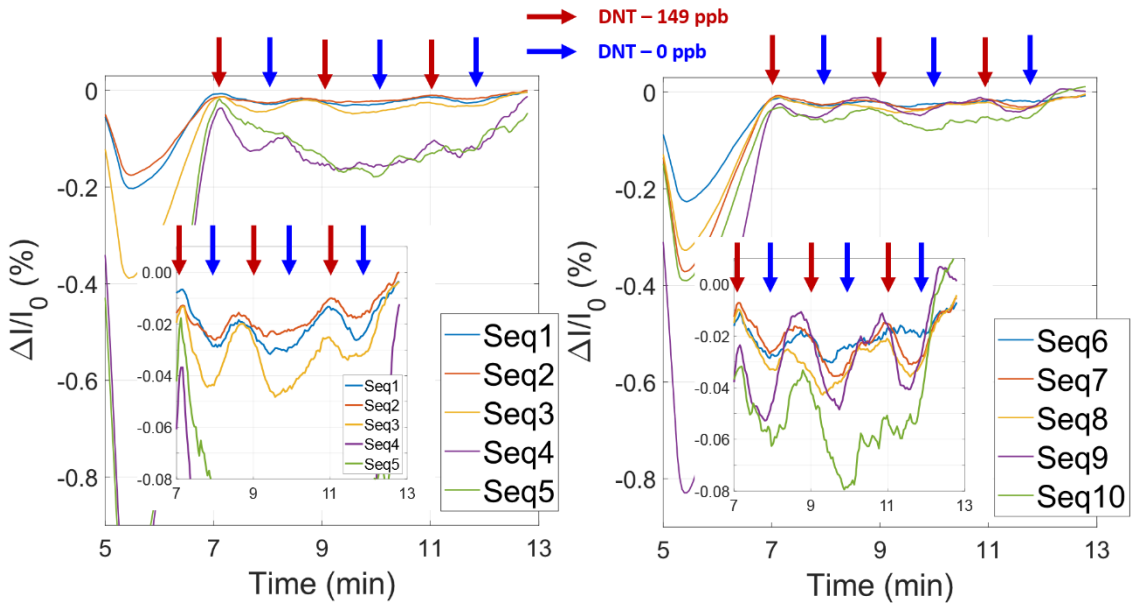

Figure 5.5. DNA-NT responses to pulses of 149 ppb DNT in 152 ppm DMMP background refreshed with 152 ppm DMMP instead of pure nitrogen. All responses were smoothed using 7-point boxcar averaging. The first pulse overresponded significantly, possibly due to unregulated buildup of

the DMMP headspace, and was therefore ignored. For the remaining responses, small yet apparent negative current shifts are observed in the absence of DNT in the mixture.

In conclusion, our DNA-NTs demonstrated promising performance in terms of sensitivity, fast response, long lifetime, and ability to detect a dilute target in a highly concentrated complex background. Though X Development did not disclose their particular interests in this technology, it is conceivable that DNA-NT arrays can be used for long-term monitoring of crops or mounted on robotic devices programmed to roam warzones or other potentially hazardous environments in search of explosives or chemical weapons. Sensor performance can be further optimized according the specific application by tuning DNA-NT affinities towards the desired targets via specially engineered DNA sequences. Further performance enhancements may be realized by using more sophisticated baseline-fitting algorithms and noise reduction techniques.

References

1. Wilson, A. D., Diverse applications of electronic-nose technologies in agriculture and forestry. *Sensors (Basel)* **2013,** *13* (2), 2295-348.
2. Greenshields, M. W.; Mamo, M. A.; Coville, N. J.; Spina, A. P.; Rosso, D. F.; Latocheski, E. C.; Destro, J. G.; Pimentel, I. C.; Hummelgen, I. A., Electronic detection of Drechslera sp. fungi in charentais melon (Cucumis melo Naudin) using carbon-nanostructure-based sensors. *J Agric Food Chem* **2012,** *60* (42), 10420-5.
3. Wiederoder, M. S.; Nallon, E. C.; Weiss, M.; McGraw, S. K.; Schnee, V. P.; Bright, C. J.; Polcha, M. P.; Paffenroth, R.; Uzarski, J. R., Graphene Nanoplatelet-Polymer Chemiresistive Sensor Arrays for the Detection and Discrimination of Chemical Warfare Agent Simulants. *Acs Sensors* **2017,** *2* (11), 1669-1678.
4. Brudzewski, K.; Osowski, S.; Pawlowski, W., Metal oxide sensor arrays for detection of explosives at sub-parts-per million concentration levels by the differential electronic nose. *Sensors and Actuators B-Chemical* **2012,** *161* (1), 528-533.
5. Abel, T.; Ungerbock, B.; Klimant, I.; Mayr, T., Fast responsive, optical trace level ammonia sensor for environmental monitoring. *Chem Cent J* **2012,** *6* (1), 124.

6. Baby, R. E.; Cabezas, M.; de Reca, E. N. W., Electronic nose: a useful tool for monitoring environmental contamination. *Sensors and Actuators B-Chemical* **2000,** *69* (3), 214-218.
7. Grimes, C. A.; Stoyanov, P. G.; Liu, Y.; Tong, C.; Ong, K. G.; Loiselle, K.; Shaw, M.; Doherty, S. A.; Seitz, W. R., A magnetostatic-coupling based remote query sensor for environmental monitoring. *J Phys D Appl Phys* **1999,** *32* (12), 1329-35.
8. Khan, S. B.; Faisal, M.; Rahman, M. M.; Jamal, A., Exploration of CeO(2) nanoparticles as a chemi-sensor and photo-catalyst for environmental applications. *Sci Total Environ* **2011,** *409* (15), 2987-92.
9. Lieberzeit, P. A.; Dickert, F. L., Sensor technology and its application in environmental analysis. *Anal Bioanal Chem* **2007,** *387* (1), 237-47.
10. Makombe, M.; van der Horst, C.; Silwana, B.; Iwuoha, E.; Somerset, V., Antimony film sensor for sensitive rare earth metal analysis in environmental samples. *J Environ Sci Health A Tox Hazard Subst Environ Eng* **2016,** *51* (8), 597-606.
11. Pella, P. A., Measurement of the vapor pressures of TNT, 2,4-DNT, 2,6-DNT, and EGDN. *The Journal of Chemical Thermodynamics* **1977,** *9* (4), 301-305.
12. Staii, C.; Johnson, A. T., Jr.; Chen, M.; Gelperin, A., DNA-decorated carbon nanotubes for chemical sensing. *Nano Lett* **2005,** *5* (9), 1774-8.
13. Butrow, A. B., Buchanan, J. H., & Tevault, D. E., Vapor Pressure of Organophosphorus Nerve Agent Simulant Compounds. *Journal of Chemical & Engineering Data* **2009,** *54* (6), 1876-1883.

CHAPTER 6: DNA-NT Screening for SARS-CoV-2 Based on Human Sweat

In this chapter, we discuss the deadly implications of the new coronavirus SARS-CoV-2 that spread rapidly across the world and the need for efficient and accurate screening of large populations for the COVID-19 disease. Early identification of infection would allow medical experts and government officials to take effective preventative measures, such as enforcing strict quarantine regulations for densely populated urban centers that are at high risk of an outbreak.

Section 6.1 introduces a pilot study initiated by our collaborator Dr. Cynthia Otto involving human sweat samples provided by individuals who had tested positive for COVID-19 (CoV+) as well as those who had been confirmed as COVID-negative (CoV-). The aim was to train scent detection dogs to discriminate between VOCs associated with the CoV+ and CoV- groups. All samples were shared with the Johnson Group for a complementary investigation using DNA-NT arrays.

Section 6.2 describes experiments involving DNA-NT measurements of headspace VOCs from 30 sweat samples: fifteen from COVID-positive subjects and fifteen from COVID-negative individuals. The ensuing DNA-NT responses were processed using linear discriminant analysis (LDA) which revealed strong discrimination between the CoV+ and CoV- groups. An LDA-based binary classifier was constructed using estimated normal distributions associated with the two groups. The classifier demonstrated degrees of sensitivity and specificity which are comparable to what is currently reported for reverse

transcription polymerase chain reaction (RT-PCR), one of the established screening techniques used for COVID-19 today. A receiver operating characteristic (ROC) curve was also generated using the CoV+ and CoV- distributions from which a 99% classification accuracy was estimated for the LDA binary classifier. These compelling results suggest significant promise for reliable screening based on DNA-NT sensor arrays and motivate further investigation.

6.1 Introduction

A new strain of coronavirus, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), emerged in Wuhan, China during the winter of 2019 and spread rapidly around the world.[1] Those infected with the virus develop coronavirus disease 2019 (COVID-19) with potentially severe or deadly outcomes.[2] This is the third coronavirus outbreak within the past twenty years and is by far the most infectious to humans with over 58 million reported cases and nearly 1.4 million deaths worldwide as of November 15, 2020.[1-2] The United States leads the world in both figures with over 12 million cases and 250,000 deaths.[1] Though there has been progress towards a vaccination for preventing the illness, none have yet been approved by the U. S. Food and Drug Administration.[3] Furthermore, developing accurate screening tests is challenging due to the novelty of COVID-19. Current detection approaches utilize reverse transcription polymerase chain reaction (RT-PCR) as well as antigen and antibody tests.[4] Reports of sensitivity for RT-PCR tests are largely uncertain, ranging from about 70% to 98%, while specificities are

estimated at roughly 95%.[5] The accuracies of antigen and antibody tests are even less certain as each suffers from high false negative rates.[4] A sensing technology that can permit effective and efficient screening of large populations would help reduce the rate at which COVID-19 spreads, particularly in airports, hospitals, and other potentially densely populated areas.

A pilot study was initiated by our collaborator Dr. Cynthia Otto at the Penn Vet Working Dog Center, University of Pennsylvania, with the aim of training scent detection dogs and using DNA-NT arrays to discriminate between human sweat samples from COVID-19 positive (CoV+) subjects and COVID-19 negative (CoV-) individuals. Dr. Otto established an online system through which volunteers could apply to donate sweat samples through the mail. Only volunteers who had been tested for COVID-19 were eligible for selection. Each volunteer was provided a brand new 100% cotton T-shirt and was given instructions to wear it to sleep for one night. The T-shirt would absorb sweat generated from the volunteer throughout the night. Each shirt was isolated inside a Ziploc bag for 24 hours before being returned to Dr. Otto in order to extinguish any lingering active viruses inhabiting the samples.[6] All shirts were transported and handled in a BSL2 biosafety hood to satisfy University of Pennsylvania regulations for COVID-19 research. Due to their high sweat content, portions from the sleeves, back, and chest areas were cut from each shirt and placed in individual glass jars for VOC preservation. Samples were split evenly and shared between the Otto and Johnson research groups.

6.2 DNA-NT Measurements of Human Sweat Samples

DNA-NT sensor arrays functionalized with ten DNA oligomers listed in Table 3.1 were used to measure headspace VOCs from fifteen COVID positive (CoV+) and fifteen COVID negative (CoV-) T-shirt samples. Strips of cloth were cut from the T-shirt sleeves using scissors cleaned with isopropanol and loaded into separate 25 mL round-bottom flasks. The samples were heated to 50 °C and were allowed 15 minutes to generate headspace VOCs. These were subsequently delivered to a ten-channel DNA-NT array for two minutes at a time using 67% RH nitrogen. Small yet distinguishable responses to the sample VOCs were observed (Figure 6.1).

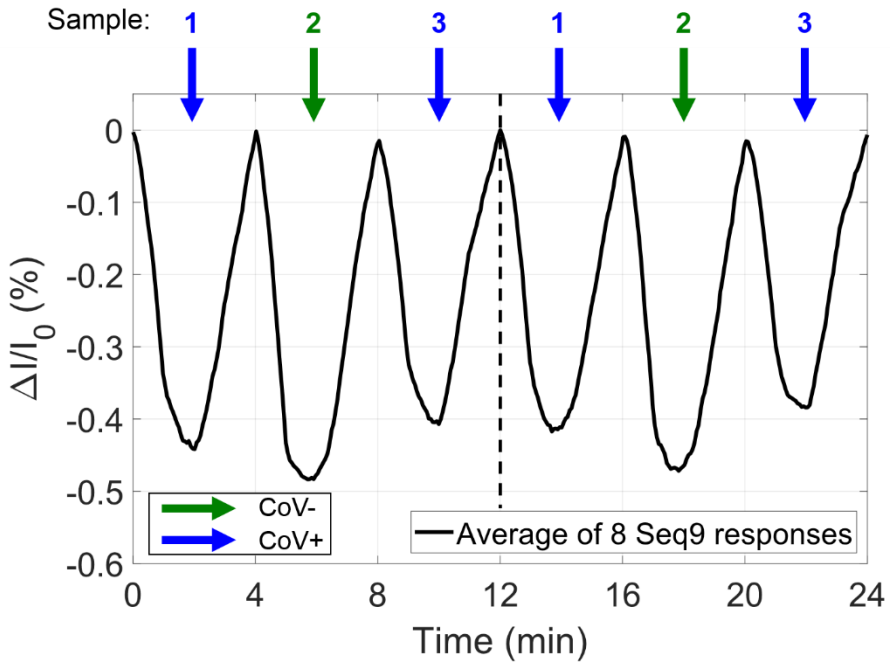

Figure 6.1. Average responses to eight Seq9 DNA-NT sensors exposed to headspace VOCs from two CoV+ and one CoV- T-shirt sleeve samples. Two cycles of measurements are shown to demonstrate signal reproducibility. The responses are small compared to those associated with

blood plasma, which are typically on the order of 10%. Nevertheless, the signal-to-noise ratios of the response averages are comparable to those observed for plasma.

The ten-channel DNA-NT response data was processed using linear discriminant analysis (LDA). Response projections onto the first discriminant revealed strong separation between the CoV+ and CoV- groups (Figure 6.2). Given the lack of overlap between the clusters, a classification boundary line between them could easily be imagined that would result in perfect binary classification of all data. For a more informed estimation of the classifier's diagnostic accuracy, Gaussian distributions were approximated for each cluster based on their respective means and standard deviations in order to simulate overlap between the groups. The false positive (false negative) rate was calculated by dividing the area of the overlapped region of the CoV+ (CoV-) distribution by the area of the non-overlapped region.

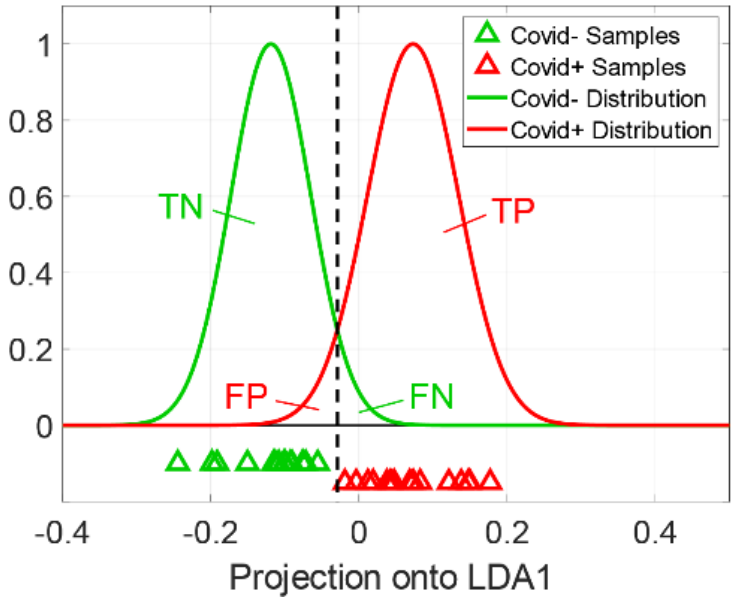

Figure 6.2. LDA projections of DNA-NT responses to fifteen CoV+ and fifteen CoV- T-shirt samples are represented as colored triangles along the first linear discriminant (LDA1) with

associated Gaussian distributions based on the computed means and standard deviations of each cluster. The dashed line through the intersection between the curves indicates a binary classification boundary which defines the true positive, false positive, true negative, and false negative regions (labeled as TP, FP, TN, and FN, respectively).

The sensitivity and specificity for the binary classifier were also estimated based on the Gaussian distributions from Figure 6.2. Sensitivity measures the fraction of CoV+ cases that were correctly identified as positive while specificity measures the fraction of CoV- cases that were correctly classified as negative (6.2.1):

$$Sensitivity = \frac{TPR}{TPR + FNR} \tag{6.2.1a}$$

$$Selectivity = \frac{TNR}{TNR + FPR} \tag{6.2.1b}$$

where TPR, TNR, FPR, and FNR are the true positive, true negative, false positive, and false negative rates, respectively. The calculated DNA-NT sensitivity was 93.5% and the selectivity was 95.8% which is comparable to the most optimistic estimates for RT-PCR tests. The diagnostic accuracy of the binary classifier based on the two Gaussian distributions was also estimated using a receiver operating characteristic (ROC) curve (Figure 6.3). The area under the curve indicated a 99% probability of correctly distinguishing CoV+ from CoV- data.

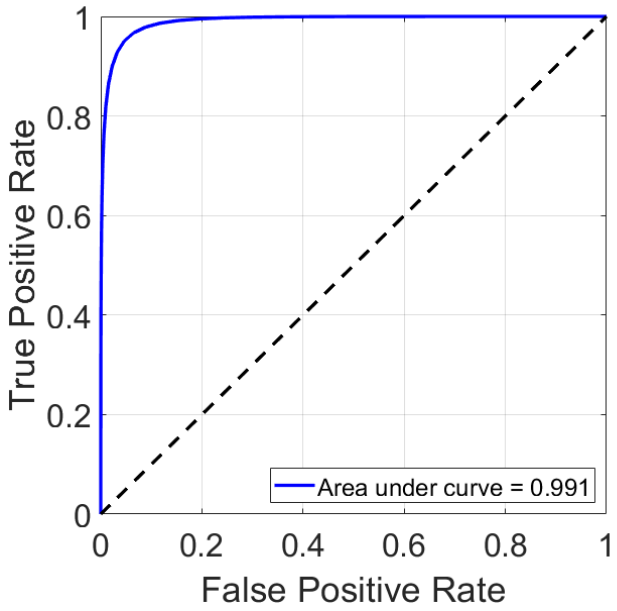

Figure 6.3. Receiver operating characteristic for the estimated normal distributions from Figure 6.2. The area under the curve indicates a 99% probability of correctly distinguishing CoV+ and CoV- data.

In conclusion, DNA-NT sensors demonstrated a strong capacity to differentiate between headspace VOCs from human CoV+ sweat and CoV- sweat in this initial proof-of-concept investigation. The success of the DNA-NT sensors is particularly impressive in this case because they were immediately adaptable to discriminating between VOCs associated with such a new disease, alluding to the flexibility of the technology. Advantages of this e-nose approach include speed of development, on-site prompt testing potential, guaranteed safety to both the patient and the caretaker, and relatively low cost of deployment. Furthermore, because only human sweat is needed as a specimen, this approach is non-intrusive and compatible with other testing approaches.

These compelling results motivate further development of an e-nose approach to screening populations for the COVID-19 disease. In order to build further acclaim for this screening method, we plan to increase our sample database up to 400 CoV- and 200 CoV+ individuals. As is discussed in previous chapters, we also have the opportunity to improve our selection of DNA oligomers to further increase the discrimination power of the DNA-NTs, perhaps by targeting specific characteristics of biomarker molecules discovered by GC/MS that are associated with COVID-19. Finally, the success of LDA-based data processing suggests the possibility of using more sophisticated machine learning algorithms that may improve the performance of this e-nose based diagnostic.

References

1. Coronavirus disease (COVID-2019) situation reports. 2020. (Accessed 22nd of November, 2020, at https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/.).
2. Yang, Y.; Peng, F.; Wang, R.; Yange, M.; Guan, K.; Jiang, T.; Xu, G.; Sun, J.; Chang, C., The deadly coronaviruses: The 2003 SARS pandemic and the 2020 novel coronavirus epidemic in China. *J Autoimmun* **2020,** *109*, 102434.
3. U. S. Food & Drug Administration COVID-19 Vaccines. 2020. (Accessed 25th of November, 2020, at https://www.fda.gov/emergency-preparedness-and-response/coronavirus-disease-2019-covid-19/covid-19-vaccines).
4. Which test is best for COVID-19? Harvard Health Blog. 2020. (Accessed 25th of November, 2020, at https://www.fda.gov/emergency-preparedness-and-response/coronavirus-disease-2019-covid-19/covid-19-vaccines).
5. Watson, J.; Whiting, P. F.; Brush, J. E., Interpreting a covid-19 test result. *BMJ* **2020,** *369*, m1808.
6. Harbourt, D. E.; Haddow, A. D.; Piper, A. E.; Bloomfield, H.; Kearney, B. J.; Fetterer, D.; Gibson, K.; Minogue, T., Modeling the stability of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) on skin, currency, and clothing. *PLoS Negl Trop Dis* **2020,** *14* (11), e0008831.

Chapter 7: Summary, Conclusions and Future Work

In this work, progress was made in developing an e-nose system based on DNA-NT arrays capable of screening individuals for ovarian cancer and COVID-19 using human blood plasma and sweat, respectively. In addition, DNA-NTs were shown to have desirable attributes in terms of parts-per-billion (ppb) limits of detection, the ability to detect low concentrations of target VOCs in the presence of a highly concentrated background, fast responses on the order of seconds, and lifetimes on the order of months. The successes enjoyed by the DNA-NTs suggest significant promise for their utility in disease diagnostics as well as various environmental monitoring applications such as pollution detection.

These vapor sensors possess exquisite sensitivity that has only recently become available due to the technological breakthroughs and advancements in the fields of nanotechnology and bioengineering and rise up to the challenge posed by early-stage disease detection where very small concentration of one or several compounds is present. Furthermore, despite the fact that the identities of relevant VOC compounds are often not completely known, DNA-NT arrays have demonstrated a desirable ability to differentiate between informative vapor species using the same principles that govern the discrimination capabilities of mammalian olfaction systems found in nature.

In addition, DNA-NTs enjoy numerous attributes that make them ideal candidates for hospital use or remote monitoring of various environments. Because they are fast-responding, cost-efficient, and easy to use by non-technical personnel, e-nose systems based on DNA-NTs are ideal for on-site prompt testing. These characteristics are clear

advantages over expensive and highly complex analytical techniques such as gas chromatography / mass spectrometry (GC/MS) which are cumbersome and suffer from slow readouts. Given their function of probing compounds in the vapor phase, DNA-NTs are also generally non-intrusive as the possibility of direct contamination of the sensor (or contamination of the sample from the sensor) is reduced.

The fabrication of vapor sensor arrays is simple and can also be customized and tailored towards specific applications by choosing the DNA oligomers that work best for detecting certain molecules. Thus, as researchers develop a more complete understanding of the volatile metabolites in bodily tissues or fluids that indicate the presence of particular diseases of interest, the DNA-NT arrays can be evolved for improved screening performance.

In medical practice, clinical screening always yields false positives and negatives due to biological variations between individuals and countless other sources. Both forms of experimental error are of clinical concern. False negatives occur when a screening fails to identify the presence of a condition. This is especially problematic for diseases like ovarian cancer that can only be treated reliably in the early stages. A false negative in this case would often be fatal as the cancer would be able to metastasize unbeknownst to caretakers and patients alike until it is too late to treat the patient. With such high stakes, false negatives reduce the confidence of the medical community that the screening technology is worthwhile. False positives have dramatic consequences as well. False positives will put healthy individuals through unnecessary follow-up examinations such as biopsies. These tests are often harmful to the patients that receive them, and the overall

process can cause a great deal of needless anxiety. Moreover, biopsies are arduous and expensive procedures, and too many false positives will put strain on hospitals that have limited resources in terms of personnel, equipment, and funds.

The DNA-NT arrays have already demonstrated promising performance in terms of false positives and negatives. However, this technology has yet to see extensive testing in a real hospital setting. To curtail some of the worry surrounding false predictions, DNA-NTs can be combined with other techniques that could provide orthogonal information to increase the probability of identifying important features in samples that would ultimately lead to improved classification performance. This is easily imaginable, as vapor-phase chemical sensing is unlikely to interfere with other sensing modalities in terms of contamination.

APPENDIX

Appendix A: Basic Mathematical Foundations of Machine Learning Algorithms

Appendix A.1: Principal Component Analysis

The general procedure of PCA is to represent a data set with a basis that will best distinguish significant parameter correlations from signal noise. We start by imagining a data set acquired from a 10-channel NTFET array exposed to VOCs from *n* different trials. We define $\boldsymbol{X} = \{\vec{x}_1, \vec{x}_2, \ldots \vec{x}_{10}\}$ as the set of ten-dimensional NTFET outputs where the column vector $\vec{x}_1$ represents the Seq1 responses, $\vec{x}_2$ represents the Seq2 responses, and so on. Important trends are easily revealed when the signal-to-noise ratios (SNRs) of the responses are high, and much less so when the SNRs are low. The response signals are associated with the variances of $\vec{x}_1, \vec{x}_2, \ldots \vec{x}_{10}$. Meanwhile, NTFET responses are also strongly correlated because they are highly cross-sensitive. This is an indication that much of the information in the output data is redundant. Thus, we consider the covariances between the $\vec{x}_i$ parameters represented in the following covariance matrix:

$$S_X = \frac{1}{n-1}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}} = \begin{bmatrix} var(x_1) & cov(x_1, x_2) & \cdots & cov(x_1, x_{10}) \\ cov(x_2, x_1) & var(x_2) & & cov(x_2, x_{10}) \\ \vdots & & \ddots & \vdots \\ cov(x_{10}, x_1) & cov(x_{10}, x_2) & \cdots & var(x_{10}) \end{bmatrix}$$

We now want to represent $\boldsymbol{X}$ in a basis more suitable to revealing only the most important information with as few parameters as necessary. To do this, we apply a linear transformation on $\boldsymbol{X}$:

$$\boldsymbol{PX} = \boldsymbol{Y}$$

where $\boldsymbol{P}$ is an orthonormal matrix. This transformation essentially rotates the representation of $\boldsymbol{X}$ to a new set of axes encoded in the rows of $\boldsymbol{P}$. The best choice of $\boldsymbol{P}$ associates only a few principal components with high measures of SNR – that is to say, maximizing the response signals, measured by the variances, and minimizing redundancy, measured by covariances. This is achieved if the covariance matrix S_Y of $\boldsymbol{Y} = \boldsymbol{PX}$ takes a diagonal form:

$$S_Y = \frac{1}{n-1}\boldsymbol{YY}^{\mathrm{T}} = \frac{1}{n-1}(\boldsymbol{PX})(\boldsymbol{PX})^{\mathrm{T}} = \frac{1}{n-1}\boldsymbol{P}(\boldsymbol{XX^T})\boldsymbol{P^T}$$

Since $\boldsymbol{XX^T}$ is a symmetric matrix, it can be diagonalized by $\boldsymbol{XX^T} = \boldsymbol{VDV}^T$. Here, the columns of $\boldsymbol{V}$ are the eigenvectors of $\boldsymbol{XX^T}$ and $\boldsymbol{D}$ is a diagonal matrix whose elements are the eigenvalues of $\boldsymbol{XX^T}$. Setting $\boldsymbol{P} = \boldsymbol{V}^T$ simplifies our expression:

$$S_Y = \frac{1}{n-1}\boldsymbol{P}(\boldsymbol{VDV}^T)\boldsymbol{P^T} = \frac{1}{n-1}\boldsymbol{P}(\boldsymbol{P^TDP})\boldsymbol{P^T} = \frac{1}{n-1}\boldsymbol{D}$$

In this way, the procedure for PCA is reduced to solving a generalized eigenvalue for $\boldsymbol{XX^T}$.

Appendix A.2: Linear Discriminant Analysis

We imagine a data set where each data point, v^i, represents ten response averages (Seq1, Seq2, and so on). We define the basis $\boldsymbol{V} \in \mathbb{R}$ such that each 10-dimensional data

point $\{v^1, v^2, v^3, v^4, \ldots v^N\} \in \boldsymbol{V}$ where N is the total number of data points. Our mission is to find a transformation that maps $\boldsymbol{V} \rightarrow \boldsymbol{Y}$ where $\boldsymbol{Y}$ is the basis with maximized class separation. For simplicity, we assume there are only two classes with N_1 data points in Class 1 and N_2 data points in Class 2. We relate y to v by the transformation $y = w^T v$ where $v = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_m \end{bmatrix}$ and $w = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_m \end{bmatrix}$. To maximize the separation between class distributions, we will need two properties: (1) the relative positions of each distribution in space and (2) the degree of spread within each distribution.

Starting with first property, let μ_i be the centroid of distribution of the i^{th} class represented in the $\boldsymbol{V}$ basis:

$$\mu_i = \frac{1}{N_i} \sum_{j=1}^{N_i} v_j$$

Transforming this to the $\boldsymbol{Y}$ basis, we obtain:

$$\tilde{\mu}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} y_j = \frac{1}{N_i} \sum_{j=1}^{N_i} w^T v_j = w^T \cdot \frac{1}{N_i} \sum_{j=1}^{N_i} v_j = w^T \mu_i$$

We now define a measure of the "offset" between distributions:

$$(\tilde{\mu}_1 - \tilde{\mu}_2)^2 = (w^T \mu_1 - w^T \mu_2)^2 = w^T (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T w = w^T \boldsymbol{S_B} w = \tilde{\boldsymbol{S}}_{\boldsymbol{B}}$$

Here, $\boldsymbol{S_B}$ is the between-class scatter matrix for samples in the $\boldsymbol{V}$ space while $\tilde{\boldsymbol{S}}_{\boldsymbol{B}}$ is same matrix represented in the $\boldsymbol{Y}$ basis.

Now, for the second property, we define the following metric of within-class scatter:

$$s_i^2 = \sum_{j=1}^{N_i} (x_j - \mu_j)(x_j - \mu_j)^T$$

or alternatively,

$$\tilde{s}_i^2 = \sum_{j=1}^{N_i} (y_j - \tilde{\mu}_j)(y_j - \tilde{\mu}_j)^T$$

We can define a within-class scatter matrix $\boldsymbol{S_i}$ for each class as the following:

$$\tilde{s}_i^2 = \sum_{j=1}^{N_i} (y_j - \tilde{\mu}_j)(y_j - \tilde{\mu}_j)^T = \sum_{j=1}^{N_i} w^T (x_j - \mu_j)(x_j - \mu_j)^T w = w^T \boldsymbol{S_i} w$$

Then the total within-class scatter for the entire data set is defined as the sum of these terms for each class:

$$\tilde{s}_1^2 + \tilde{s}_2^2 = w^T(\boldsymbol{S_1} + \boldsymbol{S_2})w = w^T \boldsymbol{S_W} w = \tilde{\boldsymbol{S}}_{\boldsymbol{W}}$$

where $\boldsymbol{S_W}$ and $\tilde{\boldsymbol{S}}_{\boldsymbol{W}}$ are the complete within-class scatter matrices for the entire data set in bases $\boldsymbol{V}$ and $\boldsymbol{Y}$, respectively.

The transformation $y = w^T v$ the between-class separation relative to the overall within-class scatter is accomplished by finding a w that maximizes the following:

$$J(w) = \frac{|\tilde{\mu}_1 - \tilde{\mu}_2|^2}{\tilde{s}_1^2 + \tilde{s}_2^2} = \frac{w^T \boldsymbol{S_B} w}{w^T \boldsymbol{S_W} w}$$

Solving for w, we differentiate both sides and set the result equal to zero:

$$\frac{d}{dw} J(w) = \frac{d}{dw}\left(\frac{w^T \boldsymbol{S_B} w}{w^T \boldsymbol{S_W} w}\right) = 0$$

With some algebra manipulation we obtain the following:

$$\boldsymbol{S_W^{-1}} \boldsymbol{S_B} w - \left(\frac{w^T \boldsymbol{S_B} w}{w^T \boldsymbol{S_W} w}\right) w = \boldsymbol{S_W^{-1}} \boldsymbol{S_B} w - J(w) w = 0$$

We are now faced with a generalized eigenvalue problem where $\boldsymbol{S_W^{-1}S_B} = \lambda w$. Here, the eigenvalue $\lambda = J(w)$. We can finally solve for w:

$$w^* = \underset{w}{\text{argmax}}\,[J(w)] = \underset{w}{\text{argmax}}\left[\frac{w^T \boldsymbol{S_B} w}{w^T \boldsymbol{S_W} w}\right] = \boldsymbol{S_W^{-1}}(\mu_1 - \mu_2)$$

This is the general solution for the optimized basis for maximizing the separation between two classes. The result can be generalized to any number of classes with some more effort.

Appendix B: Fabrication Procedures for DNA-Functionalized NTFET Arrays

Appendix B.1: Photolithography Process

Deposit 15nm of Al_2O_3 on standard Si/SiO_2 wafer. Then follow the steps below:

1.) Pre-bake

Dehydration bake: ~150°C, ~ 2mins.

2.) Primer

Spin PMGI: 4000 rpm, 45s, 500 rpm/s

Bake at 210°C for 5 minutes

3.) Photoresist

Spin S1813: 5000 rpm, 45s, 500 rpm/s

Bake at 100°C for 2 minutes

4.) UV Light Exposure

Use a hard contact procedure with manual alignment. Exposure dose is **120** mJ/cm^2. Use the mask-aligner camera to make sure the common source electrode for column 10 is not clipped by the edge of the silicon wafer.

5.) Development

Immerse substrate in MF-319 developer for around 30-60s. Use a pipette to agitate and disperse dissolving resist away from surface. The features of the pattern should gradually appear and change color due to the changing thickness of the resist. Once the surface is a uniform dark purple, remove the wafer from the developer and immediately rinse with water for a few minutes. Blow-dry with a nitrogen gun when the wafer is sufficiently cleaned.

6.) Evaporation

Normally evaporate 5 nm Cr / 40 nm Au

7.) Lift-off

Use R1165 remover to lift-off leftover resist. Leave the wafer soaking overnight and then heat the remover to 60°C. Use a pipette to agitate the solvent so that every last bit of residual photoresist is dissolved. Sometimes sonication is needed to resolve finer features. When the desired pattern is fully resolved, transfer the wafer to an acetone bath for 10 min and then to an isopropanol bath for 10 min. After blow-drying, bake the wafer at 250°C for 1 hour.

Appendix B.2: CNT Deposition

1. Acquire one large petri dish (bottom and cover), one small petri dish (cover only), and one glass slide. Place glass slide and small cover in the large bottom.
2. If sonication is needed, pipette desired volume of CNT solution in a 1.5 mL aliquot vial and place vial in a foam "floater." Place in the small sonicator and sonicate for desired duration.
3. Choose array(s) for CNT deposition and blow off with N_2 gas. Place chips on the glass slide.
4. Start boiling DI water.
5. Drop-cast roughly 130 – 140 µL of CNT solution for each array.
6. Pour boiling water into small cover. Cover everything with the large cover and set a timer for 30 min.
7. When the timer goes off, remove the top cover and the water container. Pour isopropanol directly into the large bottom so that the chips are fully immersed in solvent.
8. Set a timer for 15 minutes. Agitate the IPA bath every 3 minutes.
9. Transfer to a water bath. Be sure to prevent the chips from drying out during the transfer by spraying each chip with water from a squirt bottle while it is exposed to the air.
10. Set timer for 5 minutes.
11. Repeat steps 9 and 10 two more times.
12. Blow each chip dry with N_2 gas.

13. Set each chip on a hot plate set at 165ºC. Set a timer for 1 hour.
14. Remove chips from hot plate. Once cool, apply silver paint to the underside of each chip.

Appendix B.3: DNA Deposition

1. Acquire materials- 1 large petri dish (bottom and cover), 1 small petri dish (cover only), and 1 glass slide. Place glass slide and small cover in the large bottom.

2. Thaw aliquots of desired DNA sequences (usually I use Seq1 through Seq10)

3. Choose array for DNA deposition and blow off with N_2 gas. Place the chip on the glass slide.

4. Start boiling DI water.

5. Pipette roughly 1.8 – 1.9 µL of DNA onto the regions shown below:

6. Fill the small petri dish with hot water and close the cover of humidity bath.

7. Set a timer for 30 minutes.

8. Remove the chip from the humidity bath. Use the nitrogen gun to carefully blow the five leftmost droplets off the left side of the chip one at a time (corresponding to Seq1 – Seq5 in the diagram). Then blow off the remaining five droplets simultaneously.

Appendix C: Tabulated Ovarian Plasma Responses

Current responses to VOCs from the 58 individual plasma samples. Each response value represents the percent variation from baseline, $\Delta I/I_0$, for all sensors of a given sensor type. The sample set consists of 21 healthy samples, 16 benign samples, and 21 malignant samples, of which five were from women with Stage I/II ovarian cancer. Variations in average responses within each class is believed to be caused by biological variations between individuals.

Sample #	Seq1	Seq2	Seq3	Seq4	Seq5	Seq6	Seq7	Seq8	Seq9	Seq10
Healthy										
5	-3.84	-0.60	-2.24	-1.79	-2.95	-2.17	-2.26	-4.31	-3.37	-4.45
3	-4.74	-1.71	-3.44	-3.21	-7.06	-3.39	-2.55	-6.60	-4.65	-5.13
13	-3.26	-0.90	-2.00	-1.82	-3.71	-2.03	-1.92	-4.06	-2.98	-3.61
15	-3.04	-0.95	-1.91	-1.90	-4.42	-1.45	-2.36	-4.45	-2.99	-3.19
7	-3.38	-0.72	-2.12	-1.84	-3.82	-2.03	-2.18	-4.91	-2.80	-4.05
14	-1.97	-1.12	-0.81	-0.88	-1.85	-0.85	-0.81	-1.84	-1.10	-1.29
9	-1.87	-1.26	-0.77	-0.79	-1.36	-0.75	-0.84	-1.84	-0.94	-1.24
18	-1.65	-1.11	-0.83	-0.81	-1.66	-0.67	-0.95	-1.89	-1.13	-1.32
17	-1.72	-1.08	-0.69	-0.75	-1.24	-0.58	-0.85	-1.66	-1.04	-1.20
B10	-2.91	-1.17	-1.79	-1.33	-3.68	-1.84	-1.57	-3.47	-2.00	-2.68
B16	-3.08	-1.10	-1.76	-1.65	-3.01	-1.57	-1.54	-4.09	-2.81	-2.63
D21	-2.71	-0.96	-1.40	-1.50	-3.67	-1.47	-1.79	-3.22	-2.57	-2.52
C7	-2.75	-1.03	-1.61	-1.49	-3.38	-1.60	-1.93	-3.69	-2.21	-2.99
C24	-2.97	-1.07	-1.28	-1.24	-3.51	-1.25	-1.55	-2.88	-2.21	-3.06
C30	-3.12	-1.04	-1.57	-1.57	-2.56	-1.63	-1.65	-3.30	-2.45	-2.72
D8	-3.39	-1.03	-1.58	-1.43	-3.18	-1.62	-1.64	-3.28	-2.59	-3.12
D26	-3.03	-1.02	-1.69	-1.73	-3.00	-1.25	-1.69	-3.62	-2.94	-2.47
D31	-3.13	-1.06	-1.81	-1.67	-3.64	-1.49	-1.68	-3.79	-2.45	-3.16
D36	-2.69	-0.91	-1.62	-1.48	-3.47	-1.45	-1.36	-3.75	-2.21	-2.50
D56	-2.60	-0.98	-1.71	-1.45	-2.99	-1.31	-1.66	-3.95	-2.36	-2.34
D59	-3.25	-1.20	-1.93	-1.61	-3.16	-1.67	-1.56	-3.79	-2.68	-3.22
Benign										
54	-2.18	-1.57	-1.33	-1.09	-1.46	-1.24	-0.96	-2.72	-1.95	-2.00
53	-2.27	-1.80	-1.38	-1.18	-1.24	-1.27	-1.19	-2.70	-2.01	-2.21
58	-2.33	-1.91	-1.38	-1.27	-2.14	-1.32	-1.17	-2.94	-2.08	-2.06
64	-1.32	-1.21	-0.57	-0.50	-0.96	-0.42	-0.94	-1.09	-0.98	-1.02

56	-1.97	-1.64	-1.16	-0.96	-1.07	-0.93	-1.16	-2.36	-1.56	-1.79
H5837	-2.19	-1.69	-1.12	-0.98	-1.30	-1.05	-1.15	-2.14	-1.61	-1.68
H6032-8	-2.06	-1.13	-1.18	-1.05	-1.44	-0.85	-1.13	-2.47	-1.63	-1.86
H6001-11	-2.12	-1.59	-1.14	-1.06	-1.48	-0.71	-1.01	-2.37	-1.62	-1.67
H6130-10	-1.89	-1.65	-1.21	-1.05	-1.49	-1.01	-1.08	-2.07	-1.71	-1.84
H6153-6	-1.86	-1.63	-1.11	-0.98	-1.32	-1.15	-1.07	-2.45	-1.71	-1.84
H5869	-2.05	-1.65	-1.22	-0.97	-1.39	-0.99	-1.01	-2.40	-1.70	-1.71
H6019-4	-2.08	-1.71	-1.13	-0.89	-1.37	-0.83	-1.14	-2.23	-1.63	-1.89
H6044-5	-2.12	-1.48	-1.11	-1.08	-1.48	-1.05	-1.14	-2.33	-1.76	-1.89
H6057-7	-1.96	-1.64	-1.17	-0.98	-1.49	-0.84	-0.93	-2.28	-1.53	-1.95
H6145-2	-1.94	-1.47	-1.24	-1.13	-1.32	-0.98	-1.11	-2.43	-1.85	-1.63
H6151-6	-2.17	-1.69	-1.14	-0.99	-1.19	-1.12	-1.12	-2.39	-1.75	-1.69
Malignant										
49	-2.25	-1.54	-3.15	-3.33	-2.26	-2.62	-1.10	-4.61	-5.44	-3.87
46	-6.67	-2.16	-4.40	-7.15	-9.63	-3.97	-6.97	-7.54	-5.30	-7.83
52	-2.02	-0.95	-2.97	-2.14	-2.07	-2.33	-1.19	-2.83	-3.50	-3.00
57	-1.39	-0.72	-3.17	-1.88	-2.62	-0.31	-1.85	-2.16	-3.27	-1.74
50	-1.39	-0.74	-3.55	-1.89	-4.16	-1.04	-2.80	-2.56	-2.23	-2.26
48	-1.63	-1.60	-0.94	-1.15	-1.36	-0.89	-1.54	-1.84	-1.35	-1.48
45	-2.21	-2.39	-1.42	-1.59	-2.41	-1.75	-1.17	-2.98	-1.98	-1.80
66*	-2.19	-1.56	-1.08	-1.14	-1.54	-1.16	-1.83	-2.25	-1.62	-1.69
68	-1.96	-1.29	-1.00	-1.07	-1.58	-1.03	-1.51	-2.14	-1.56	-1.54
65	-1.9	-2.11	-0.87	-0.96	-1.29	-0.78	-1.46	-2.14	-1.38	-1.46
H6035-14	-2.39	-1.53	-2.07	-2.04	-3.00	-1.64	-1.93	-2.74	-2.92	-2.42
H5919-10	-2.47	-1.61	-2.62	-2.49	-2.53	-1.59	-2.24	-2.65	-2.90	-2.67
H6017-16	-2.24	-1.57	-2.24	-2.28	-2.45	-1.50	-2.04	-3.11	-3.09	-3.11
H6133-9	-2.13	-1.44	-2.23	-1.94	-2.55	-1.74	-2.10	-3.28	-2.53	-2.72
4451-C-1210*	-2.12	-1.45	-2.20	-1.83	-2.40	-1.72	-2.50	-3.13	-2.75	-2.47
H5877*	-2.26	-1.57	-2.32	-2.23	-2.78	-1.55	-1.97	-3.23	-2.79	-2.50
H5926-16	-2.44	-1.49	-2.20	-1.87	-3.41	-1.61	-2.20	-3.27	-2.74	-2.26
H5899*	-2.46	-1.54	-2.35	-2.27	-3.16	-1.82	-2.29	-3.55	-2.91	-2.36
H6011-9*	-2.59	-1.44	-2.21	-2.44	-2.94	-1.83	-2.18	-3.45	-2.84	-2.55
H5940-10*	-2.13	-1.56	-2.06	-2.25	-3.36	-1.83	-2.05	-3.11	-2.58	-2.81
H6083-9	-2.54	-1.39	-2.27	-2.65	-3.06	-1.99	-2.89	-3.25	-2.78	-2.74

* *Stage I/II ovarian cancer*

BIBLIOGRAPHY

Abel, T.; Ungerbock, B.; Klimant, I.; Mayr, T., Fast responsive, optical trace level ammonia sensor for environmental monitoring. *Chem Cent J* **2012,** *6* (1), 124.

Alon, U., *An introduction to systems biology : design principles of biological circuits*. Chapman & Hall/CRC: Boca Raton, FL, 2007; p xvi, 301 p., 4 p. of plates.

American Cancer Society, Cancer Facts and Figures 2019. 2019.

Aqel, A.; Abou El-Nour, K. M. M.; Ammar, R. A. A.; Al-Warthan, A., Carbon nanotubes, science and technology part (I) structure, synthesis and characterisation. *Arabian Journal of Chemistry* **2012,** *5* (1), 1-23.

Arnold, M. S.; Green, A. A.; Hulvat, J. F.; Stupp, S. I.; Hersam, M. C., Sorting carbon nanotubes by electronic structure using density differentiation. *Nat Nanotechnol* **2006,** *1* (1), 60-5.

Baby, R. E.; Cabezas, M.; de Reca, E. N. W., Electronic nose: a useful tool for monitoring environmental contamination. *Sensors and Actuators B-Chemical* **2000,** *69* (3), 214-218.

Benitez-Martinez, S.; Lopez-Lorente, A. I.; Valcarcel, M., Graphene quantum dots sensor for the determination of graphene oxide in environmental water samples. *Anal Chem* **2014,** *86* (24), 12279-84.

Bishop, C. M., *Pattern recognition and machine learning*. Springer: New York, 2006; p 561-565.

Breaker, R. R., Natural and engineered nucleic acids as tools to explore biology. *Nature* **2004,** *432* (7019), 838-45.

Brudzewski, K.; Osowski, S.; Pawlowski, W., Metal oxide sensor arrays for detection of explosives at sub-parts-per million concentration levels by the differential electronic nose. *Sensors and Actuators B-Chemical* **2012,** *161* (1), 528-533.

Butrow, A. B., Buchanan, J. H., & Tevault, D. E., Vapor Pressure of Organophosphorus Nerve Agent Simulant Compounds. *Journal of Chemical & Engineering Data* **2009,** *54* (6), 1876-1883.

Buys, S. S.; Partridge, E.; Black, A.; Johnson, C. C.; Lamerato, L.; Isaacs, C.; Reding, D. J.; Greenlee, R. T.; Yokochi, L. A.; Kessel, B.; Crawford, E. D.; Church, T. R.; Andriole, G. L.; Weissfeld, J. L.; Fouad, M. N.; Chia, D.; O'Brien, B.; Ragard, L. R.;

Clapp, J. D.; Rathmell, J. M.; Riley, T. L.; Hartge, P.; Pinsky, P. F.; Zhu, C. S.; Izmirlian, G.; Kramer, B. S.; Miller, A. B.; Xu, J. L.; Prorok, P. C.; Gohagan, J. K.; Berg, C. D., Effect of screening on ovarian cancer mortality: the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Randomized Controlled Trial. *JAMA* **2011,** *305* (22), 2295-303.

Chu, D.; Zabet, N. R.; Mitavskiy, B., Models of transcription factor binding: sensitivity of activation functions to model assumptions. *J Theor Biol* **2009,** *257* (3), 419-29.

Coronavirus disease (COVID-2019) situation reports. 2020. (Accessed 22nd of November, 2020, at https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/.).

Ding, S.; Sachs, F., Single channel properties of P2X2 purinoceptors. *J Gen Physiol* **1999,** *113* (5), 695-720.

Gelperin, A. and Hopfield, J.J., Electronic and Computational Olfaction, in Chemistry of Taste: Mechanisms, Behaviors and Mimics. (eds. P. Given & D. Paredes) 289-317 (American Chemical Society, Washington DC; 2002).

Goldsmith, B. R.; Mitala, J. J.; Josue, J.; Castro, A.; Lerner, M. B.; Bayburt, T. H.; Khamis, S. M.; Jones, R. A.; Brand, J. G.; Sligar, S. G.; Luetje, C. W.; Gelperin, A.; Rhodes, P. A.; Discher, B. M.; Johnson, A. T. C., Biomimetic Chemical Sensors Using Nanoelectronic Readout of Olfactory Receptor Proteins. *Acs Nano* **2011,** *5* (7), 5408-5416.

Greenshields, M. W.; Mamo, M. A.; Coville, N. J.; Spina, A. P.; Rosso, D. F.; Latocheski, E. C.; Destro, J. G.; Pimentel, I. C.; Hummelgen, I. A., Electronic detection of Drechslera sp. fungi in charentais melon (Cucumis melo Naudin) using carbon-nanostructure-based sensors. *J Agric Food Chem* **2012,** *60* (42), 10420-5.

Grimes, C. A.; Stoyanov, P. G.; Liu, Y.; Tong, C.; Ong, K. G.; Loiselle, K.; Shaw, M.; Doherty, S. A.; Seitz, W. R., A magnetostatic-coupling based remote query sensor for environmental monitoring. *J Phys D Appl Phys* **1999,** *32* (12), 1329-35.

Harbourt, D. E.; Haddow, A. D.; Piper, A. E.; Bloomfield, H.; Kearney, B. J.; Fetterer, D.; Gibson, K.; Minogue, T., Modeling the stability of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) on skin, currency, and clothing. *PLoS Negl Trop Dis* **2020,** *14* (11), e0008831.

Hartwell, L. H.; Hopfield, J. J.; Leibler, S.; Murray, A. W., From molecular to modular cell biology. *Nature* **1999,** *402* (6761 Suppl), C47-52.

Hierlemann, A.; Zellers, E. T.; Ricco, A. J., Use of linear solvation energy relationships for modeling responses from polymer-coated acoustic-wave vapor sensors. *Analytical Chemistry* **2001,** *73* (14), 3458-3466.

Hoover, K. C., Evolution of olfactory receptors. *Methods Mol Biol* **2013,** *1003*, 241-9.

Hopfield, J. J., Odor space and olfactory processing: collective algorithms and neural implementation. *Proc Natl Acad Sci U S A* **1999,** *96* (22), 12506-11.

Horvath, G.; Andersson, H.; Nemes, S., Cancer odor in the blood of ovarian cancer patients: a retrospective study of detection by dogs during treatment, 3 and 6 months afterward. *BMC Cancer* **2013,** *13*, 396.

Johnson, K. J.; Rose-Pehrsson, S. L., Sensor Array Design for Complex Sensing Tasks. *Annual Review of Analytical Chemistry, Vol 8* **2015,** *8*, 287-310.

Johnson, R. R.; Johnson, A. T. C.; Klein, M. L., Probing the structure of DNA-carbon nanotube hybrids with molecular dynamics. *Nano Letters* **2008,** *8* (1), 69-75.

Johnson, R. R.; Kohlmeyer, A.; Johnson, A. T. C.; Klein, M. L., Free Energy Landscape of a DNA-Carbon Nanotube Hybrid Using Replica Exchange Molecular Dynamics. *Nano Letters* **2009,** *9* (2), 537-541.

Joliffe, I. T. (n.d.). *Principal Component Analysis*. Springer.

Khan, S. B.; Faisal, M.; Rahman, M. M.; Jamal, A., Exploration of CeO(2) nanoparticles as a chemi-sensor and photo-catalyst for environmental applications. *Sci Total Environ* **2011,** *409* (15), 2987-92.

Kodogiannis, V.; Wadge, E., The use of gas-sensor arrays to diagnose urinary tract infections. *Int J Neural Syst* **2005,** *15* (5), 363-76.

Kong, J.; Dai, H. J., Full and modulated chemical gating of individual carbon nanotubes by organic amine compounds. *Journal of Physical Chemistry B* **2001,** *105* (15), 2890-2893.

Krestel, D.; Passe, D.; Smith, J. C.; Jonsson, L., Behavioral Determination of Olfactory Thresholds to Amyl Acetate in Dogs. *Neuroscience and Biobehavioral Reviews* **1984,** *8* (2), 169-174.

Kwak, J.; Gallagher, M.; Ozdener, M. H.; Wysocki, C. J.; Goldsmith, B. R.; Isamah, A.; Faranda, A.; Fakharzadeh, S. S.; Herlyn, M.; Johnson, A. T.; Preti, G., Volatile

biomarkers from human melanoma cells. *J Chromatogr B Analyt Technol Biomed Life Sci* **2013,** *931*, 90-6.

Laird, E. A. K., F.; Steele, G. A.; Grove-Rasmussen, K.; Nygård, J.; Flensberg, K.; Kouwenhoven, L. P., Quantum Transport in Carbon Nanotubes. *Reviews of Modern Physics* **2015,** *87* (3), 703-764.

Lerner, M. B.; Matsunaga, F.; Han, G. H.; Hong, S. J.; Xi, J.; Crook, A.; Perez-Aguilar, J. M.; Park, Y. W.; Saven, J. G.; Liu, R. Y.; Johnson, A. T. C., Scalable Production of Highly Sensitive Nanosensors Based on Graphene Functionalized with a Designed G Protein-Coupled Receptor. *Nano Letters* **2014,** *14* (5), 2709-2714.

Lerner, M. B.; Resczenski, J. M.; Amin, A.; Johnson, R. R.; Goldsmith, J. I.; Johnson, A. T. C., Toward Quantifying the Electrostatic Transduction Mechanism in Carbon Nanotube Molecular Sensors. *Journal of the American Chemical Society* **2012,** *134* (35), 14318-14321.

Lieberzeit, P. A.; Dickert, F. L., Sensor technology and its application in environmental analysis. *Anal Bioanal Chem* **2007,** *387* (1), 237-47.

Makombe, M.; van der Horst, C.; Silwana, B.; Iwuoha, E.; Somerset, V., Antimony film sensor for sensitive rare earth metal analysis in environmental samples. *J Environ Sci Health A Tox Hazard Subst Environ Eng* **2016,** *51* (8), 597-606.

McCulloch, M.; Jezierski, T.; Broffman, M.; Hubbard, A.; Turner, K.; Janecki, T., Diagnostic accuracy of canine scent detection in early- and late-stage lung and breast cancers. *Integr Cancer Ther* **2006,** *5* (1), 30-9.

McCulloch, M.; Turner, K.; Broffman, M., Lung cancer detection by canine scent: will there be a lab in the lab? *Eur Respir J* **2012,** *39* (3), 511-2.

Natelson, D., Nanostructures and nanotechnology. Cambridge University Press: 2015; p 182-191.

Naylor, C. H.; Kybert, N. J.; Schneier, C.; Xi, J.; Romero, G.; Saven, J. G.; Liu, R. Y.; Johnson, A. T. C., Scalable Production of Molybdenum Disulfide Based Biosensors. *Acs Nano* **2016,** *10* (6), 6173-6179.

Niimura, Y., Olfactory receptor multigene family in vertebrates: from the viewpoint of evolutionary genomics. *Curr Genomics* **2012,** *13* (2), 103-14.

Novak, J. P.; Snow, E. S.; Houser, E. J.; Park, D.; Stepnowski, J. L.; McGill, R. A., Nerve agent detection using networks of single-walled carbon nanotubes. *Applied Physics Letters* **2003,** *83* (19), 4026-4028.

Padodara, R., Olfactory Sense in Different Animals. *The Indian Journal off Veterinary Science* **2014,** *2*, 1-14.

Patel, D. J.; Suri, A. K.; Jiang, F.; Jiang, L.; Fan, P.; Kumar, R. A.; Nonin, S., Structure, recognition and adaptive binding in RNA aptamer complexes. *J Mol Biol* **1997,** *272* (5), 645-64.

Pella, P. A., Measurement of the vapor pressures of TNT, 2,4-DNT, 2,6-DNT, and EGDN. *The Journal of Chemical Thermodynamics* **1977,** *9* (4), 301-305.

Persaud, K.; Dodd, G., Analysis of discrimination mechanisms in the mammalian olfactory system using a model nose. *Nature* **1982,** *299* (5881), 352-5.

Petricoin, E. F.; Ardekani, A. M.; Hitt, B. A.; Levine, P. J.; Fusaro, V. A.; Steinberg, S. M.; Mills, G. B.; Simone, C.; Fishman, D. A.; Kohn, E. C.; Liotta, L. A., Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* **2002,** *359* (9306), 572-577.

Radosavljević, M., Freitag, M., Thadani, K. V., & Johnson, A. T., Nonvolatile Molecular Memory Elements Based on Ambipolar Nanotube Field Effect Transistors. *Nano Letters* **2002,** *2* (7), 761–764.

Rusling, J. F.; Kumar, C. V.; Gutkind, J. S.; Patel, V., Measurement of biomarker proteins for point-of-care early detection and monitoring of cancer. *Analyst* **2010,** *135* (10), 2496-511.

Russell, R. A., Tracking chemical plumes in constrained environments. *Robotica* **2001,** *19*, 451-458.

Rutkauskas, D.; Zhan, H.; Matthews, K. S.; Pavone, F. S.; Vanzi, F., Tetramer opening in LacI-mediated DNA looping. *Proc Natl Acad Sci U S A* **2009,** *106* (39), 16627-32.

Sarafoleanu, C.; Mella, C.; Georgescu, M.; Perederco, C., The importance of the olfactory sense in the human behavior and evolution. *J Med Life* **2009,** *2* (2), 196-8.

Scikit-Learn: Machine Learning in Python, Scikit-Learn, July 2017, https://scikit-learn.org/stable/.

SEER Cancer Stat Facts: Ovarian Cancer. National Cancer Institute. Bethesda, MD, https://seer.cancer.gov/statfacts/html/ovary.html.

Staii, C.; Johnson, A. T., Jr.; Chen, M.; Gelperin, A., DNA-decorated carbon nanotubes for chemical sensing. *Nano Lett* **2005,** *5* (9), 1774-8.

Teif, V. B., Ligand-induced DNA condensation: Choosing the model. *Biophysical Journal* **2005,** *89* (4), 2574-2587.

U. S. Food & Drug Administration COVID-19 Vaccines. 2020. (Accessed 25th of November, 2020, at https://www.fda.gov/emergency-preparedness-and-response/coronavirus-disease-2019-covid-19/covid-19-vaccines).

Walker, D. B.; Walker, J. C.; Cavnar, P. J.; Taylor, J. L.; Pickel, D. H.; Hall, S. B.; Suarez, J. C., Naturalistic quantification of canine olfactory sensitivity. *Applied Animal Behaviour Science* **2006,** *97* (2-4), 241-254.

Watson, J.; Whiting, P. F.; Brush, J. E., Interpreting a covid-19 test result. *BMJ* **2020,** *369*, m1808.

What are volatile organic compounds (VOCs)?, U. S. Environmental Protection Agency, August 2019, https://www.epa.gov/indoor-air-quality-iaq/what-are-volatile-organic-compounds-vocs.

Which test is best for COVID-19? Harvard Health Blog. 2020. (Accessed 25th of November, 2020, at https://www.fda.gov/emergency-preparedness-and-response/coronavirus-disease-2019-covid-19/covid-19-vaccines).

Wiederoder, M. S.; Nallon, E. C.; Weiss, M.; McGraw, S. K.; Schnee, V. P.; Bright, C. J.; Polcha, M. P.; Paffenroth, R.; Uzarski, J. R., Graphene Nanoplatelet-Polymer Chemiresistive Sensor Arrays for the Detection and Discrimination of Chemical Warfare Agent Simulants. *Acs Sensors* **2017,** *2* (11), 1669-1678.

Wiederoder, M. S.; Nallon, E. C.; Weiss, M.; McGraw, S. K.; Schnee, V. P.; Bright, C. J.; Polcha, M. P.; Paffenroth, R.; Uzarski, J. R., Graphene Nanoplatelet-Polymer Chemiresistive Sensor Arrays for the Detection and Discrimination of Chemical Warfare Agent Simulants. *Acs Sensors* **2017,** *2* (11), 1669-1678.

Willis, C. M.; Church, S. M.; Guest, C. M.; Cook, W. A.; McCarthy, N.; Bransbury, A. J.; Church, M. R.; Church, J. C., Olfactory detection of human bladder cancer by dogs: proof of principle study. *BMJ* **2004,** *329* (7468), 712.

Wilson, A. D., Diverse applications of electronic-nose technologies in agriculture and forestry. *Sensors (Basel)* **2013,** *13* (2), 2295-348.

Wilson, A. D.; Baietto, M., Advances in electronic-nose technologies developed for biomedical applications. *Sensors (Basel)* **2011,** *11* (1), 1105-76.

Wilson, A. D.; Baietto, M., Applications and advances in electronic-nose technologies. *Sensors (Basel)* **2009,** *9* (7), 5099-148.

Yang, Y.; Peng, F.; Wang, R.; Yange, M.; Guan, K.; Jiang, T.; Xu, G.; Sun, J.; Chang, C., The deadly coronaviruses: The 2003 SARS pandemic and the 2020 novel coronavirus epidemic in China. *J Autoimmun* **2020,** *109,* 102434.

Zheng, M.; Jagota, A.; Semke, E. D.; Diner, B. A.; McLean, R. S.; Lustig, S. R.; Richardson, R. E.; Tassi, N. G., DNA-assisted dispersion and separation of carbon nanotubes. *Nat Mater* **2003,** *2* (5), 338-42.

Zhong, Y.; He, Y.; Ge, Y.; Song, G., beta-Cyclodextrin protected Cu nanoclusters as a novel fluorescence sensor for graphene oxide in environmental water samples. *Luminescence* **2017,** *32* (4), 596-601.