

Portland State University

PDXScholar

Dissertations and Theses

Dissertations and Theses

7-23-2020

Sensing Malignancy: Expanding Upon the Nematode Scent Detection Test

Morgan Alexandra Thompson
Portland State University

Follow this and additional works at: https://pdxscholar.library.pdx.edu/open_access_etds



Part of the [Biology Commons](#)

Let us know how access to this document benefits you.

Recommended Citation

Thompson, Morgan Alexandra, "Sensing Malignancy: Expanding Upon the Nematode Scent Detection Test" (2020). *Dissertations and Theses*. Paper 5528.

<https://doi.org/10.15760/etd.7402>

This Thesis is brought to you for free and open access. It has been accepted for inclusion in Dissertations and Theses by an authorized administrator of PDXScholar. Please contact us if we can make this document more accessible: pdxscholar@pdx.edu.

Sensing Malignancy: Expanding upon the Nematode Scent Detection Test

by

Morgan Alexandra Thompson

A thesis submitted in partial fulfillment of the
requirements for the degree of

Master of Science
in
Biology

Thesis Committee:
Suzanne Estes, Chair
Jason Podrabsky
Brooke Napier

Portland State University
2020

© 2020 Morgan Alexandra Thompson

Abstract

With the growing prevalence of cancer worldwide, research has increasingly focused on methods of early detection since the early detection of cancer can dramatically improve health outcomes for affected individuals. Ideally, early detection systems would be highly reliable yet cost-effective and easy to use. Toward development of such a system, Hirotsu et al. (2015), developed a “Nematode Scent Detection Test” (NSDT) that capitalizes upon the attraction of *Caenorhabditis elegans* nematodes to various tissue samples from cancer patients to detect multiple types and stages of cancer. However, it is currently unknown which G-protein coupled receptors (GPCRs), and/or ligands interacting with these receptors, cause this attractive behavior in *C. elegans*. By determining the GPCRs and ligands responsible, we can develop a detection system that no longer requires the use of an animal. To lay the groundwork necessary to achieve these goals, my thesis sought to: **1)** validate and expand upon the NSDT for samples derived from a wider variety of patient types (e.g., tobacco users) and cancer and benign disease types than those tested by Hirotsu et al. (2015); **2)** determine whether nematodes’ chemotactic response to cancer is strain- or species-specific; **3)** and perform a metabolomics literature review to find possible candidate VOCs that are significantly different in concentration between cancer and healthy patients.

Acknowledgements

Thank you to Dr. Suzanne Estes, my graduate advisor, for taking me on as a graduate student when I was ready to leave the university. I also want to thank Dr. Estes for her wonderful edits on my many pieces of writing and guidance on my project. I want to acknowledge Jennifer Sullins, Ally Yoshioka, Noemi Sarabia Feria, Ariana Montoya and Brent Bever for all their help on my project from plate making to coming in on weekends to age-sync. Thank you to Dr. Josiah Wagner and the CEDAR group that provided funding, expertise and biological samples for this project. I would like to thank my family and friends for being supportive of me during my graduate program. Finally, thank you to Portland State University for the Vision 2025 grant to Dr. Suzanne Estes that helped to fund this project. And finally, I want to thank my committee members for all their advice on my project.

Table of Contents

Abstract.....	i
Acknowledgements.....	ii
List of Figures.....	v
List of Abbreviations.....	vii
Introduction.....	1
Methods.....	15
Results.....	24
Discussion.....	36
Conclusions & Future Research.....	46
References.....	49
Appendix A. Table of prostate urine sample cohort patient information.....	59
Appendix B. Table of metabolomic studies on VOC profiles from both cancer and healthy patients.....	71
Appendix C. Table of specific VOCs from the metabolomic studies that were significantly different between cancer and healthy patients.....	82

Appendix D. Multivariate correlation of entire data set with three sample types (Healthy, Negative Screen and PAC)	83
Appendix E. Multivariate correlation of PAC patient data.....	84
Appendix F. Multivariate correlation of healthy patient data.....	85
Appendix G. Multivariate correlation of negative screen patient data.....	86

List of Figures

Figure 1: Estimated cancer-related deaths in the United States in 2018.....	1
Figure 2: Weaknesses of PSA testing for the early detection system for prostate cancer.....	3
Figure 3: Neurological network in <i>C. elegans</i>	8
Figure 4: Chemotaxis behavior of <i>C. elegans</i> with a uniform concentration of attractant...	10
Figure 5: GPCR activation via extracellular ligand binding.....	11
Figure 6: <i>C. elegans</i> chemotactic response to urine from cancer and healthy patients.....	12
Figure 7: Hirotsu et al. (2015) <i>C. elegans</i> chemotaxis assay.....	16
Figure 8: Chemotaxis assay plate set up.....	19
Figure 9: A dilution of 10^{-2} elicits the strongest attractive response of wildtype N2 <i>C. elegans</i> to HCC and PAC urine.....	24
Figure 10: Wildtype N2 <i>C. elegans</i> are more attracted to urine from HCC patients than to urine of healthy and liver cirrhosis patients.....	26
Figure 11: Wildtype N2 <i>C. elegans</i> are more attracted to urine from HCC patients than to urine from healthy and negative screen patients.....	26
Figure 12: Wildtype N2 <i>C. elegans</i> are more attracted to urine from PAC patients than to healthy and negative screen patient urine.....	27

Figure 13: Wildtype N2 <i>C. elegans</i> are more attracted to urine from PAC patients than to healthy and negative screen patient urine.....	28
Figure 14: Wildtype N2 <i>C. elegans</i> are slightly more attracted to urine from PAC patients categorized as having tumor types 2 and 3 than to those with type 1 or no tumor (type 0).....	29
Figure 15: N2 Ancestral <i>C. elegans</i> are more attracted to urine from PAC patients than to healthy and negative screen patient urine.....	31
Figure 16: <i>C. briggsae</i> are attracted to urine from PAC, healthy and negative screen patients.....	31
Figure 17: N2 Ancestral <i>C. elegans</i> are attracted to urine from PAC patients.....	32
Figure 18: <i>C. briggsae</i> are attracted to urine independent of type (PAC, healthy and negative screen).....	32

List of Abbreviations

BMI – Body mass index

CI - Chemotaxis Index

CIs – Chemotaxis Indices

DRE – Digital rectal exam

GPCR – G-protein coupled receptor

HCC – Hepatocellular carcinoma

NaOH – Sodium hydroxide

NGML – Nematode growth medium, lite

NSDT - Nematode Scent Detection Test

PAC – Prostate adenocarcinoma

PSA – Prostate specific antigen

VOC – Volatile organic compound

Introduction

The Worldwide Impact of Cancer and the Need for an Early Detection System

Owing to a lack of early detection diagnostics and ineffective treatments for later-stage cancerous tumors, cancer has become a global phenomenon (Hirotsu et al., 2015). Cancer is a leading cause of death worldwide, with an estimated 17 million deaths by 2030 (Boyle & Levin, 2008). In the United States alone, there were an estimated 610,000 deaths and 1.7 million

new cases of cancer that arose during 2018, with the majority of these cases resulting from breast and

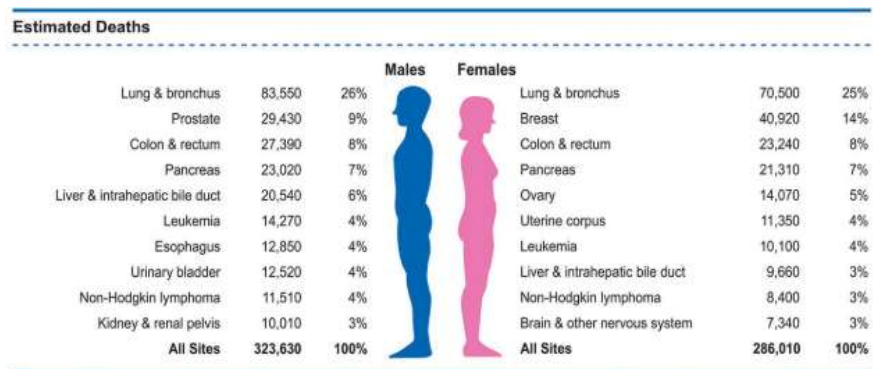


Figure 1: Estimated cancer-related deaths in the United States in 2018. The top three cancer types for men were lung, prostate and colon cancer; the top three cancer types for women were lung, breast and colon cancer. From Siegel et al. (2019).

prostate cancer (Fig. 1) (Siegel, Miller & Jemal, 2019). To combat this epidemic, the National Institutes of Health (NIH) estimates the United States will spend approximately \$158 billion in 2020 on health care specifically related to cancer, which is an increase of 27 percent from 2010 (<https://www.nih.gov/news-events/news-releases/cancer-costs-projected-reach-least-158-billion-2020>). Individually, cancer has the highest per-person medical costs out of the five health conditions that most contribute to high medical costs in the United States (Narang & Nicholas, 2017). Unfortunately, these individualized costs

were expected to grow 40% from 2010 to 2020, a trend attributed primarily to increasing costs for cancer treatments, health services, and expected changes in demographics. Additionally, since most individuals who undergo cancer treatment are normally over the age of 65, Medicare, a federally funded health care program for individuals over 65, will hold most of the burden (Narang & Nicholas, 2017).

As the burden of cancer increases both domestically and internationally (Siegel, Miller & Jemal, 2019), research has increasingly focused on methods of early detection since early detection dramatically improves health outcomes for affected individuals. Unlike other illnesses, such as diabetes, laboratory diagnostics for cancer remain limited (Schiffman, Fisher & Gibbs, 2015). Currently, tumor markers—cancer-derived substances that may be elevated in blood, urine or tissues of individuals with cancer—are the primary system for early screening of potential cancer cases (Schiffman, Fisher & Gibbs, 2015). For instance, prostate-specific-antigen (PSA) testing was first utilized in 1979 for diagnosing prostate cancer (Thompson et al., 2004). A value of 2.6 ng/ml was originally determined to be the upper end of the normal range for PSA levels in patient serum. Prior to PSA testing, prostate cancer was diagnosed via a digital rectal exam (DRE), which was ineffective as it was only able to detect cancer after it had already metastasized. However, if the DRE results were abnormal, the patient underwent a prostate biopsy with digital guidance to take

THE PSA TEST AND WHY ITS RESULTS CAN BE CONFUSING

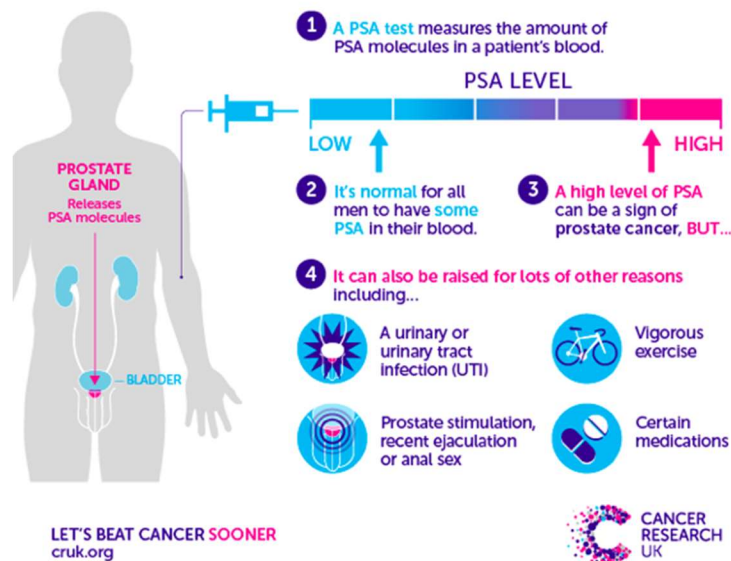


Figure 2: Weaknesses of PSA testing for the early detection system for prostate cancer. PSA occurs naturally in both cancerous and non-cancerous prostate tissue, and there are many factors that may either decrease or increase PSA values, independent of cancer within the body, including exercise, urinary tract infections and various medications. From cruk.org.

up to four biopsy samples. A study including data from multiple institutions examined PSA values and DREs from 6,630 men over the age of 50 and concluded that the upper limit for the normal range of PSA should be increased to 4 ng/ml of PSA present in patient serum (Thompson et al., 2004). However, PSA is excreted by both malignant

as well as nonmalignant prostate cells and can therefore be present in the serum of patients (Fig. 2) who have benign diseases such as prostatitis and prostatic hyperplasia (Eastham, 2017). A meta-analysis by Mistry & Cable (2003) found that the sensitivity and specificity of PSA tests were 72.1% and 93.2%, respectively. Sensitivity is a measurement of how well a diagnostic test correctly detects a patient with the disease of interest while specificity shows how well a test correctly distinguishes a healthy patient from a patient with the disease of interest (Parikh et al., 2007).

Additionally, tumor markers for cancer, many of which can also be produced by normal cells, have shown to be inaccurate and inefficient, especially among the most common types of cancer (Schiffman, Fisher & Gibbs, 2015). The most common forms of cancer detection currently in use include the PSA test for prostate cancer, carcinoembryonic antigen for breast cancer, cancer imaging for multiple cancer types, and biopsies, which are invasive and often expensive tests that can also lead to overdiagnosis and overtreatment of different cancer types and stages (Schiffman, Fisher & Gibbs, 2015).

Recently, early cancer detection studies using animal models have become more prevalent. These studies trained animals with advanced olfactory systems, such as dogs (Cronu et al., 2011) and mice (Sato et al., 2017), to correctly distinguish cancer from control individuals using patient urine. Trained dogs were able to detect prostate cancer patient from control patient urine, independent of cancer stage, with sensitivity and specificity values of 91% (Cronu et al., 2011). Similarly, mice were trained to distinguish bladder cancer patient urine from controls and were able to discriminate all eleven bladder cancer patient urine samples from control patient urine (Sato et al., 2017). However, the cost and time involved in breeding, housing and training animals to detect cancer are prohibitive. Additionally, the olfactory systems of such mammals are extremely complex, giving little hope of determining which receptors or ligands are involved in these animals' ability to detect cancer.

Cell Cycle Dysregulation and the Progression of Cancer

The normal progression of the cell cycle involves the faithful duplication and eventual segregation of genetic material into daughter cells. This process ultimately allows for proper gene expression and production of proteins that regulate metabolic pathways. When genetic mutations arise, however, cell cycle dysregulation can result. Such mutations can result from intrinsic sources such as DNA replication errors, or from extrinsic sources such as exposure to UV radiation from sunlight, carcinogens, or ionizing radiation. The cell cycle contains many checkpoints to preserve the integrity of the replication process and ensure that mutations are not transmitted. However, when a mutation negatively affects checkpoint systems in the cell cycle, outcomes can range from cell death to reprogramming of the cell cycle, which can lead to cancer formation (Barnum and O'Connell, 2014).

Due to the high turnover rate of many somatic cell types, the majority of cancers are sporadic rather than heritable, caused by mutations resulting from replication errors within these cells (Bertram, 2000). Since adult humans contain around 10^{15} cells, a large target for such mutations exists. Cancers therefore tend to affect cell types with greater proliferative potential and opportunity for exposure to carcinogens; e.g., epithelial tissue (Bertram, 2000).

Cancer progresses by reprogramming the metabolic pathways of affected cells to allow excessive expansion and invasion of various tissue types by cancerous cells (Sarkar et al., 2013). This uncontrolled cell growth occurs as a result of oncogene activation alongside deactivation of genes that suppress tumor growth, as well as the down-regulation of

apoptotic systems. Cancer hijacks metabolic and protein synthesis pathways to overexpress various cellular components. For instance, these cells can exhibit a shift to aerobic glycolysis, consuming massive amounts of glucose for conversion to lactate, which is secreted from the cell as a byproduct (Vazquez et al., 2016). This process is known as the Warburg effect (Warburg et al., 1927). As such, these proliferating cells begin to overproduce glucose transporters and glycolytic enzymes (Heiden & DeBerandinis, 2017). However, the Warburg effect, also known as aerobic glycolysis, also appears in non-cancerous cells that are involved with the immune system, angiogenesis and pluripotency (Abdel-Haleem et al., 2017). Non-cancerous cells invoke rapid proliferation in response to many processes such as inflammation, autoimmune disorders and for regulation of physiological process such as T-cell production (Abdel-Haleem et al., 2017). As a result, the Warburg effect is associated with rapid proliferation of both cancerous and non-cancerous cells (Abdel-Haleem et al., 2017).

Fuel sources beyond glucose are also required for energy conversion, biomass generation, and control of redox reactions within cancerous cells (Heiden & DeBerandinis, 2017). Fatty acids, branched chain amino acids, serine, glycine and many other nutrients have been found necessary for cancer metabolism (Heiden & DeBerandinis, 2017). As byproducts of this metabolism, cancer cells also secrete a variety of compounds including volatile organic compounds (VOCs), such as various amino acids, ammonia and urea, which help to form the tumor microenvironment (Dang, 2017). Such discoveries began a new area of study known as volotomics, the study profiles of VOCs excreted in the

presence and absence of a disease (Borza et al., 2015). VOCs can be measured from various bodily fluids including, but not limited to: urine, serum, breath, skin, feces and/or saliva. These VOCs are categorized into exogenous and endogenous. Exogenous VOCs originate from external factors such as smoking, drugs, ingested foods, and absorption through the skin, whereas endogenous VOCs arise from both normal and abnormal metabolic processes (Borza et al, 2015). Many VOC profiles have been determined from different studies across many cancer types, but currently, no single VOC(s) has proven to reliably distinguish cancer from healthy sample types.

There is a clear need for an accurate and efficient system for earlier cancer prognoses (Schiffman, Fisher & Gibbs, 2015). Strides have been made toward increasing and improving technology available to detect cancer progression in the early stages, saving lives, money and time; however, diagnostics continue to be specific to advanced stages and cancer types (Schiffman, Fisher & Gibbs, 2015). Ideally, future early detection systems would need to be highly reliable across many stages and types of cancer while remaining cost-effective and easy to use, unlike the currently available technology.

C. elegans Olfactory System

Caenorhabditis elegans is commonly utilized as an experimental model in neurobiology because of its simple nervous system, which is primarily devoted to chemosensation—as is a sizeable fraction of its genome (Bargmann, 2006). This species has the exceptional ability to detect a wide variety

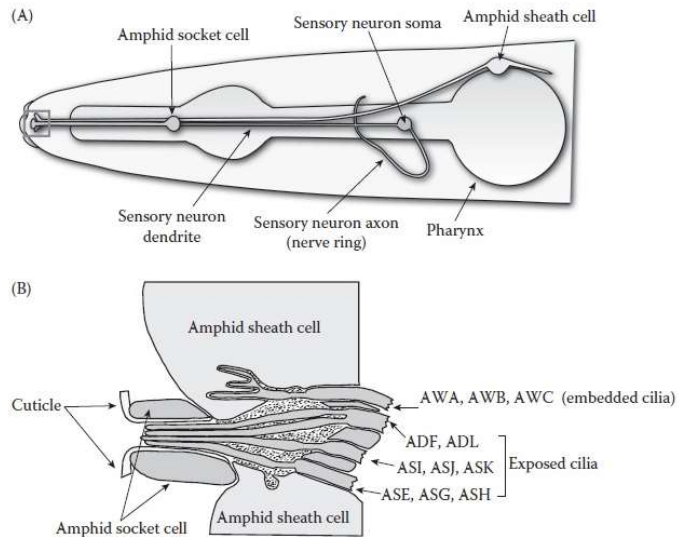


Figure 3: Neurological network in *C. elegans*. Within the amphid sheath cells of the neurological network in the head of *C. elegans*, there are a series of neurons that are involved in chemotaxis behaviors, both attractive (AWA and AWC) and repellant (ASH, ADL and AWB). From [A] Hall et al. (1991) & [B] Perkins (1986).

of volatile and water-soluble compounds necessary to differentiate food, mates, pathogens and predators encountered in its natural environment (Bargmann et al., 1993; Bargmann, 2006). *C. elegans* is a bacterivorous species whose natural habitat is decaying vegetation, which involves low oxygen and high humidity (Schulenburg & Félix, 2017). Thus, *C. elegans* can be reliably found in vegetable gardens, orchards, wooded areas and compost heaps across all continents except Antarctica (Schulenburg & Félix, 2017). However, since limited natural sampling of *C. elegans* has been conducted, there may be additional habitats in which *C. elegans* thrive (Schulenburg & Félix, 2017; Kiontke and Sudhaus 2006).

C. elegans hermaphrodites have 302 neurons, which fall into 14 different categories, five of which detect volatile chemicals. One group, known as amphid neurons, are located in the head of the worm and include the AWA, AWC, ASH, ADL and AWB neurons (Bargmann et al., 1993; Fukuto et al., 2004). ASH, ADL and AWB neurons are used to detect repellent chemicals whereas AWA and AWC neurons detect attractive substances (Fig. 3) (Fukuto et al., 2004). Currently, there are two known ligands for *C. elegans* G-protein coupled receptors (GPCRs): diacetyl and 2-heptanone (Sengupta, Chou & Bargmann, 1996; Zhang et al., 2016). Diacetyl binds to odr-10, a GPCR embedded in the attractive AWA amphid neuron, causing an attractive behavioral response (Sengupta, Chou & Bargmann, 1996). 2-heptanone binds str-2, a GPCR found embedded in the attractive AWC amphid neuron, eliciting an attractive response from *C. elegans* (Zhang et al., 2016). In response to chemical stimuli, *C. elegans* movement involves two behaviors: sinusoidal-swimming movements called “runs” and sharp turns called “pirouettes” (Pierce-Shinmomura, Morse & Lockery, 1999). Pirouettes occur when the *C. elegans* head curls back towards its tail, known as an “omega turn”. When a *C. elegans* is placed on an agar plate with a uniform dispersal of 2 mM NH₄Cl, a known attractant, they respond making many runs and pirouettes, advancing toward the chemical stimulus (Fig. 4a). By tracking and analyzing the movements of multiple such *C. elegans*, Pierce-Shinmomura, Morse & Lockery (1999) developed a map showing the probability of *C. elegans* movement to various locations on an agar plate containing a uniform concentration of attractant (Fig. 4b). From this, they determined that *C. elegans* primarily remained in a 1 cm diameter from

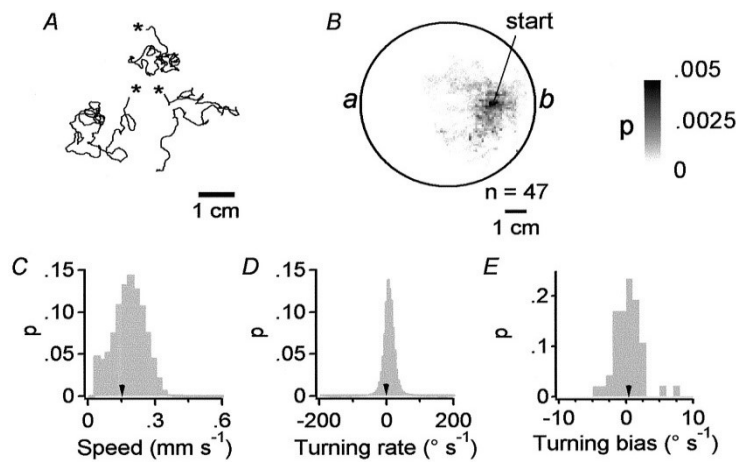


Figure 4: Chemotaxis behavior of *C. elegans* with a uniform concentration of attractant. [A] Three *C. elegans* movement tracks on an agar plate with a uniform concentration of 2 mM NH_4Cl , run independently. [B] The probability per unit of area movement of individual *C. elegans* based on 47 individual runs. [C-E] Speed, turning rate, and turning bias for the individual *C. elegans* tested in B. From Pierce-Shinmura, Morse & Lockery (1999).

their initial starting point (Fig. 4b), also documenting the mean and variance of animal speed, turning rate and turning bias (Fig. c-e). From such data, the authors showed that, in the presence of a uniform attractant, *C. elegans* will remain at their initial locus unless a stronger attractant is present, thus demonstrating their chemical-dependent

olfactory system.

The detection of attractive compounds is thought to be mediated by binding and subsequent phosphorylation of G protein-coupled receptors (GPCRs) expressed in the AWA and AWC chemosensory neurons (Fig. 5). GPCRs are responsible for processing olfactory signals (Fukuto et al., 2004) and transducing signals for a variety of other biological processes. For olfaction to occur, a ligand must interact with a G-protein coupled receptor (GPCRs) on the membrane of neurological cells (Fig. 5) (Sengupta et al., 1996). Only two of the estimated 1,000 – 2,000 GPCRs in *C. elegans* have known ligands as named above (Bargmann et al., 1993; Tobin, 2008; Vidal et al., 2018).

C. elegans Use in Early Cancer Detection

Hirotsu et al. (2015) developed a “Nematode Scent Detection Test” (NSDT) that capitalizes on the ability of *C. elegans* to sense cancer via olfaction. Cancer cells excrete numerous

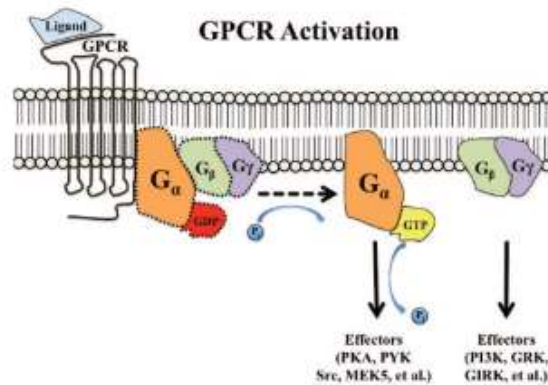


Figure 5: GPCR activation via extracellular ligand binding. When the ligand for the GPCR binds, the GPCR changes conformation and allows for the activation of a G protein-coupled receptors (GPCRs) downstream in the pathway when GDP is exchanged for GTP on the alpha subunit. Once the alpha subunit is active, the beta and gamma subunits detach from the alpha subunit to allow their signal transduction. From Belmonte & Burns (2011).

compounds as metabolic byproducts—many of the same byproducts generated by lactic acid bacteria, a primary *C. elegans*’ food source to which it is strongly attracted (Lee et al., 2017). For instance, bacteria emit a variety of volatile organic compounds (VOCs), and several of these VOCs (e.g., acetone and benzaldehyde) are also found in exhaled breath from patients with several cancer types including lung and pancreatic

cancer (Barash et al., 2009; Hirotsu et al., 2015; Markar et al., 2018).

Hirotsu et al. (2015) found that, when *C. elegans* were given the choice between healthy and cancerous urine samples, they had an overwhelming preference for the samples from cancer patients regardless of the stage (including stage 0) or type of cancer presented (Fig. 6) (Hirotsu et al., 2015). Specifically, *C. elegans* tended to chemotax toward cancer urine

samples at a dilution of 10^{-1}

¹ and away from healthy

urine samples presented

on a Petri plate, which

were quantified as

positive and negative

“chemotaxis indices”

(CIs), respectively.

Specifically, they showed

that urine samples from patients

with different types and stages of

cancer at a 10^{-1} dilution,

generally elicited positive CIs

while control samples elicited either no response or were slightly repulsive to worms,

represented by a negative CI (Fig. 6). More specifically, the chemotaxis index (CI)

measured for a population of age-synchronous worms describes the proportion of worms

that migrated to a stimulus as [(number migrating to sample of interest) – (number not

migrating to sample of interest)] / (all worms). A positive CI indicates an overall preference

for the sample of interest. Hirotsu et al. measured a positive average CI of around 0.3 for

urine samples taken from cancer patients diagnosed with different types and stages of

cancer, suggesting that there are specific compounds shared among the cancer types tested

that are attractive to *C. elegans*. Nematodes were similarly attracted to diluted urine from

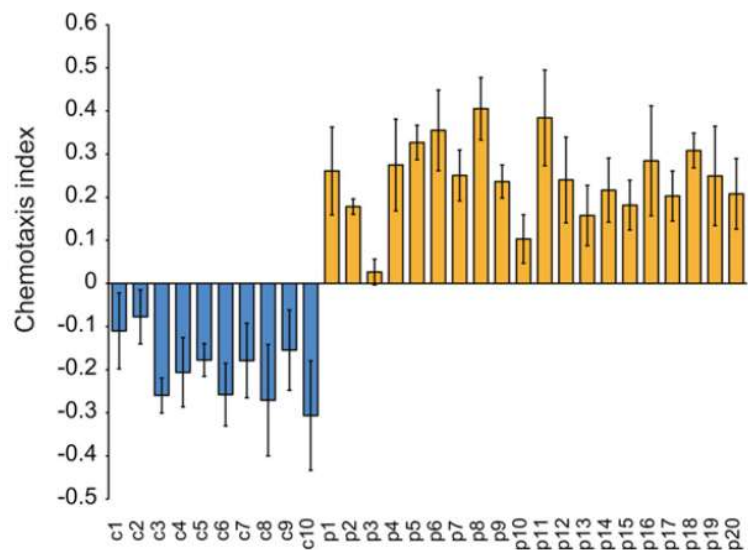


Figure 6: *C. elegans* chemotactic response to urine from cancer and healthy patients. Chemotaxis of wild type *C. elegans* to diluted urine samples from cancer patients (yellow) compared to healthy individuals (blue) at a 10^{-1} dilution. Error bars represented SEM. From Hirotsu et al. (2015).

cancer patients and repulsed by urine from healthy individuals (Fig. 6). Conversely, *C. elegans* could not reliably distinguish serum samples from cancer affected versus healthy individuals, perhaps due to other interfering substances present in serum.

Hirotsu et al. (2015) utilized urine samples stored at -20°C prior to testing from only pre-treatment patients and excluded patients on the basis of age (> 20 years old), physical symptoms (e.g., headache, chest or abdominal distension, appetite, weariness, cough, bloody feces, constipation, and diarrhea), pregnancy, current medicine usage, alcohol usage (<3 days per week a drink was consumed), and tobacco use (less than 2 weeks prior to participating). Whether the NSDT is useful for additional cancer types and in wider variety of patient and sample types remains untested.

Research Aims

Because the ligand-receptor interactions that fuel *C. elegans* chemotactic behavior are not currently known, we have essentially no known candidate GPCRs or ligands that may be responsible for *C. elegans* cancer detection. Identifying these interactions (or the cancer metabolite that *C. elegans* is detecting when it “smells” cancer) would allow for development of an animal-free early detection system. To lay the groundwork necessary to achieve these goals, my thesis expands upon the work of Hirotsu et al. (2015) to: **1)** validate and expand upon the NSDT for samples derived from a wider variety of patient types (e.g., tobacco smokers) and cancer and benign disease types; **2)** determine whether nematodes’ chemotactic response to cancer is strain- or species-specific; **3)** and perform a metabolomics literature review to find possible candidate VOCs that are significantly different in concentration between cancer and healthy patients.

Methods

Nematode strains and culture conditions. Three separate strains of nematodes were utilized in chemotaxis assays: the standardly utilized wildtype Bristol N2 (referred to hereafter as “N2”), an “ancestral N2” strain, and a strain of *C. elegans*’ congener species, *C. briggsae*, PB800. The ancestral N2 strain was isolated and cryogenically stored in 1968, while the standardly utilized N2 experienced many generations of lab culture before being cryogenically stored as separate strain in 1980 (WormBase).

Following Hirotsu et al. (2015), nematodes were maintained at 20 °C on NGM Lite (NGML) plates seeded with NA22 *E. coli* as a food source. Prior to all experiments, strains were allowed to recover from freezing for two to three generations prior to being age synchronized by standard bleach treatment.

Age-synchronization. In accordance with Hirotsu et al. (2015), young adult stage nematodes were used for all chemotaxis assays. Mixed populations of nematodes were grown on NGM Lite (NGML) plates seeded with NA22 *E. coli*. Nematodes were washed into 15 ml conical tubes with S-Basal buffer (100mM NaCl, 5.75mM K₂HPO₄, 45mM KH₂PO₄, and Milli-Pore water). The conical tubes were spun at 800 rpm for 1 minute, and the supernatant was poured off. Nematodes were washed 3 more times with 15 ml of S-Basal buffer at 800 rpm for 1 minute prior to bleaching for age synchronization. To release eggs from the adult nematodes, samples were treated with a 1:2 solution of 5M sodium hydroxide (NaOH) and 2.75% bleach, respectively. After the third wash, 5 ml of S-Basal buffer was added to each conical tube, followed by 2 ml of the NaOH-bleach solution.

Every 2 minutes, the conical tubes were inverted 4-6 times. After approximately 12 minutes, the adults break apart, releasing eggs. The conical tubes were then quickly spun down at 800 rpm for 1 minute, and the supernatant poured off. Samples were again washed 3 times with 15 ml of S-Basal buffer at 800 rpm for 1 minute. After the final wash, the supernatant was poured off, and the remaining 1 ml of buffer and egg solution, was pipetted onto NGML plates seeded with NA22 *E. coli*. The plated eggs were kept at 20 °C for 72 hours at which time they had reached the young adult stage.

Specific Aim 1. Validate and expand on Hirotsu et al. NSDT for samples derived from a variety of patient, cancer and benign disease types.

Hirotsu et al. (2015) NSDT. I

first utilized the chemotaxis assay of Hirotsu et al. (2015), which used unseeded 9 cm Petri plates containing NGML and 200 mg/ mL streptomycin that were marked to contain three sections (Fig. 7). Points A and B are 2.5 centimeters from the center of the plate. Prior to plating the diluted urine samples, 1 µl of 1 M sodium azide (NaAz)

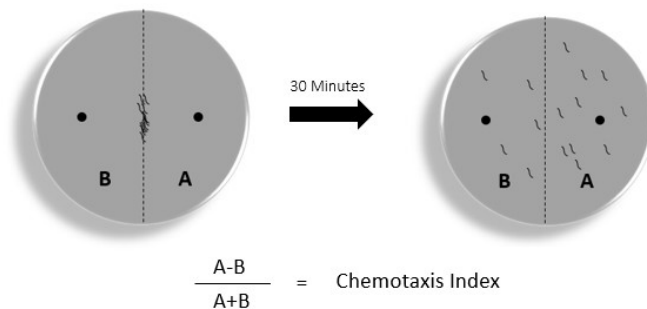


Figure 7: Hirotsu et al. (2015) *C. elegans* chemotaxis assay. 50-100 wildtype N2 *C. elegans* are plated onto the midline of a 9 cm plate treated with an attractant (side A) and control (side B), each 2.5 cm from the center. 1 µl of NaAz was plated on points A and B prior to plating the attractant and control to kill *C. elegans* once they reached either point A or B. A center area 0.5 cm wide is marked to control for *C. elegans* that are unable to move or chemotax and will therefore not be included in the CI calculations. *C. elegans* are allowed to chemotax for 30 minutes. A CI was determined from counts of animals that did and did not move toward the attractant.

was plated on points A and B to kill *C. elegans*' chemotaxing towards either point A or B. On point A, 1 μ l of sample was plated while on point B, 1 μ l of control was plated. To control for *C. elegans* that were incapable of movement or scent detection, a third section was created 0.5 cm from the center of the plate. After 30 minutes, *C. elegans* that chemotaxed into sections A and B were counted and a CI was determined for each sample as $CI = [(number\ migrating\ to\ sample\ of\ interest) - (number\ not\ migrating\ to\ sample\ of\ interest)] / (all\ worms)$.

Urine and serum samples. Urine samples from patients were obtained from collaborators at The Knight Cancer Institute's Cancer Early Detection Advanced Research Center (CEDAR, Portland, OR). We obtained two cohorts of urine samples: "HCC" and "PAC". The HCC cohort contained 3 patients for each of three sample types: hepatocellular carcinoma (HCC), healthy, and liver cirrhosis. The PAC cohort contained 20 urine samples from prostate adenocarcinoma (PAC) patients, 19 samples from healthy patients, and 28 samples from "negative screen" patients (Appendix A). For the HCC cohort, healthy patients were deemed as those who did not have HCC or liver cirrhosis. For the PAC cohort, healthy patients had high PSA levels but upon a biopsy of their prostate tissue, their tissue was either normal or benign; and negative screen patients, had low PSA values (Appendix A). For the HCC cohort, urine samples were frozen immediately upon collection, thawed and aliquoted, then stored at -80°C as 50 μ l aliquots prior to being thawed for each chemotaxis assay setup. For the PAC cohort, urine samples were frozen

immediately upon collection, thawed at room temperature, and freshly diluted using autoclaved milliQ water prior to each chemotaxis assay setup.

For the PAC cohort, we received clinical data for each patient including at least some of the following: PSA (ng/ml), clinical diagnosis, prostate volume (cm³), tobacco use (yes = 1/no = 0), body mass index (BMI) (kg/m²), age at time of sample collection, and tumor grade (Gleason score for each tumor) (Appendix A). If a patient develops a cancerous tumor, the “Gleason grading system” is used to score its severity by evaluating the different histological patterns that arise in various stages of prostate cancer (Gordetsky & Epstein, 2016). The Gleason grading system was derived from a study conducted through the Veteran’s Affairs Cooperative Research Group between 1959 and 1964 by pathologist, Donald Gleason. Gleason noted that prostate tissue was composed of primarily two histological patterns and therefore added the two grade patterns together, each on a scale of 1-5, to create an overall Gleason score on a scale of 2-10 (Gleason & Mellinger, 1974; Gordetsky & Epstein, 2016). With both histological grade groups (1-5) and Gleason score values (2-10) the higher the value the more severe the prostate cancer prognosis. Two patients can have the same Gleason score value but have different severities. For example, a Gleason score of 7 can result from two grade group combinations of either 3+4 or 4+3, with 4+3 being more severe than 3+4. This means that a patient with 4+3, and a Gleason score of 7, would have a histological grade group of 4 most prevalent and 3 the second most prevalent. Additionally, higher stages of prostate cancer (i.e., stages 3 and 4) usually result in a Gleason score of 7 and above, while lower stages of prostate cancer (i.e., 1 and

2) usually result in a Gleason score of less than 7 (Gleason & Mellinger, 1974; Gordetsky & Epstein, 2016).

Alterations to NSDT. After an extensive series of pilot tests demonstrating that *C. elegans* exhibited weak and variable

chemotactic responses to HCC

and PAC samples alongside a

weak attraction to 1 M NaAz,

which was used to kill

nematodes in the Hirotsu assay,

I devised a modified NSDT.

Following Margie et al. (2013),

unseeded NGML plates

prepared without streptomycin

were divided into 4 equally

sized quadrants (Fig. 8). A

circle with a diameter of 1 cm

was drawn around the center of

the plate. Quadrants across from one another, were labelled with “T” for the sample and

the other two quadrants were labelled “C” for control. Each quadrant contained a dot

exactly 2.5 cm from the center of the plate for placement of 1 µl of each sample or control.

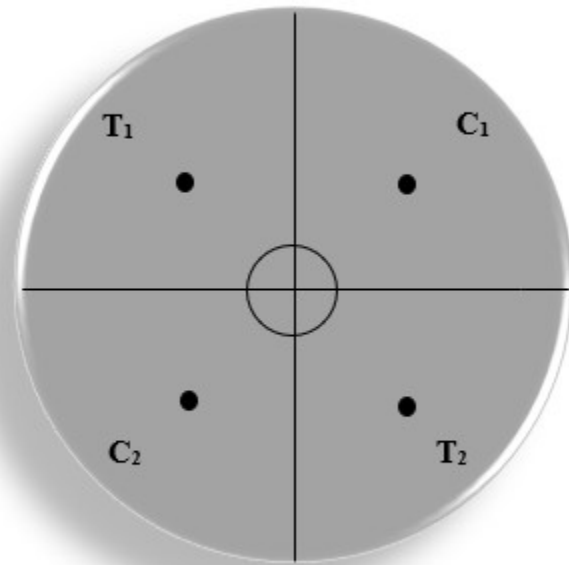


Figure 8: Chemotaxis assays plate set up. 1µl of a diluted urine sample was plated on the T₁ and T₂ spots, which were 2.5 cm from the origin. 1 µl of autoclaved milliQ water was placed on both the C₁ and C₂ spots. The origin circle had a diameter of 1cm, where ~50 young adult *C. elegans* were plated.

After each chemotaxis plate was prepared, approximately 50 washed, young adult *C. elegans* were plated in the center (Fig. 8). We further controlled for light and surrounding odors by using clean, unscented towels to cover the chemotaxis plates during assays. The plates were left on the benchtop, covered with towels for 1 hour at 21 °C. After 1 hour, the plates were individually imaged and the number of *C. elegans* in each quadrant was recorded.

To determine the CI, the previously described equation was used where $T_1 + T_2$ was the number of *C. elegans* that chemotaxed into the quadrants with the diluted urine samples, and $C_1 + C_2$ is the number of *C. elegans* that chemotaxed into the quadrants with the controls (Fig. 8). The CI in this case is the total number of *C. elegans* that chemotaxed towards a diluted urine sample, minus the total number of *C. elegans* that chemotaxed towards the control, divided by the total number of *C. elegans* in each quadrant. To control for *C. elegans* that were immobile or incapable of olfaction, all animals in the 1 cm circle near the center were left uncounted (Fig. 8).

Data analysis. Using a nested analysis of variance (ANOVA), differences in nematode chemotactic responses to different urine types (healthy, benign disease and cancer), and to different patients within each sample type, were determined for the liver (HCC) and prostate (PAC) cancer cohorts separately. The model $y = \mu + \text{type} + \text{individual}(\text{type}) + \varepsilon$ was tested in each case. ANOVAs were followed by post hoc Tukey's honestly significant difference (HSD) tests ($\alpha = 0.05$) using least squares means to test for differences between individual pairs of sample types.

Since a variety of clinical data were available for the PAC cohort, pairwise correlation coefficients were calculated between various clinical traits and between these traits and the average CIs for each patient. For these analyses, we also created a trait, “tumor type”, coded as 0 for normal tissue, 1 for benign tissue, 2 for cancerous tumor with a Gleason score of 3+3 or 3+4, and 3 for any cancerous tumor with a Gleason score of 4+3 or above. Correlations were calculated using the entire dataset and, where appropriate, each urine type (PAC, healthy, negative screen).

Sensitivity and specificity calculations. Following Hirotsu et al. (2015), sensitivity and specificity were calculated for each sample set and each nematode strain. Sensitivity was calculated as the total number of true positives over the combined number of true positives and false negatives, while specificity was calculated as the total number of true negatives over the combined number of true negatives and false positives. True positives were positive CI values that correctly detected cancer, while true negatives were negative CI values that correctly detected samples from either healthy or benign disease patients. False positives were negative CI values that incorrectly detected samples from either healthy or benign disease patients, while false negatives were positive CI values that incorrectly detected cancer.

Specific Aim 2. Determine whether nematodes’ chemotactic response to cancer is strain- or species-specific. Chemotaxis assays were performed as described above using N2 ancestral *C. elegans* and *C. briggsae* using urine from a subset of PAC cohort patients.

The same nested ANOVA model was tested for each strain to test the effect of sample type (PAC, healthy, negative screen) on CIs. These analyses were again followed by Tukey's HSD tests.

Specific Aim 3. Identify candidate VOC ligands.

Toward the goal of identifying candidate ligands to utilize in future NSDT assays, I conducted a preliminary literature review of 19 cancer metabolomic studies. Metabolomic studies performed on multiple types of cancer were analyzed for VOCs that were statistically significantly over- or underrepresented in cancer patients compared to healthy patients. These studies were included based on the age of the paper, number of control samples compared to the number of cancer samples tested, whether the samples came from individuals treated for cancer or not, how many VOCs were found in their study, and how many factors were controlled for (Appendix A). More specifically, I chose papers that were published no earlier than 2012 seeing that the capabilities of research for detecting specific VOCs has matured extensively since that time. I also found that papers published prior to 2012 often stated that VOC profiles were different overall between cancer and healthy patients without identifying any specific VOCs. For the number of samples tested, I eliminated papers that had fewer than 15 total patient urine, serum or breath samples; however, for papers that used cell culture, I eliminated papers that had fewer than three different cell lines. I also eliminated papers that controlled for fewer than three external variables. These variables included smoking status, body mass index, exercise habits, diet

and many others. Additionally, I eliminated publications that reported fewer than two VOCs in their findings to prevent papers with biases for specific VOCs. The VOCs found to be statistically significantly different between control samples and cancer samples among these studies were compiled. I also used the Human Metabolomic Database (<http://www.hmdb.ca/>) to compile VOC information including chemical formula, natural state of the VOC, and whether the VOC is considered to be of endogenous or exogenous origin.

Results

Specific Aim 1. Validate and expand Hirotsu et al. NSDT for samples derived from a variety of patient, cancer and benign disease types

Following the exact methods of Hirotsu et al. (2015), we tested a series of dilutions (0, 1:10, 1:25, 1:50) made from healthy, liver cirrhosis and hepatocellular carcinoma (HCC) sample types using autoclaved MilliQ water. These assays included many technical replicates from three individuals of each type. We found no significant difference between average CIs for any dilution of HCC urine compared to the urine from patients with liver cirrhosis and patients that were healthy (data not shown). Because we were unable to

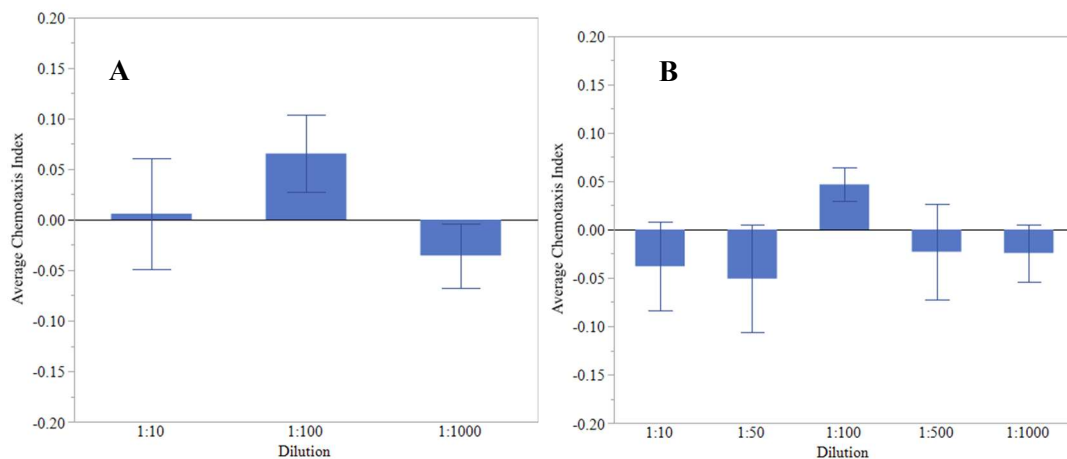


Figure 9: A dilution of 10^{-2} elicits the strongest attractive response of wildtype N2 *C. elegans* to HCC and PAC urine. (A) Average CI obtained from HCC patient urine samples at three dilutions ($n = 3$ patients, 18-22 assays each). (B) Average CI obtained from PAC urine samples at five dilutions. Error bars represent 1 S.E.M.

replicate the results of Hirotsu et al. (2015), we adapted the methodology of Margie et al. (2013) for our chemotaxis assays, which allowed us to additionally control for random movement, surrounding odors and light that could affect *C. elegans* behavior.

Using the methodology of Margie et al. (2013), we tested two cancer types: hepatocellular carcinoma (HCC), and prostate adenocarcinoma (PAC). For HCC controls, we again used urine from patients diagnosed with liver cirrhosis and patients that were confirmed healthy. For PAC comparators, we tested urine from patients who were determined healthy from a PSA blood test (“Negative Screen”) and urine from patients who had high PSA levels but, when biopsied, were determined not to have cancer (“Healthy”). We utilized the same dilution series used by Hirotsu et al. (2015) 10^{-1} , 10^{-2} and 10^{-3} and, like this paper, found that a dilution of 10^{-2} elicited the greatest and most consistent attractive response of worms to patient HCC and PAC samples compared to those from benign disease and healthy samples (Fig. 9). Therefore, a dilution of 10^{-2} was used in all remaining experiments.

Hepatocellular Carcinoma (HCC) sample cohort

Individually, assays of HCC patient urine showed primarily positive CIs on average whereas those using liver cirrhosis and healthy patient urine exhibited negative CIs (Figs. 10 & 11). Our modified NSDT had a sensitivity of 67% and a specificity of 83% for distinguishing HCC from

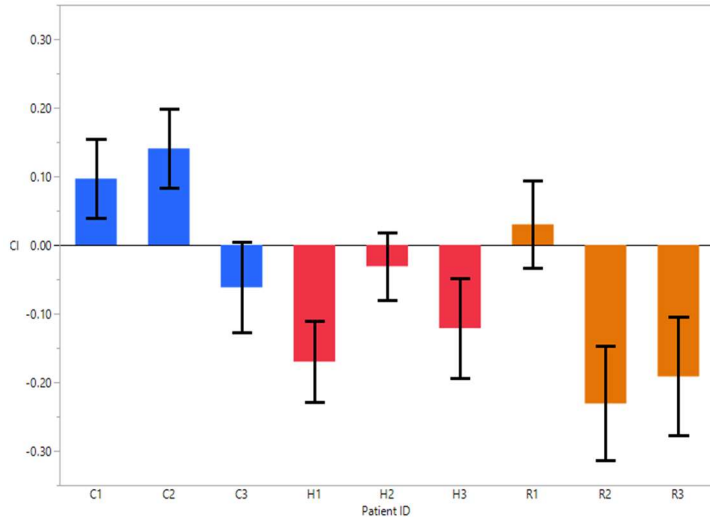


Figure 10: Wildtype N2 *C. elegans* are more attracted to urine from HCC patients than to urine of healthy and liver cirrhosis patients. Average CI for individual patient urine samples: HCC (blue), healthy (red) and liver cirrhosis (orange) (n = 6-10 assays each). Error bars represent 1 S.E.M.

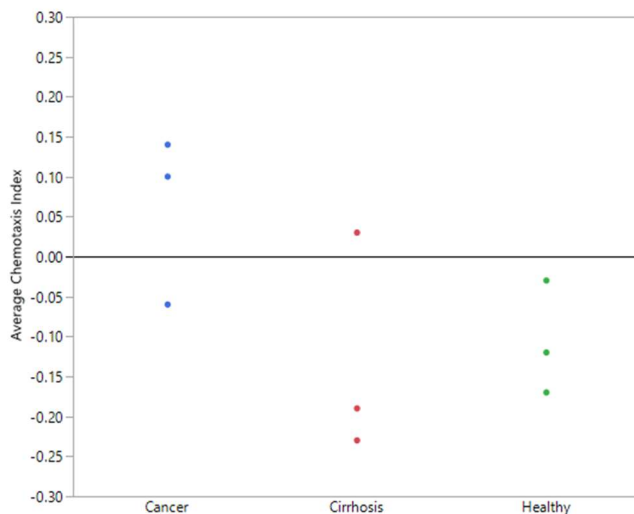


Figure 11: Wildtype N2 *C. elegans* are more attracted to urine from HCC patients than to urine from healthy and negative screen patients. A dot plot of average CIs of each patient urine sample (n=6-10 assays per patient).

both healthy urine and urine from patients with liver cirrhosis. Using a nested ANOVA testing the model $y = \mu + \text{type} + \text{individual}(\text{type}) + \varepsilon$, average CIs varied significantly among both sample types ($F_2 = 6.77$; $p = 0.002$) and among individuals within each type ($F_6 = 2.73$, $p = 0.021$), and a post hoc Tukey's HSD test ($\alpha =$

0.05) using least squares means revealed that HCC samples elicited significantly more positive chemotactic responses than either cirrhosis or healthy types.

Prostate adenocarcinoma (PAC) sample cohort

Since a dilution of 10^{-2} yielded the best results in preliminary assays (Fig. 9), we used this dilution for remaining PAC cohort assays. Considering all individuals in the full PAC cohort ($n = 67$), patients of different types (PAC, healthy, negative screen) exhibited no statistically significant differences from each other in terms of tobacco use, BMI, age, or prostate size (1-way ANOVA testing the model: $y = \mu + \text{type} + \varepsilon$); however, healthy

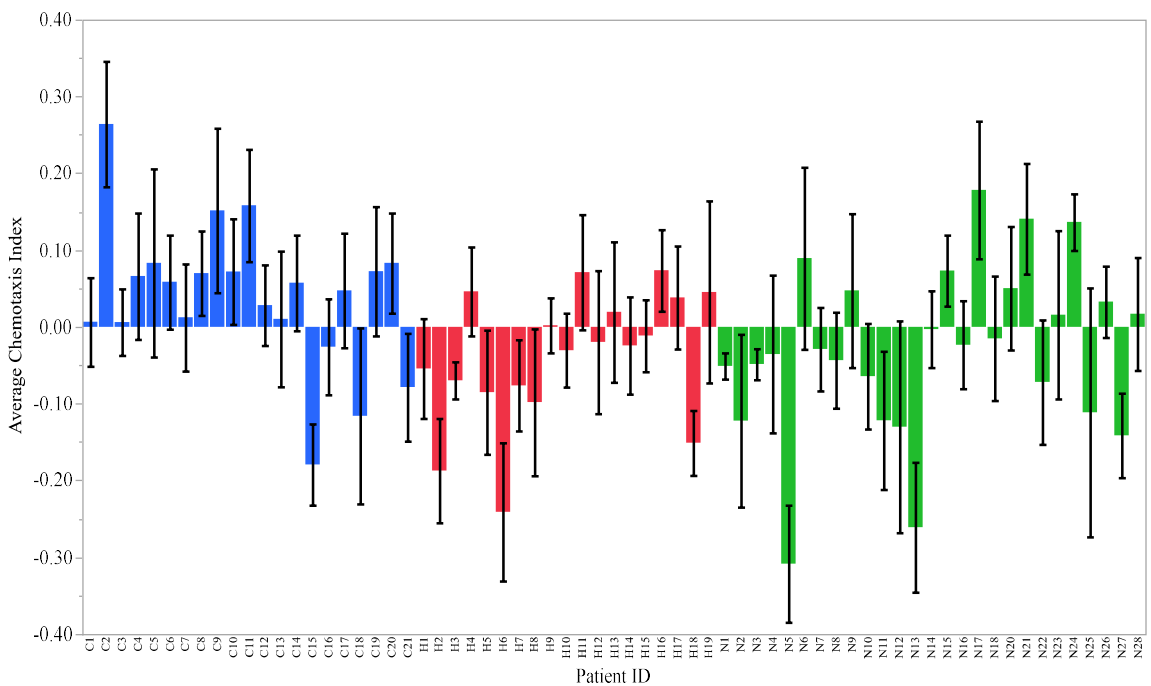


Figure 12: Wildtype N2 *C. elegans* are more attracted to urine from PAC patients than to healthy and negative screen patient urine. Average CI for each patient: PAC (blue), healthy (red) and negative screen (green) ($n = 4-24$ assays per individual). Error bars represent 1 S.E.M.

individuals tended to have slightly larger prostate sizes on average (55.1 cm^3 compared to

40.0 cm³ for PAC and 40.7 cm³ for negative screen individuals). As expected, significant differences among types existed for their defining features: 1) PSA level ($F_2 = 20.85$, $p < 0.0044$) such that negative screen individuals had substantially lower PSA levels (1.79 ng/mL on average) compared to PAC and healthy types (7.07 and 7.17 ng/mL on average, respectively) (Tukey's HSD, $\alpha = 0.05$); and 2) tumor type, with PAC having tumor types of 2 or 3; healthy with type 1; negative screen with types 0 or 1.

Independent of cancer stage, urine from PAC patients tended to elicit generally positive CIs compared to urine from healthy and negative screen individuals (Figs. 12 & 13). Using a nested ANOVA testing the model $y = \mu + \text{type} + \text{individual}(\text{type}) + \varepsilon$, there was significant variance among types ($F_2 = 6.22$, $p =$

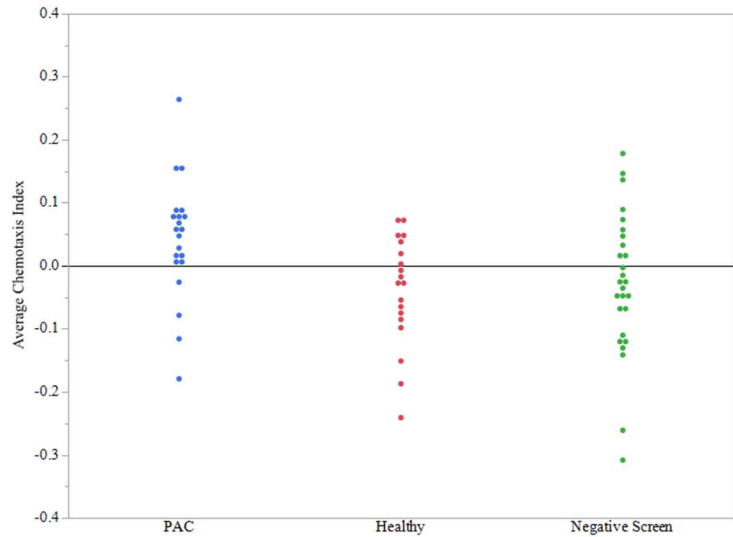


Figure 13: Wildtype N2 *C. elegans* are more attracted to urine from PAC patients than to healthy and negative screen patient urine. A dot plot of average CIs for each patient. (n=19-27 patients; 4-24 replicate assays per patient).

0.002) and among individuals within type ($F_{64} = 1.48$, $p = 0.014$). A post hoc Tukey's HSD test ($\alpha = 0.05$) using least-squares means showed that average CIs for PAC types were significantly higher than both that of healthy and negative screen types.

From the average CIs, a sensitivity of 81% and a specificity of 61% was calculated for detecting PAC from controls: healthy and negative screen urine samples. In comparison, using the PSA values from clinical diagnostics associated with each patient yielded a sensitivity and specificity of the PSA blood test as 88% and 59%, respectively.

Looking across “tumor type” categories (0-3), chemotactic indices were somewhat predictive of types 2 and 3, but completely unable to distinguish benign tumors (type 1) from healthy (type 0) (Fig. 14). A 1-way

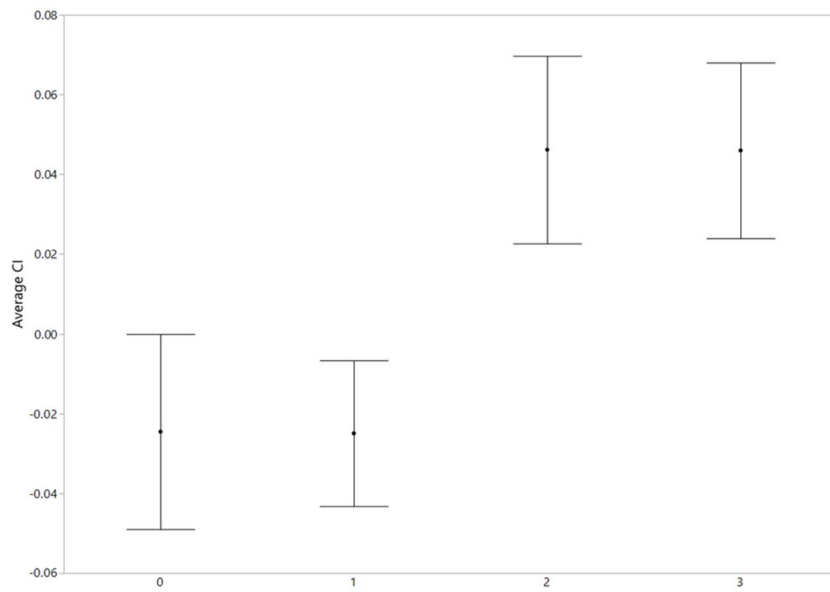


Figure 14: Wildtype N2 *C. elegans* are slightly more attracted to urine from PAC patients categorized as having tumor types 2 and 3 than to those with type 1 or no tumor (type 0). Average CIs (1 S.E.M.) for each patient type (n = 8-23 per type): 0 (-0.025, 0.020), 1 (-0.025, 0.020), 2 (0.046, 0.024), 3 (0.046, 0.035).

ANOVA testing the model $y = \mu + \text{type} + \varepsilon$ found only a marginally significant difference among tumor categories with respect to average CI ($F_3 = 2.75$, $p = 0.049$). A post hoc Tukey’s HSD comparison could not detect statistically significant differences between CIs of any pair of samples types.

Associations between CI values and patient clinical profile features

We assessed pairwise associations between average CIs and aspects of patient clinical profiles; i.e., PSA, tobacco use, body mass index (BMI), prostate size, age at time of urine collection, and tumor type (0 = none; benign = 1; Gleason score cancerous 3+3 and 3+4 = 2; Gleason score 4+3 and above=3) for the entire set and by each type (PAC, healthy or negative screen) (Appendices D-G). First considering the entire data set, statistically significant correlations included those between average CI and prostate size ($r = -0.307$, $p = 0.027$), average CI and tumor type ($r = 0.251$, $p = 0.041$), PSA value and prostate size ($r = 0.275$, $p = 0.040$), and PSA value and tumor type ($r = 0.602$, $p < 0.0001$) (Appendix D). Next assessing bivariate relationships within separate patient groups, there were no statistically significant correlations for PAC patients (Appendix E). However, there was a statistically significant correlation between PSA value and prostate size ($r = 0.5447$, $p = 0.0227$) for healthy individuals (Appendix F), and between PSA value and prostate size ($r = 0.5035$, $p = 0.0473$) for negative screen individuals (Appendix G).

Specific Aim 2. Strain- and species-specific chemotactic responses

To understand whether the results described above would be particular to the laboratory-adapted wildtype N2 *C. elegans* strain, we performed the same chemotaxis assays—again using PAC-cohort urine diluted at 10^{-2} —with an “ancestral” strain of N2 *C.*

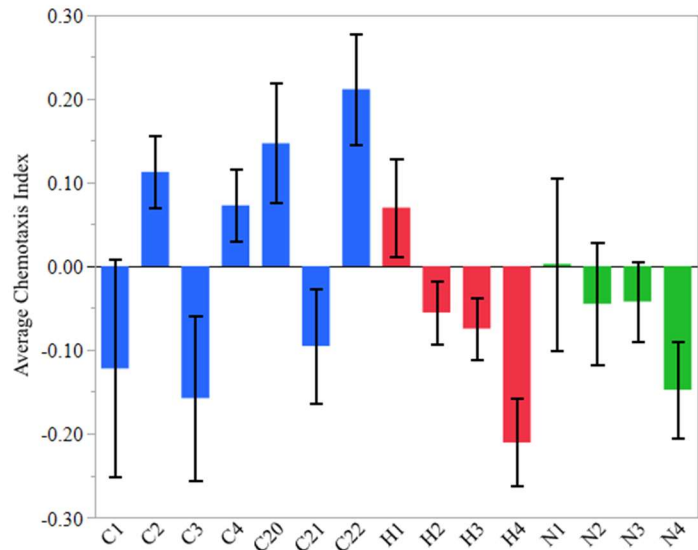


Figure 15: N2 Ancestral *C. elegans* are more attracted to urine from PAC patients than to healthy and negative screen patient urine. An average CI for each individual patient urine sample: PAC (blue), healthy (red) and negative screen (green) (n=6-10 assays). Error bars represent 1 S.E.M.

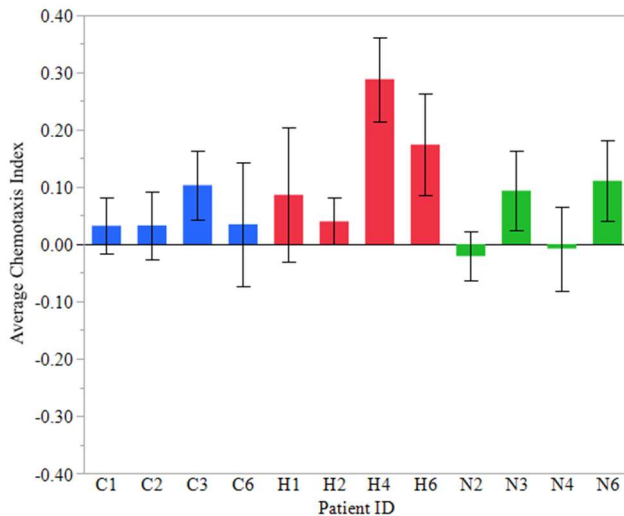


Figure 16: *C. briggsae* are attracted to urine from PAC, healthy and negative screen patients. Average CI for each patient urine sample: PAC (blue), healthy (red) and negative screen (green) (n = 6-10 assays each). Error bars represent 1 S.E.M.

elegans that should be less lab adapted, and with a congener, *C. briggsae*. Similar to the lab-adapted strain, ancestral *C. elegans* exhibited a preference for PAC urine over healthy and negative screen but were less able to distinguish between cancer and

control samples (Figs. 15 & 17).

Using a nested ANOVA testing the

model $y = \mu + \text{type} +$

individual(type) + ε , there was significant variance among types ($F_2 = 3.506$, $p = 0.0320$) and among individuals within type ($F_{12} = 2.286$, $p < 0.0001$). But a post hoc Tukey's HSD comparison could not detect statistically significant differences between CIs of any pair of samples types.

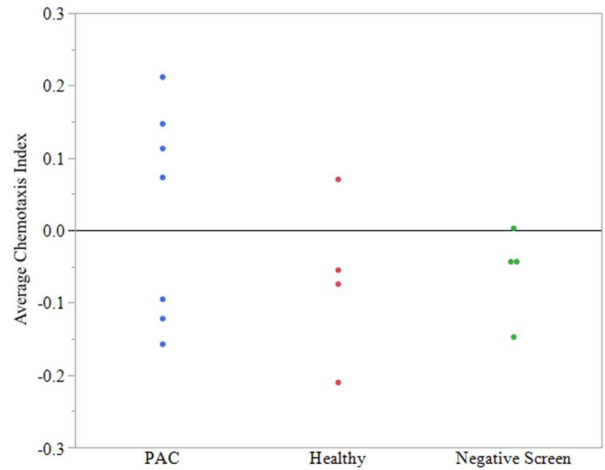


Figure 17: N2 Ancestral *C. elegans* are attracted to urine from PAC patients. A dot plot of the average CI of each individual: PAC (blue), healthy (red) and negative screen (green) ($n = 4$ patients; 6-10 assays each).

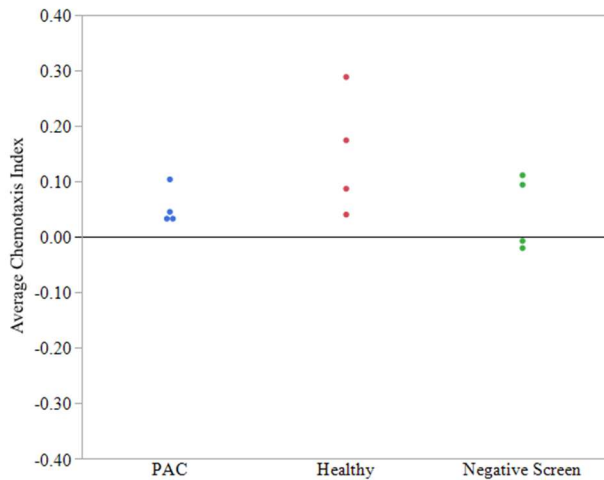


Figure 18: *C. briggsae* are attracted to urine independent of type (PAC, healthy and negative screen). A dot plot of the average CI of each individual: PAC (blue), healthy (red) and negative screen (green) ($n = 4$ patients; 6-10 assays each).

C. briggsae had a different response to the urine from PAC, healthy and negative screen patients compared to that observed for *C. elegans*. Generally, *C. briggsae* were determined to have an overall

attractive response to urine independent of type (PAC, healthy and negative screen) (Figs. 16 & 18). Using the same

nested ANOVA model $y = \mu + \text{type} +$

individual(type) + ε , there was no significant variation in CIs among types ($F_2 = 1.578$, $p = 0.212$) or among individuals within type ($F_9 = 0.917$, $p = 0.514$).

Specific Aim 3. Identifying candidate VOC ligands

I conducted a literature review of 19 metabolomic studies of various types of cancer. From these studies, several compounds were detected at elevated concentrations in the metabolomic profiles of cancer patients compared to those of healthy individuals (Appendix B). These studies encompassed many types and stages of cancer, as well as tissue types, with the metabolomic profiles being derived from fluids including serum, exhaled breath, cell culture and urine (Appendix B). Several the studies identified benzaldehyde and decane based compounds as having significantly different concentrations between cancer and healthy patients (Appendix B).

Decane was found by seven studies to be significantly differentially expressed within the VOC profiles of individuals with multiple types of cancer (oral, lung, prostate, breast, and renal) compared to healthy patients (Appendices B & C). However, decane appeared in multiple forms across these studies, including the isomers: 3,7-dimethylundecane, dodecane, hexadecane, nonadecane, tetradecane and undecane (Appendix C). All of these compounds were found to be overrepresented in all cancer and sample types tested (Bouza et al., 2017; Thriumani et al., 2018; Handa et al., 2014; Wang et al., 2012; Hadi et al., 2017; Wang et al., 2016).

Among the above compounds, 3,7-dimethylundecane was found to be overrepresented in the exhaled breath of patients with oral squamous cell carcinoma along with dodecane, hexadecane, and undecane (Bouza et al., 2017). Dodecane was found at elevated concentrations in the breath and bronchoscopic tissue of adenocarcinoma lung cancer

patients with a specific epidermal growth factor receptor (EGFR) mutation (Handa et al., 2014), and in the exhaled breath of breast cancer patients (Phillips et al., 2010). Wang et al. (2012) found nonadecane to be present in exhaled breath, lung cell culture and lung tissue from patients with either lung adenocarcinoma, small cell lung, or lung squamous carcinoma cancer. Finally, tetradecane concentration was found to be significantly elevated in the serum of pre-operative breast cancer patients (Hadi et al., 2017) as well as in urine from patients with renal cell carcinoma (Wang et al., 2016).

Additionally, various aldehydes, a group previously identified as a potential biomarker for cancer (Rodrigues et al., 2018; Fuchs et al., 2010; Poli et al., 2010) were identified across multiple studies including benzaldehyde, decanal, hexadecanal, nonanal, and pentanal (Appendix C). Within the first group, 4-methylbenzaldehyde was present at significantly reduced concentrations in prostate cancer cell lines compared to healthy prostate epithelial cell lines (Lima et al., 2018), while five other studies found benzaldehyde concentrations to be significantly elevated in cancer samples (bladder, lung, oral) compared to healthy samples regardless of sample type (urine, serum, cell culture, breath) (Rodrigues et al., 2018; Jia et al., 2018; Roś-Mazurczyk et al., 2017; Bouza et al., 2017). Jia et al. (2018) found that benzaldehyde concentrations were elevated in lung large cell carcinoma compared to both lung adenocarcinoma and lung squamous carcinoma cell culture lines; however, benzaldehyde was generally elevated in lung cancer patients, regardless of type, compared to healthy patients. Similarly, Roś-Mazurczyk et al. (2017) determined that benzaldehyde was significantly overexpressed in

serum from patients with both lung adenocarcinoma and squamous cell carcinoma. In a study of exhaled breath from patients with oral squamous cell carcinoma, Bouza et al. (2017) determined that benzaldehyde concentrations were significantly increased compared to the exhaled breath from healthy patients. Rodrigues et al. (2018) found that benzaldehyde was also significantly overexpressed among three bladder cancer cell lines compared to healthy bladder cells.

Finally, several ketones were identified as potential differentiators of cancer from healthy VOC profiles, including: 2-octanone; 3-octanone; 2,6-di-*t*-butyl-4-hydroxymethylene-2,3,5,6-tetrahydrocyclohexanone; cyclohexanone; 2-butanone; 3-heptanone; 4-heptanone; and 2-nonanone (Appendix C). Of these compounds 2,6-di-*t*-butyl-4-hydroxymethylene-2,3,5,6-tetrahydrocyclohexanone; 2-butanone; 3-heptanone; and 2-nonanone were elevated in prostate (urine), lung (exhaled breath), renal (urine), and bladder cancer (cell culture), while 3-octanone was elevated in lung cancer (urine) but reduced in prostate cancer (urine) compared to healthy (Appendices B & C). Last, 2-octanone, and cyclohexanone were both reduced in prostate cancer (urine) while 4-heptanone was reduced in renal cancer (urine) (Appendices B & C).

Discussion

There is a demand for a cost-efficient and non-invasive cancer detection system that is effective across cancer types and stages. Although there are a few early detection tests such as PSA for prostate cancer, non-invasive tests for other cancer types do not exist. Rather, diagnostic methods for most cancer types are both invasive and rarely begun until a patient has reached later stages of disease progression (Hirotsu et al., 2015; Thompson et al., 2004). It has been well established through metabolomics studies that VOC profiles of patients with cancer are different from those of healthy patients (Bouza et al., 2017; Hadi et al., 2017; Wang et al., 2012). However, these studies have so far failed to identify particular VOCs that consistently and reliably diagnose cancer. Taking advantage of the advanced olfactory capabilities of certain organisms may provide the foundation for non-invasive early detection tests (Sato et al., 2017; Cronu et al., 2011).

Recent research has attempted to exploit the olfactory systems of both mice (Sato et al., 2017) and dogs (Cronu et al., 2011) for cancer detection; however, both animals require behavioral training to be effectively utilized for this purpose. Conversely, *C. elegans* relies almost exclusively on its primal olfactory system to engage with and navigate its environment. This system is highly attuned for detecting VOCs, as its primary functions include detection of bacterial food sources and pheromones produced by potential mates (Bargmann et al., 1993; Bargmann, 2006). As such, *C. elegans* have shown promise for detecting cancer patients from healthy individuals with no requirement for behavioral training (Hirotsu et al., 2015).

Failure to validate Hirotsu et al. (2015) NSDT

Despite considerable effort, I failed to replicate the results of Hirotsu et al. (2015). There are several potential reasons for this. First, Hirotsu et al. (2015) controlled for a variety of variables among their patient cohort, while our cohort was more diverse. For example, they excluded participants based upon various physical symptoms (i.e., headache, bloody feces, alcohol consumption, appetite, chest or abdominal distention). Although our PAC patient cohort was more diverse in terms of clinical variables, it was nonetheless restricted to men between the ages of 47 and 75 residing near Portland, Oregon who had not received chemotherapy at time of urine collection (Appendix A). Second, we were unable to obtain the specific wildtype N2 *C. elegans* line used by Hirotsu et al. (2015); it is therefore possible that minor but important genetic differences existed between our wildtype N2 *C. elegans* and that of Hirotsu et al. (2015) that explain the inconsistency in our results. Third, Hirotsu et al. (2015) studied exclusively Japanese individuals, while our cohort is likely to have consisted primarily of white Americans. If differences between typical eastern and western diet can lead to differences in the VOC profiles of patient urine, this could also explain *C. elegans*' reduced and less consistent attraction to cancer patient urine in our study.

Wildtype N2 C. elegans are attracted to liver (HCC) and prostate (PAC) cancer patient urine

The research reported by this thesis extended the work of Hirotsu et al. (2015), showing that wildtype N2 *C. elegans* could, with reasonable accuracy, distinguish urine derived from liver (HCC) and prostate (PAC) cancer patients from that derived from healthy individuals and those with benign disease (Figs. 10-13). For both types of cancer, a dilution of 10^{-2} was found to have the strongest attractive behavioral response from the *C. elegans* compared to the other dilutions tested (Fig. 9). At this dilution, *C. elegans* were also able to best distinguish cancer patient urine from that of the other patient types, to which they were repulsed on average (Figs. 10-13). For the PAC cohort, we found that *C. elegans* were attracted to “tumor type” scores of 2 and 3, which represented cancerous tissue with Gleason scores of 3+3 and above, but were repulsed on average by samples from individuals scored as 0 and 1, corresponding to normal and benign tissue (Fig. 14). We reason that the attractive behavioral response, resulting in higher average CI values, is a result of a differing VOC profiles between normal, benign and cancerous tissue types.

Despite the moderate success of our NSDT for distinguishing cancer from other sample types, we found large variability in CIs both within and among urine samples regardless of patient type. Variation in CI within individuals (i.e., among technical replicates) might be attributed to factors including the innate variability of animal behavior, pipetting error, and micro-environmental variations. Variation in CI among individuals could result from the same factors in addition to those including patient diet, lifestyle, whether or not they

had benign tumors, and the initial concentration of the collected urine sample. However, Hirotsu et al. (2015) found no relationship between nematode chemotactic responses and creatine concentration, a proxy for initial urine concentration. We also noticed that the attractive behavior of the *C. elegans* to cancer patient urine appeared to depend upon the number of times a urine sample was freeze-thawed prior to use in the chemotaxis assays, with the attractive response to cancer patient urine declining with increasing freeze-thaw cycles (data not shown). This might be as a result of VOCs being lost from the urine samples during such cycles.

We determined the NSDT test had a sensitivity of 81% and a specificity of 61% for the PAC cohort. While these values are much lower than those yielded by the NSDT of Hirotsu et al. (2015)—sensitivity of 95.8% and specificity of 95.0%—our NSDT appears to be comparable to the PSA test, which yielded a sensitivity and specificity of 88% and 59%, respectively, for our PAC cohort. Unlike the PSA test, however, the NSDT test can be used for other types of cancer. Although the sensitivity of NSDT was lower for HCC compared to PAC cohorts, the limited sample size for the former and lack of clinical diagnostics information on these patients leaves us unable to expand on the reasons behind this outcome.

Although we found no differences among patient types in the PAC cohort in tobacco use, BMI, or age, healthy individuals in this cohort tended to have enlarged prostates, 55.1 cm³ on average, compared to the widely accepted average size for normal prostates, 45 cm³ (Appendix A) (Mitterberger et al., 2010). Specifically, 17 of the 20 PAC patients for

which we received prostate volume information had prostates smaller than 45 cm³ (Appendix A). Conversely, PAC and negative screen patients had smaller than average prostate sizes of 40.0 cm³ and 40.7 cm³, respectively. The underlying cause of the above relationship is unknown, but may be spurious (a result of sampling from a small population) or related to the fact that patients in the “healthy” group were those who presented with elevated PSA, levels of which are known to be positively correlated with prostate size (e.g., Coric et al., 2015). Likewise, we found a positive correlation between prostate size and PSA value for both patients categorized as healthy and negative screen (Appendix F & G). The same relationship did not exist for PAC patients, who also had elevated PSA compared to negative screen patients (Appendices A & E). However, we found a significant negative correlation between average CI and prostate size among the entire data set of patients, meaning that *C. elegans* was increasingly attracted to urine derived from patients with smaller prostate sizes (Appendix D). Curiously, prostate size appears to be negatively related to PAC incidence (Al-khalil et al, 2016) with PAC more often invading smaller prostatic glands compared to larger ones (Chen et al., 1999). Though the PSA test remains a valuable indicator of prostate cancer, these complex relationships alongside the fact that benign prostate tumors (Punglia et al., 2006) and many other factors can result in elevated PSA highlight its shortcomings for the reliable prediction of PAC.

Nematode strain- and species-specific chemotactic responses

We further extended the work of Hirotsu et al. (2015) to test whether the efficacy of the NSDT would demonstrate *Caenorhabditis* species- or strain-specificity using a subset of the PAC cohort of urine samples. The “ancestral” N2 *C. elegans* strain performed in a manner that was qualitatively similar to the standard wildtype N2 laboratory strain that has experienced many generations of laboratory adaptation, but the lab-adapted wildtype strain was better at distinguishing PAC from healthy and negative screen urine (Figs. 12, 13, 15 & 17). The causal nature of this difference is unknown but is consistent with the presence of within-species genetic variation for chemotactic response to cancer-specific VOCs. Notably, smaller pilot chemotaxis assays were performed with two *C. elegans* natural isolates, JU2524 and MY16, neither of which exhibited any response during NSDTs. In particular, these nematodes showed very little to no movement on average from the point of origin on chemotaxis plates (data not shown). Chemotaxis responses were also species-specific in that *C. briggsae* demonstrated no ability to differentiate between PAC, healthy and negative screen samples, and indeed, demonstrated a weak attraction to all three over the water control (Figs. 16 & 18). *C. briggsae* is quite distantly related to *C. elegans*; the two species are as genetically dissimilar as human and mouse (Stein et al., 2003; Waterston et al., 2002). These findings imply the presence of both within- and among-species genetic variation for cancer-specific chemotactic response, suggesting fruitful avenues for future work to identify relevant genetic loci using existing recombinant inbred or nearly isogenic lines (RILs and NILs, respectively).

Identifying candidate VOC ligands

It is now well understood that the human body excretes numerous VOCs in breath, blood, skin and urine as a result of an individual's specific metabolism (Shirasu & Touhara, 2011). The profile of such VOCs can be altered when an individual has an infection or a metabolic, genetic or other disease (Shirasu & Touhara, 2011). By reviewing the recent cancer metabolomics literature, I was able to identify VOCs that both differentiate cancer patients from healthy individuals and that may also overlap those with biological relevance for *C. elegans* based on their natural history and bacterivorous nature (Appendix B). Specifically, three functional groups: alkanes, aldehydes, and ketones were the most relevant compound types in that they were identified by multiple studies (Appendix C). Within these functional groups, there were ten relevant primary compounds that arose in multiple studies including: benzaldehyde, butanone, decane, decanal, heptanone, hexanone, nonanal, nonanone, octanone, and pentanal which had the most significant differences in concentration between healthy and cancer patients (Appendix C). Of these, two compound types stood out: decane, an alkane, was identified in 7/19 studies addressing five different cancer types (breast, lung, oral prostate and renal), and benzaldehyde, the same aldehyde utilized as a positive control in this study, was identified by 5/19 studies on four cancer types (bladder, lung, prostate and oral) (Appendices B & C). Additionally, benzaldehyde and various compounds containing decane (undecane; dodecane; 3,7-dimethylundecane; and hexadecane) appeared in the exhaled breath of patients with oral cancer (Appendix B).

Cellular concentrations of alkanes (dodecane, tetradecane and hexadecane) and benzaldehyde are known to be elevated in cancer cells as a result of lipid peroxidation, which occurs with inflammation and oxidative stress and is associated with higher concentrations of reactive oxygen species (Rodrigues et al., 2018; Phillips et al., 2000; Silva et al., 2011; Poli et al., 2010). Therefore, elevated concentrations of benzaldehyde and alkanes signify higher oxidative stress within cancer cells (Rodrigues et al., 2018). As a group, aldehydes are often generated as intermediates during cytotoxic processes as a result of signal transduction, gene regulation and cell proliferation (Bouza et al., 2017). Finally, ketones have also been identified as potential VOC biomarkers of cancer. Matsumura et al. (2010) found that ketones were more prevalent in the urine of mice with induced lung cancer tumors than healthy control mice, suggesting that the ketogenesis pathway might be useful for identifying lung cancer tumors. Another literature review of 52 lung cancer VOC profile studies by Saalberg and Wolff (2016) found that 2-butanone was the most prevalent VOC with a significant difference in concentration across these papers. Ketones typically function within cells as fuel for mitochondria. Ketone bodies are typically produced by mitochondria within the liver cells as a result of fasting or exercise and transported via the bloodstream to the brain and other organs for conversion into acetyl-CoA (Henderson, 2008; Laffel, 1999; Martinez-Outschoorn et al., 2012). Within the brain, however, ketone bodies are also generated by astrocytes for mitochondrial metabolism of neurons in a process known as “neuron-glia metabolic coupling” (Guzmán, 2001; Martinez-Outschoorn et al., 2012). It has been shown that

tumors exhibit a similar metabolic process, generating ketone bodies that fuel oxidative phosphorylation critical for anabolic tumor growth (Martinez-Outschoorn et al., 2012).

Bacteria also emit specific VOC profiles composed of metabolic byproducts released during growth, in addition to VOCs that provide protection against predation and are utilized for intercellular communication (Chen et al., 2017). Likely because bacteria are the primary food source of *C. elegans*, this species can detect hundreds of VOCs including ketones, aldehydes, esters, alcohols and various aromatic compounds (Liao et al., 2010). Specifically, *E. coli*, one of *C. elegans*' preferred food sources in the lab, emits volatile compounds like 2-heptanone, 2-nonanone, 2-undecanone, 1-octanol, 1-dodecanol and indole among many others (Chen et al., 2017). Generally, methyl ketones produced through the decarboxylation of fatty acids during pyruvate fermentation through anaerobic conditions (Audrain et al., 2015; Ryu et al., 2003). Other microbes emit short-chain alkanes, such as decane or tetradecane, while longer-chain hydrocarbons are found in cyanobacteria since they are capable of synthesizing branched hydrocarbons (Audrain et al., 2015; Tellez, Schrader & Kobaisy, 2001; Ladygina, Dedyukhina & Vainshtein, 2006). The formation of alkanes in microbes requires intermediate aldehydes which are converted into various alkanes via aldehyde deformylating oxygenase (Rodriguez & Atsuml, 2014). In *E. coli* specifically, high levels of endogenous aldehyde reductases generally convert aldehydes to alcohols (Rodriguez & Atsuml, 2014). This could be the reason *C. elegans* are highly attracted to benzaldehyde (Bargmann et al., 1993).

It is likely that N2 *C. elegans* ' preference for urine prostate and liver cancer (Figs. 10-13) is owed to VOC profiles of cancer patients mimicking those of bacterial or other food microbial food sources. If true, too little is known about the differences in natural history among *C. elegans* isolates and among nematode species to understand why such an attractive response would not be universal. In any case, future research could determine which of these candidate compounds is attractive to wildtype N2 *C. elegans*, and whether any of them distinguish the HCC or PAC urine sample VOC profiles from those of control samples.

Conclusions & Future Research

This research intended to expand upon the NSDT developed by Hirotsu et al. (2015) to include a previously untested cancer type (HCC), control for further variables (random nematode movement, light, surrounding odors), determine whether nematode response to cancer was strain- or species-specific, and identify candidate cancer-specific VOCs that may serve as attractive cues for *C. elegans*. We were unable to validate the findings of Hirotsu et al. (2015) using the chemotaxis assay methodology described (Fig. 7); however, results improved upon applying a modified version of Margie et al.'s (2013) methodology (Fig. 8). We showed that standard wildtype N2 *C. elegans* and, to a lesser extent, an ancestral strain of N2 *C. elegans*, were attracted to urine from patients with either HCC or PAC compared to urine from healthy patients and those diagnosed with a benign disease (Figs. 10-13). Additionally, our modified NSDT produced a similar sensitivity and specificity for PAC (81% and 61%, respectively) compared to PSA tests of the same samples (88% and 59% respectively). Beyond showing promise for early detection of multiple cancer types, our test was also the first demonstrate that *C. elegans* can distinguish samples derived from cancer patients from those of individuals with benign disease. Further research is needed to determine if cancer-specific VOC profiles are a result of the tumor itself or from the immunological response to the cancerous tumor. Furthermore, since the Warburg effect is known to occur both in cancerous and non-cancerous cell types during rapid cell proliferation (Abdel-Haleem et al., 2017), future work should also include testing samples from patients with sepsis and other

inflammatory diseases as a control. Indeed, when individuals with benign tumors are removed from the PAC cohort dataset, the specificity of our NSDT increases slightly to 65%, while sensitivity remains unchanged at 81%. This could signify that *C. elegans* are attracted to compounds associated with general inflammation in addition to those specific to cancer.

We found strain- and species-specificity in chemotactic responses that point to natural genetic variation in underlying olfactory receptor genes. Future work to perform NSDTs on existing RIL and NIL nematode libraries, or systematic RNAi knockdown of olfactory genes, could uncover the source of this variation. Such work, combined with chemical interrogation of urine samples, could ultimately lead to identifying VOC ligands and their associated neurological receptor proteins. This result also suggests that artificial selection could yield *C. elegans* strains with superior ability to distinguish cancer-derived from healthy samples. Further work would be needed to determine if within- and between-species variation in response is maintained across other cancer types and stages.

Despite the promise of our results, the large variability in chemotactic responses both among technical replicates and individuals within sample type (cancer, control) suggests that development of an animal-based assay using non-modified wildtype nematodes is unlikely to be worthwhile. Parallel experiments conducted by OHSU collaborators aimed at determining which *C. elegans* GPCRs are activated in the presence of cancer patient urine have been unsuccessful owing to the difficulty of enriching for membrane bound proteins, and GPCRs specifically, in *C. elegans*. Rather, future GC-MS analysis of

fractionated HCC or PAC urine could identify candidate VOCs that distinguish cancer from healthy samples, which could then be validated using *C. elegans* bioassays. Future work may also include testing *C. elegans*' chemotactic responses to the biologically relevant concentrations of candidate compounds identified by this study as another means of identifying the key ligand(s) that animals use to distinguish cancer from healthy and benign tissues.

References

- Abdel-Haleem, A. M., Lewis, N. E., Jamshidi, N., Mineta, K., Gao, X. and Gojobori, T. (2017) The Emerging Facets of Non-Cancerous Warburg Effect. *Front Endocrinol*, 8(279).
- Al-Khail, S., Ibilbor, C., Cammack, J. T. & de Reise, W. (2016) Association of prostate volume with incidence and aggressiveness of prostate cancer. *Research and Reports in Urology*, 8, 201-208.
- Audrain, B., Farag, M. A., Ryu, C-M. & Ghigo, J-M. (2015) Role of bacterial volatile compounds in bacterial biology. *FEMS Microbiology Reviews*, 39, 222-233.
- Barash, O., Peled, N., Hirsch, F. R. & Haick, H. (2009) Sniffing the unique “odor print” of non-small-cell lung cancer with gold nanoparticles. *Small*, 5(22), 2618–2624.
- Bargmann, C. I. Chemosensation in *C. elegans*. WormBook (2006).
- Bargmann, C. I., Hartwig, E. & Horvitz, H. R. (1993) Odorant-selective genes and neurons mediate olfaction in *C. elegans*. *Cell*, 74(3), 515–527.
- Barnum K.J., O’Connell M.J. (2014) Cell Cycle Regulation by Checkpoints. In: Noguchi E., Gadaleta M. (eds) Cell Cycle Control. Methods in Molecular Biology (Methods and Protocols), vol 1170. Humana Press, New York, NY
- Barry MJ. Clinical practice. Prostate-specific-antigen testing for early diagnosis of prostate cancer. *N. Engl. J. Med.*, 344(18), 1373-1377.
- Bertram, J. (2000). The Molecular Biology of Cancer. *Molecular Aspects of Medicine*, 21(6), 167–223.

Borza, Y.Y., Mochalski, P., Ruzsanyi, V., Amann, A. & Haick, H. (2015) Hybrid volatolomics and disease detection. *Angew Chem Int Ed Engl*, 54(38), 11036-48.

Bouza, M., Gonzalez-Soto, J., Pereiro, R., de Vicente, J.C. & Sanz-Medel, A. (2017) Exhaled breath and oral cavity VOCs as potential biomarkers in oral cancer patients. *J. Breath Res.* 016015.

Boyle, P. & Levin, B. (2008) World cancer report 2008. *International Agency for Research on Cancer*, 511.

Chen, J., Tang, J., Shi, H., Tang, C. & Zhang, R. (2017) Characteristics of volatile organic compounds produced from five pathogenic bacteria by headspace-solid phase micro-extraction/gas chromatography-mass spectrometry. *J Basic Microbiol*, 57, 228-237.

Chen, M. E., Troncoso, P., Johnston, D., Tang, K. & Babaian, R. J. (1999) Prostate Cancer Detection: Relationship To Prostate Size. *Urology*, 53, 764-768.

Coric, J., Mujic, J., Kucukalic, E. & Ler, D. (2015) Prostate-Specific Antigen (PSA) and Prostate Volume: Better Predictor of Prostate Cancer for Bosnian and Herzegovina Men. *The Open Biochemistry Journal*, 9, 34-36.

Cronu, J.N., Cancel-Tassin, G., Ondet, V., Girardet, C. & Cussenot, O. (2010) Olfactory detection of prostate cancer by dogs sniffing urine: a step forward in early diagnosis. *Eur Urol*, 59(2), 197-201.

Dang, C. V. (2017). Feeding frenzy for cancer cells. *Science*. 358(6365), 862-863.

Eastham, J. (2017) Prostate cancer screening. *Investig Clin Urol*, 58, 217-219.

Fages, A., Duarte-Salles, T., Stepien, M., Ferrari, P., ..., Jenab, M. (2015) Metabolomic profiles of hepatocellular carcinoma in European prospective cohort. *BMC Med.*, 13(242).

Friday, P., Schiman, J. D., Fisher, P. G., & Gibbs, P. (2015). Early Detection of Cancer: Past, Present, and Future. *ASCO Educational Book*, 57-65.

Fuchs, P., Loeseken, C., Schubert, J. K. & Miekisch, W. (2009) Breath gas aldehydes as biomarkers of lung cancer. *Int. J. Cancer*, 126, 2663-2670.

Fukuto, H. S., Ferkey, D.M., Apicella, A.J., Lans, H., Sharmeen, T., Chen, W., Lefkowitz, R.J., Jansen, G., Schafer, W.R., & Hart, A.C. (2004). G Protein-Coupled Receptor Kinase Function Is Essential for Chemosensation in *C. elegans*. *Neuron*, 42(4), 581-593.

Gao, Q., Su, X., Annabi, M. H., Schreiter, B. R., Prince, T., Ackerman, A., Morgas, S., Williams, H. & Lee, W. Y. (2019) Application of Urinary Volatile Organic Compounds (VOCs) for the Diagnosis of Prostate Cancer. *Clin Genitourin Cancer*, 17(3) 183-190.

Gleason, D.F. & Mellinger, G.T. (1974) Prediction of prognosis for prostatic adenocarcinoma by combined histological grading and clinical staging. *J Urol.*, 111 58–64.

Gordetsky, J. & Epstein, J. (2016) Grading of prostatic adenocarcinoma: current state and prognostic implications. *Diagnostic Pathology*, 11(25).

Guzmán, M. & Blázquez, C. (2001) Is there an astrocyte-neuron ketone body shuttle? *Trends Endocrinol Metab*, 12, 169-7.

Hadi, N. I., Jamal, Q., Iqbal, A., Shaikh, F., Somroo, S. & Musharraf, S. G. (2017) Serum Metabolomic Profiles for Breast Cancer Diagnosis, Grading and Staging by Gas Chromatography-Mass Spectrometry. *Scientific Reports*, 7(1715).

Handa, H., Usuba, A., Maddula, S., Baumbach, J. I., Mineshita, M. & Miyazawa, T. (2014) Exhaled Breath Analysis for Lung Cancer Detection Using Ion Mobility Spectrometry. *PLOS One*, 9(12).

Heiden, M. G. Vander, & Deberardinis, R. J. (2017). Understanding the intersections between metabolism and cancer biology. *Cell*, 168(4), 657-669.

Henderson, S. T. (2008) Ketone bodies as a therapeutic for Alzheimer 's disease. *Neurotherapeutics*, 5, 470-80.

Hirotsu, T., Sonoda, H., Uozumi, T., Shinden, Y., Mimori, K., Maehara, Y., & Hamakawa, M. (2015). A Highly Accurate Inclusive Cancer Screening Test Using *Caenorhabditis elegans* Scent Detection. *PLOS One*, 10(3), e0118699.

Huang, J., Weinstein, S. J., Moore, S. C., Derkach, A., Hua, X., Mondul, A. M., Sampson, J. N. & Albanes, D. (2019) Pre-diagnostic Serum Metabolomic Profiling of Prostate Cancer Survival. *J Gerontol A Biol Sci Med Sci*, 74(6), 853-859.

Itoi, T., Sugimoto, M., Umeda, J., Sofuni, A., Tsuchiya, T., Tsuji, S., Tanaka, R., Tonozuka, R., Honjo, M., Moriyasu, F., Kasuya, K., Nagakawa, Y., Abe, Y., Takano, K., Kawachi, S., Shimazu, M., Soga, T., Tomita, M. & Sunamura, M. (2017) Serum Metabolomic Profiles for Human Pancreatic Cancer Discrimination. *Int. J. Mol. Sci.*, 18(767).

Jia, Z., Zhang, H., Ong, C. N., Patra, A., Lu, Y., Lim, C. T. & Venkatesan, T. (2018) Detection of Lung Cancer: Concomitant Volatile Organic Compounds and Metabolomic Profiling of Six Cancer Cell Lines of Different Histological Origins. *ACS Omega*, 3, 5131-5140.

Khalid, T., Aggio, R., White, P., Costello, B., Persad, R., Al-Kateb, H., Jones, P., Probert, C. S. & Ratcliffe, N. (2015) Urinary Volatile Organic Compounds for the Detection of Prostate Cancer. *PLOS One*, 10(11), e0143283.

Kiontke, K. and Sudhaus, W. Ecology of *Caenorhabditis* species (January 09, 2006), WormBook, ed. The *C. elegans* Research Community, WormBook, doi/10.1895/wormbook.1.37.1, <http://www.wormbook.org>.

Ladygina, N., Dedyukhina, E. G. & Vainshtein, M. B. (2006) A review on mimicrobial synthesis of hydrocarbons. *Process Biochem*, 41, 1001-14.

Laffel, L. (1999) Ketone bodies: a review of physiology, pathophysiology and application of monitoring to diabetes. *Diabetes Metab Res Rev*, 15, 412-26.

Lee, J.H., Kim Y.G., Kim, M., Kim, E., Choi, H., Kim, Y., & Lee, J. (2017). Indole-associated predator-prey interactions between nematode *Caenorhabditis elegans* and bacteria. *Environ Microbiol*, 19(5), 1776-1790.

Liao, C., Gock, A., Michie, M., Morton, B., Anderson, A. & Trowell, S. (2010) Behavioural and Genetic Evidence for *C. elegans*' Ability to Detect Volatile Chemicals Associated with Explosives. *PLOS One*, 5(9), e12615.

Lima, A. R., Araújo, A. M., D., Pinto, J., Jerónimo, C., Henrique, R., de Lourdes Bastos, M., Carvalho, M. & de Pinho, P. G. (2018) Discrimination between the human prostate normal and cancer cell exometabolome by GC-MS. *Scientific Reports*, 8(5539).

Margie, O., Palmer, C. & Sang-Chin, I. (2013) *C. elegans* Chemotaxis Assay. *J. Vis. Exp.* (74), e50069.

Markar, S. R., Brodie, B., Chin, S.T., Romano, A., Spalding, D., & Hanna, G.B. (2018) Profile of exhaled-breath volatile organic compounds to diagnose pancreatic cancer. *Br. J. Surg.* 105(11), 1493–1500.

Martinez-Outschoorn, U. E., Lin, Z., Whitaker-Menezes, D., Howell, A., Sotgia, F. & Lisanti, M. P. (2012) Ketone body utilization drives tumor growth and metastasis. *Cell Cycle*, 11(21), 3964-71.

Matsumura, K., Opiekun, M., Oka, H., Vachani, A., Albelda, S. M., Yamazaki, K., Beauchamp, G. K. (2010) Urinary Volatile Compounds as Biomarkers for Lung Cancer: A Proof of Principle Study Using Odor Signatures in mouse Models of Lung Cancer. *PLOS One*, 5(1), e8819.

Mistry, K. & Cable, G. (2003) Meta-analysis of prostate-specific antigen and digital rectal examination as screening tests for prostate carcinoma. *JABFP*, 16(2), 95-101.

Mitterberger, M., Horninger, W., Aigner, F., Pinggera, G. M., Stepphan, I., Rehder, P. & Frauscher, R. (2010) Ultrasound of the prostate. *Cancer Imaging*, 10, 40-48.

Narang, A., & Nicholas, L. H. (2017). Out of Pocket Spending and Financial Burden Among Medicare Beneficiaries with Cancer. *JAMA Oncology*, 3(6), 757–765.

Ossoliński, K., Nizioł, J., Arendowski, A., Ossolińska, A., Ossoliński, T., Kucharz, J., Wiechno, P. & Ruman, T. (2019) Mass spectrometry-based metabolomic profiling of prostate cancer- a pilot study. *J Cancer Metastasis Treat*, 5(1).

Parikh, R., Mathai, A., Parikh, S., Chandra Seekhar, G. & Thomas, R. (2008) Understanding and using sensitivity, specificity and predictive values. *Indian Journal of Ophthalmology*, 56 (1), 45.

Phillips, M., Cataneo, R. N., Greenberg, J., Gunawardena, R., Naidu, A., & Rahbari-Oskoui, F. (2000). Effect of age on the breath methylated alkane contour, a display of apparent new markers of oxidative stress. *The Journal of Laboratory and Clinical Medicine*, 136(3), 243–249

Phillips, M., Cataneo, R. N., Saunders, C., Hope, P., Schmitt, P. & Wai, J. (2010) Volatiles biomarkers in the breath of women with breast cancer. *J. Breath Res*, 4(026003).

Pierce-Shinmomura, J.T., Morse, T. M. & Lockery, S.R. (1999) The Fundamental Roles of Pirouettes in *Caenorhabditis elegans* Chemotaxis. *J. Neurosci.*, 19(21), 9557-9569.

Poli, D., Goldoni, M., Corradi, M., Acampa, O., Carbognani, P., Internullo, E., Casalini, A. & Mutti, A. (2010) Determination of aldehydes in exhaled breath of patients with lung

cancer by means of on-fiber-derivation SPME-GC/MS. *J. Chromatography B*, 878 (27), 2643-2651.

Punglia, R. S., D'Amico, A. V., Catalona, W. J., Roehl, K. A. & Kuntz, K. M. (2006) Impact of Age, Benign Prostatic Hyperplasia, and Cancer on Prostate-Specific Antigen Level. *Cancer*, 106(7), 1507-13.

Rodrigues, D., Pinto, J., Araújo, A. M., Monteiro-Reis, S., Jerónimo, C., Henrique, R., de Lourdes Bastos, M., de Pinho, P. G. & Carvalho, M. (2018) Volatile metabolomic signature of bladder cancer cell lines based on gas chromatography-mass spectrometry. *Metabolomics*, 14(62).

Rodriguez, G. M. & Atsuml, S. (2014) Toward aldehyde and alkane production by removing aldehyde reductase activity in *Escherichia coli*. *Metab Eng*, 25, 227-237.

Roś-Mazurczyk, M., Wojakowska, A., Marczak, L., Polański, K., Pietrowska, M., Polanska, J., Dziadziuszko, R., Jassem, J., Rzyman, W. & Widlak, P. (2017) Panel of serum metabolites discriminates cancer patients and healthy participants of lung cancer screening—a pilot study. *Acta Biochimica Polonica*, 64.

Ryu, C-M., Farag, M. A., Hu, C-H, Reddy, M. S., Wei, H-X. Paré, P. W. & Kloepper, J. W. (2003) Bacterial volatiles promote growth in *Arabidopsis*. *PNAS*, 100(8), 4927-4932.

Saalberg, Y. & Wolff, M. (2016) VOC breath biomarkers in lung cancer. *Clinica Chimica Acta*, 459, 5-9.

Santos, P. M., Sánchez, M. N., Pozas, A. P. C., Pavón, J. L. P., Cordero, B. M. (2017) Determination of ketones and ethyl acetate—a preliminary study for the discrimination of patients with lung cancer. *Anal Bioanal Chem*, 409, 5689-5696.

Sarkar, S., Horn, G., Moulton, K., Oza, A., Byler, S., Kokolus, S., & Longacre, M. (2013). Cancer Development, Progression, and Therapy: An Epigenetic Overview. *Int. J. Mol. Sci*, 14(10), 21087-21113.

Sato, T. Katsuoka, Y., Yoneda, K., Nonomura, M., Uchimoto, S., Kobayakawa, R., Kobayakawa, K. & Mizutani, Y. (2017) Sniffer mice discriminate urine odours of patients with bladder cancer: A proof-of-principle study for non-invasive diagnosis of cancer-induced odours. *Scientific Reports*, 7(1).

Schiffman, J.D., Fisher, P.G. & Gibbs, P. (2015) Early detection of cancer: past, present, and future. *ASCO*, 57-65.

Schulenburg, H. & Félix, M-A. (2017) The Natural Biotic Environment of *Caenorhabditis elegans*. *Genetics*, 206, 55-86.

Sengupta, P., Chou, J. H. & Bargmann, C. I. (1996) odr-10 Encodes a Seven Transmembrane Domain Olfactory Receptor Required for Responses to the Odorant Diacetyl. *Cell*, 84(6), 875–887.

Shirasu, M. & Touhara, K. (2011) The scent of disease: volatiles organic compounds of the human body related to disease and disorder. *J Biochem*, 150(3), 257-266.

Siegel, R. L., Miller, K. D., & Jemal, A. (2019). Cancer statistics, 2018. *CA: A Cancer Journal for Clinicians*, 69(1), 7–34.

Silva, M., Gallego, M. & Silva, M. (2011) Analysis of endogenous aldehydes in human urine by static headspace gas chromatography-mass spectrometry. *Journal of Chromatography A*, 1437, 241-246.

Stein, L. D., Bao, Z., ..., & Waterston, R. H. (2003) The Genome Sequence of *Caenorhabditis briggsae*: A Platform for Comparative Genomics. *PLOS Biology*, 1(2), 166-192.

Tellez, M. R., Schrader, K. K. & Kobaisy, M. (2001) Volatile components of the cyanobacterium *Oscillatoria perornata*. *J Agr Food Chem*, 49, 5989-92.

Thompson IM, ..., Coltman Jr., C.A. (2004) Prevalence of prostate cancer among men with a prostate-specific antigen level ≤ 4.0 ng per milliliter. *N Engl J Med*, 350(22):2239-2246.

Thrumani, R., Zakaria, A., Hashim, Y. Z. H-Y., Jeffree, A. I., Helmy, K. M., Kamarudin, L. M., Omar, M. I., Shakaff, A. Y. M., Adom, A. H. & Persaud, K. (2018) A study on volatile organic compounds emitted by in-vitro lung cancer cultured cells using gas sensor array and SPME-GCMS. *BMC Cancer*, 18(362).

Tisch, U., Billan, S., Ilouze, M., Phillips, M., Peled, N. & Haick, H. (2012) Volatile Organic Compounds in Exhaled Breath as Biomarkers for the Early Detection and Screening of Lung Cancer. *CML-Lung Cancer*, 5(4), 107-17.

Tobin, A. B. (2008) G-protein-coupled receptor phosphorylation: where, when and by whom. *Br. J. Pharmacol.* 153, 167–176.

Vazquez, A., Kamphorst, J. J., Markert, E. K., Schug, Z. T., Tardito, S., & Gottlieb, E. (2016). Cancer metabolism at a glance. *Journal of Cell Science*. 129, 3367-3373.

Vidal, B., Aghayeva, U., Sun, H., Wang, C., Glenwinkel, L., Bayer, E., & Hobert, O. (2018) An atlas of *Caenorhabditis elegans* chemoreceptor expression. *PLoS Biol*, 16(1).

Wang, D., Wang, C., Pi, X., Guo, L., Wang, Y., Li, M., Feng, Y., Lin, Z., Hou, W. & Li, E. (2016) Urinary volatile organic compounds as potential biomarkers for renal cell carcinoma. *Biomedical Reports*, 5, 68-72.

Wang, Y., Hu, Y., Wang, D., Yu, K., Wang, L., Zou, Y., Zhao, C., Zhang, X., Wang, P. & Ying, K. (2012) The analysis of volatile organic compounds biomarkers for lung cancer in exhaled breath, tissues and cell lines. *Cancer Biomarkers*, 11, 129-137.

Wang, Y., Jacobs, E. J., Carter, B. D., Gapstur, S. M. & Stevens, V. L. (2019) Plasma Metabolomic Profiles and Risk of Advanced and Fatal Prostate Cancer. *Eur Urol Oncol*, 246.

Warburg, O., Wind, F. & Negelein, E. (1927) The metabolism of tumors in the body. *J Gen Physiol*, 8 (6) 519-30.

Waterston, R. H., Lindblad-Toh, K., ..., & Lander, E. S. (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature*, 420, 520-561.

Zhang, C., Zhao, N., Chen, Y., Zhang, D., Yan, J., Zou, W., Zhang, K. & Huang, X. (2016) The Signaling Pathway of *Caenorhabditis elegans* Mediates Chemotaxis Response to the Attractant 2-Heptanone in a Trojan Horse-like Pathogenesis. *J Biol Chem*, 291(45), 23618-23627.

Appendix A. Table of prostate urine sample cohort patient information.

Type	Patient ID	Average CI	PSA (ng/ml)	Clinical Diagnosis	Tumor Grade	Prostate Volume (cm3)	Tobacco Use	Body Mass Index (kg/m2)	Age at Time of Collection
PAC	1	0.08	4.13	Adenocarcinoma	Gleason Score: 3+4	50	Current Smoker	31.07	61
Negative Screen	2	-0.31	1.94	Normal		Unknown	Former Smoker	25.58	68
PAC	3	0.03	3.95	Adenocarcinoma	Gleason Score: 3+4	42	Never Smoker	31.25	74
Healthy	4	-0.09	5.70	Benign prostatic hyperplasia		44	Never Smoker	28.19	55
Healthy	5	-0.01	2.58	Benign prostatic hyperplasia		53	Former Smoker	24.75	66

Type	Patient ID	Average CI	PSA (ng/ml)	Clinical Diagnosis	Tumor Grade	Prostate Volume (cm3)	Tobacco Use	Body Mass Index (kg/m2)	Age at Time of Collection
Healthy	6	-0.24	14.21	Benign		90	Never Smoker	24.84	53
Healthy	7	-0.04	5.27	Benign		43	Never Smoker	24.10	65
PAC	8	0.06	5.78	Adenocarcinoma	Gleason Score: 3+3	61	Never Smoker	24.83	68
PAC	9	0.08	7.37	Adenocarcinoma	Gleason Score: 4+3	16	Former Smoker	24.03	66
PAC	10	0.07	6.67	Adenocarcinoma	Gleason Score: 4+3	35	Former Smoker	29.14	63
Negative Screen	11	0.07	0.70	Normal		35	Former Smoker	25.18	70

Type	Patient ID	Average CI	PSA (ng/ml)	Clinical Diagnosis	Tumor Grade	Prostate Volume (cm3)	Tobacco Use	Body Mass Index (kg/m2)	Age at Time of Collection
Negative Screen	12	0.09	3.63	Normal		40	Never Smoker	32.84	67
Healthy	13	-0.10	8.17	Benign prostatic hyperplasia		90	Never Smoker	33.55	66
PAC	14	0.15	4.52	Adenocarcinoma	Gleason Score: 3+4	41	Former Smoker	30.94	67
Negative Screen	15	-0.05	0.96	Normal		15	Never Smoker	24.33	64
PAC	16	0.07	12.66	Adenocarcinoma	Gleason Score: 4+3	18	Current Smoker	26.54	58
Negative Screen	17	-0.04	2.01	Normal		30	Never Smoker	33.96	65

Type	Patient ID	Average CI	PSA (ng/ml)	Clinical Diagnosis	Tumor Grade	Prostate Volume (cm3)	Tobacco Use	Body Mass Index (kg/m2)	Age at Time of Collection
Negative Screen	18	0.05	6.97	Benign prostatic hyperplasia		60	Former Smoker	29.53	58
Negative Screen	19	0.02	3.93	Normal		48	Never Smoker	27.48	69
PAC	20	0.08	5.88	Adenocarcinoma	Gleason Score: 4+5	44	Never Smoker	28.65	62
Negative Screen	21	-0.06	4.77	Normal		85	Never Smoker	26.76	64
Negative Screen	22	-0.12	1.05	Normal		40	Former Smoker	24.85	63
Negative Screen	23	-0.13		Benign prostatic hyperplasia		Unknown	undocumented	26.31	68

Type	Patient ID	Average CI	PSA (ng/ml)	Clinical Diagnosis	Tumor Grade	Prostate Volume (cm3)	Tobacco Use	Body Mass Index (kg/m2)	Age at Time of Collection
PAC	24	0.16	16.49	Adenocarcinoma	Gleason Score: 3+4	30	Never Smoker	28.55	70
Negative Screen	25	-0.26	0.57	Normal		30	Former Smoker	37.10	70
Healthy	26	-0.03	3.55	Benign prostatic hyperplasia		31	Never Smoker	31.57	73
Negative Screen	27	0.00	0.33	Normal		Unknown	Current Smoker	24.56	59
Negative Screen	28	-0.02	1.94	Normal		Unknown	Never Smoker	24.81	69
Negative Screen	29	0.18		Normal		15	Never Smoker	31.18	60

Type	Patient ID	Average CI	PSA (ng/ml)	Clinical Diagnosis	Tumor Grade	Prostate Volume (cm3)	Tobacco Use	Body Mass Index (kg/m2)	Age at Time of Collection
Negative Screen	30	-0.02	1.02	Normal		Unknown	Never Smoker	38.97	66
Healthy	31	0.05	8.16	Benign		97	Never Smoker	28.82	63
Healthy	32	0.07	1.64	Benign prostatic hyperplasia		25	Never Smoker	28.96	57
Negative Screen	33	-0.08	4.70	Adenocarcinoma	Gleason Score: 4+3	25	Former Smoker	29.11	66
Negative Screen	34	0.05	0.65	Normal		Unknown	Never Smoker	25.29	68
PAC	35	0.01	5.80	Adenocarcinoma	Gleason Score: 3+4	49	Never Smoker	31.68	64

Type	Patient ID	Average CI	PSA (ng/ml)	Clinical Diagnosis	Tumor Grade	Prostate Volume (cm3)	Tobacco Use	Body Mass Index (kg/m2)	Age at Time of Collection
Healthy	36	0.14	6.27	Benign		40	Never Smoker	36.68	62
Negative Screen	37	0.14	1.09	Normal		Unknown	Never Smoker	32.26	54
Negative Screen	38	-0.07		Normal		30	Never Smoker	26.76	55
PAC	39	0.15	5.50	Adenocarcinoma	Gleason Score: 3+3	25	Never Smoker	31.74	66
Negative Screen	40	0.02	0.75	Benign prostatic hyperplasia		40	Light Tobacco Smoker	28.00	71
Negative Screen	41	0.14	0.85	Normal		Unknown	Former Smoker	25.39	63

Type	Patient ID	Average CI	PSA (ng/ml)	Clinical Diagnosis	Tumor Grade	Prostate Volume (cm3)	Tobacco Use	Body Mass Index (kg/m2)	Age at Time of Collection
Healthy	42	0.02	13.45	Benign prostatic hyperplasia		52	Former Smoker	23.06	75
Negative Screen	43	-0.14		Benign prostatic hyperplasia		50	Never Smoker	36.00	62
Negative Screen	44	-0.11	1.77	Normal		Unknown	Former Smoker	20.00	59
PAC	45	-0.18	11.51	Adenocarcinoma	Gleason Score: 3+3	112	Former Smoker	30.02	63
PAC	46	-0.02	14.95	Adenocarcinoma	Gleason Score: 4+4	29	Never Smoker	29.51	69
Negative Screen	47	0.03	1.05	Normal		40	Never Smoker	28.00	74

Type	Patient ID	Average CI	PSA (ng/ml)	Clinical Diagnosis	Tumor Grade	Prostate Volume (cm3)	Tobacco Use	Body Mass Index (kg/m2)	Age at Time of Collection
Healthy	48	0.02	9.66	Prostatic intraepithelial neoplasia	Generic Grade: High Grade	50	Never Smoker	32.56	75
PAC	49	-0.12	5.14	Adenocarcinoma	Gleason Score: 3+3	37	Never Smoker	27.83	66
Healthy	50	-0.01	3.07	Benign		30	Never Smoker	23.85	62
PAC	51	0.05	7.30	Adenocarcinoma	Gleason Score: 4+4	Unknown	Never Smoker	24.87	57
PAC	52	0.07	5.92	Adenocarcinoma	Gleason Score: 3+3	61	Former Smoker	21.95	67
Healthy	53	0.07	4.90	Benign		Unknown	Former Smoker	38.20	47

Type	Patient ID	Average CI	PSA (ng/ml)	Clinical Diagnosis	Tumor Grade	Prostate Volume (cm3)	Tobacco Use	Body Mass Index (kg/m2)	Age at Time of Collection
Healthy	54	0.04	10.50	Benign		50	Never Smoker	24.96	68
Healthy	55	-0.15	4.20	Benign		Unknown	Former Smoker	28.76	61
Healthy	391	0.03	8.60	Benign		56	Former Smoker	23.88	56
Negative Screen	393	-0.08		Normal		67	Never Smoker	27.57	67
Healthy	415	-0.05	8.02	Benign		62	Never Smoker	29.11	62
Negative Screen	418	-0.12	0.17	Normal		72	Light Tobacco Smoker	27.48	72

Type	Patient ID	Average CI	PSA (ng/ml)	Clinical Diagnosis	Tumor Grade	Prostate Volume (cm3)	Tobacco Use	Body Mass Index (kg/m2)	Age at Time of Collection
Healthy	428	-0.05	10.05	Benign		63	Former Smoker	29.86	63
PAC	432	0.01	7.09	Adenocarcinoma	Gleason score 3+4	61	Never Smoker	26.61	61
Negative Screen	686	-0.04	1.66	Normal		72	Never Smoker	23.87	72
PAC	689	0.13	7.15	Adenocarcinoma	Gleason score 3+3	68	Current Chew User	23.22	68
PAC	692	0.04	5.32	Adenocarcinoma	Gleason score 3+4	60	Never Smoker	34.36	60
PAC	822	0.11	12.13	Adenocarcinoma	Gleason score 4+4	73	Never Smoker	23.06	73

Type	Patient ID	Average CI	PSA (ng/ml)	Clinical Diagnosis	Tumor Grade	Prostate Volume (cm3)	Tobacco Use	Body Mass Index (kg/m2)	Age at Time of Collection
Negative Screen	830	0.04	1.65	Normal		70	Never Smoker	28.96	70
Healthy	969	-0.03	8.32	Benign		61	Never Smoker	25.83	61

Appendix B. Table of metabolomic studies on VOC profiles from both cancer and healthy patients

Title	Year Published	Total Cancer Samples	Total Control Samples	Sample Type(s)	Cancer Types	VOCs
Tisch, U., Billan, S., Ilouze, M., Phillips, M., Peled, N. & Haick, H. (2012) Volatile Organic Compounds in Exhaled Breath as Biomarkers for the Early Detection and Screening of Lung Cancer. <i>CML-Lung Cancer</i> , 5(4), 107-17.	2012	65	31	Breath	Lung	2-Butanone Diacetyl 3-Butyn-2-ol Tetramethyl-urea Butane, 2-methyl 1,4-Pentadiene
Bouza, M., Gonzalez-Soto, J., Pereiro, R., de Vicente, J.C. & Sanz-Medel, A. (2017) Exhaled breath and oral cavity VOCs as potential biomarkers in oral cancer patients. <i>J. Breath Res.</i> 016015.	2017	52	52	Breath	Oral	Undecane dodecane benzaldehyde decanal 3,7-dimethyl undecane 4,5-dimethyl nonane 1-octene hexadecane

Title	Year Published	Total Cancer Samples	Total Control Samples	Sample Type(s)	Cancer Types	VOCs
Handa, H., Usuba, A., Maddula, S., Baumbach, J. I., Mineshita, M. & Miyazawa, T. (2014) Exhaled Breath Analysis for Lung Cancer Detection Using Ion Mobility Spectrometry. <i>PLOS One</i> , 9(12).	2014	50	39	Breath Tissue	Lung	n-dodecane
Wang, Y., Hu, Y., Wang, D., Yu, K., Wang, L., Zou, Y., Zhao, C., Zhang, X., Wang, P. & Ying, K. (2012) The analysis of volatile organic compounds biomarkers for lung cancer in exhaled breath, tissues and cell lines. <i>Cancer Biomarkers</i> , 11, 129-137.	2012	85	158	Breath Cell Culture Tissue	Lung	Pentadecanone nonadecane eicosane

Title	Year Published	Total Cancer Samples	Total Control Samples	Sample Type(s)	Cancer Types	VOCs
Lima, A. R., Araújo, A. M., D., Pinto, J., Jerónimo, C., Henrique, R., de Lourdes Bastos, M., Carvalho, M. & de Pinho, P. G. (2018) Discrimination between the human prostate normal and cancer cell exometabolome by GC-MS. <i>Scientific Reports</i> , 8(5539).	2018	4	1	Cell Culture	Prostate	<p>pentadecane-2-one decanoic acid</p> <p>cyclohexanone</p> <p>4-methylheptan-2-one 2-methylpentane-1,3-diol</p> <p>4-methylbenzaldehyde 1-(3,5-dimethylfuran-2-yl) ethenone</p> <p>methyl benzoate</p> <p>nonanoic acid</p>
Thrumani, R., Zakaria, A., Hashim, Y. Z. H-Y., Jeffree, A. I., Helmy, K. M., Kamarudin, L. M., Omar, M. I., Shakaff, A. Y. M., Adom, A. H. & Persaud, K. (2018) A study on volatile organic compounds emitted by in-vitro lung cancer cultured cells using gas sensor array and SPME-GCMS. <i>BMC Cancer</i> , 18(362).	2018	2	2	Cell Culture	Lung	<p>Dodecane</p> <p>Heneicosane</p> <p>Decanal</p> <p>Nonanal</p> <p>2-ethyldodecanol</p>

Title	Year Published	Total Cancer Samples	Total Control Samples	Sample Type(s)	Cancer Types	VOCs
Rodrigues, D., Pinto, J., Araújo, A. M., Monteiro-Reis, S., Jerónimo, C., Henrique, R., de Lourdes Bastos, M., de Pinho, P. G. & Carvalho, M. (2018) Volatile metabolomic signature of bladder cancer cell lines based on gas chromatography-mass spectrometry. <i>Metabolomics</i> , 14(62).	2018	3	1	Cell Culture	Bladder	Benzaldehyde 2-nonanone
Jia, Z., Zhang, H., Ong, C. N., Patra, A., Lu, Y., Lim, C. T. & Venkatesan, T. (2018) Detection of Lung Cancer: Concomitant Volatile Organic Compounds and Metabolomic Profiling of Six Cancer Cell Lines of Different Histological Origins. <i>ACS Omega</i> , 3, 5131-5140.	2018	7	1	Cell Culture	Lung	2-ethyl-1-hexanol Benzaldehyde 2,4-Decadien-1-ol

Title	Year Published	Total Cancer Samples	Total Control Samples	Sample Type(s)	Cancer Types	VOCs
Hadi, N. I., Jamal, Q., Iqbal, A., Shaikh, F., Somroo, S. & Musharraf, S. G. (2017) Serum Metabolomic Profiles for Breast Cancer Diagnosis, Grading and Staging by Gas Chromatography-Mass Spectrometry. <i>Scientific Reports</i> , 7(1715).	2017	152	155	Serum	Breast	Glucopyranoside Tetradecane 1-(1-methoxycarbonyl-ethyl)- 4-(2-methyl-2-hydroxypropyl) benzene
Fages, A., Duarte-Salles, T., Stepien, M., Ferrari, P., ..., Jenab, M. (2015) Metabolomic profiles of hepatocellular carcinoma in European prospective cohort. <i>BMC Med.</i> , 13(242).	2015	114	222	Serum	Liver	Ethanol Propylene Glycol Acetate Tyrosine Phenylalanine Glutamate

Title	Year Published	Total Cancer Samples	Total Control Samples	Sample Type(s)	Cancer Types	VOCs
Roś-Mazurczyk, M., Wojakowska, A., Marczak, L., Polański, K., Pietrowska, M., Polanska, J., Dziadziuszko, R., Jassem, J., Rzyman, W. & Widlak, P. (2017) Panel of serum metabolites discriminates cancer patients and healthy participants of lung cancer screening-a pilot study. <i>Acta Biochimica Polonica</i> , 64.	2017	31	92	Serum	Lung	Benzaldehyde Hydroxypyruvic Acid Urea

Title	Year Published	Total Cancer Samples	Total Control Samples	Sample Type(s)	Cancer Types	VOCs
Itoi, T., Sugimoto, M., Umeda, J., Sofuni, A., Tsuchiya, T., Tsuji, S., Tanaka, R., Tono-zuka, R., Honjo, M., Moriyasu, F., Kasuya, K., Nagakawa, Y., Abe, Y., Takano, K., Kawachi, S., Shimazu, M., Soga, T., Tomita, M. & Sunamura, M. (2017) Serum Metabolomic Profiles for Human Pancreatic Cancer Discrimination. <i>Int. J. Mol. Sci.</i> , 18(767).	2017	39	46	Serum	Panc-reatic	Glutamatic Acid Histidine Glutamine Trimethylamine N-oxide Serine
Wang, Y., Jacobs, E. J., Carter, B. D., Gapstur, S. M. & Stevens, V. L. (2019) Plasma Metabolomic Profiles and Risk of Advanced and Fatal Prostate Cancer. <i>Eur Urol Oncol</i> , 246.	2019	246	342	Serum	Prostate	Ethylmalonate Butyrylcarnitine Aspartate Sphingomyelin γ-glutamyl-methionine γ-glutamyl-glutamine

Title	Year Published	Total Cancer Samples	Total Control Samples	Sample Type(s)	Cancer Types	VOCs
Huang, J., Weinstein, S. J., Moore, S. C., Derkach, A., Hua, X., Mondul, A. M., Sampson, J. N. & Albanes, D. (2019) Pre-diagnostic Serum Metabolomic Profiling of Prostate Cancer Survival. <i>J Gerontol. A. Biol Sci Med Sci</i> , 74(6), 853-859.	2018	197	0	Serum	Prostate	N-Oleoyl taurine
Ossoliński, K., Nizioł, J., Arendowski, A., Ossolińska, A., Ossoliński, T., Kucharz, J., Wiechno, P. & Ruman, T. (2019) Mass spectrometry-based metabolomic profiling of prostate cancer- a pilot study. <i>J Cancer Metastasis Treat</i> , 5(1).	2019	5	15	Serum Urine Interstitial Fluid	Prostate	Cholesterol Sulfate (Urine) Triglyceride (Urine) Cholesterol and triglycerides (Serum) Amino Acids (Serum)

Title	Year Published	Total Cancer Samples	Total Control Samples	Sample Type(s)	Cancer Types	VOCs
Wang, D., Wang, C., Pi, X., Guo, L., Wang, Y., Li, M., Feng, Y., Lin, Z., Hou, W. & Li, E. (2016) Urinary volatile organic compounds as potential biomarkers for renal cell carcinoma. <i>Biomedical Reports</i> , 5, 68-72.	2016	22	25	Urine	Renal	Phenol decanal 1,6-dioxacyclo-dodecane-7,12-dione 1-bromo-1-(3-methyl-1-pentenylidene)-2,2,3,3-tetramethyl-cyclopropane nonanal 3-ethyl-3-methylheptane isolongifolene-5-ol 2,5-cyclohexadiene-1,4-dione, 2,6-bis (1,1-dimethylethyl) tetradecane aniline 2,6,10,14-tetramethyl-pentadecane styrene 4-heptanone dimethyl silanediol

Title	Year Published	Total Cancer Samples	Total Control Samples	Sample Type(s)	Cancer Types	VOCs
Santos, P. M., Sánchez, M. N., Pozas, A. P. C., Pavón, J. L. P., Cordero, B. M. (2017) Determination of ketones and ethyl acetate—a preliminary study for the discrimination of patients with lung cancer. <i>Anal Bioanal Chem</i> , 409, 5689-5696.	2017	12	12	Urine	Lung	ethyl acetate 3-heptanone 3-octanone
Khalid, T., Aggio, R., White, P., Costello, B., Persad, R., Al-Kateb, H., Jones, P., Probert, C. S. & Ratcliffe, N. (2015) Urinary Volatile Organic Compounds for the Detection of Prostate Cancer. <i>PLOS One</i> , 10(11), e0143283.	2015	59	43	Urine	Prostate	2,6-dimethyl-7-octen-2-ol pentanal (Valeraldehyde) 3-octanone 2-octanone

Title	Year Published	Total Cancer Samples	Total Control Samples	Sample Type(s)	Cancer Types	VOCs
Gao, Q., Su, X., Annabi, M. H., Schreiter, B. R., Prince, T., Ackerman, A., Morgas, S., Williams, H. & Lee, W. Y. (2019) Application of Urinary Volatile Organic Compounds (VOCs) for the Diagnosis of Prostate Cancer. <i>Clin Genitourin Cancer</i> , 17(3) 183-190.	2019	53	22	Urine	Prostate	1,1,1,5,5,5-hexamethyl-3,3-bis [(trimethylsilyl) oxy]- Trisiloxane 1-Propylpentachloro-triphosphazene 4-Nitro-4'-Chlorodiphenyl-sulfoxide 1,1,3,3,5,5,7,7,9,9-decamethyl-pentasiloxane 2,6-di-t-butyl-4-hydroxymethylene-2,3,5,6-tetrahydro cyclohexanone Phthalic acid, bis(7-methyloctyl) ester

Appendix C. Table of specific VOCs from the metabolomic studies that were significantly different between cancer and healthy patients

Primary compound	Additional compounds containing primary compound	Functional Group	Number of Papers Referenced (Out of 19 Papers)	Number of Cancer types appeared (Out of 8 Cancer types)
Decane	3,7-dimethylundecane dodecane hexadecane nonadecane tetradecane undecane	Alkane	7	5
Benzaldehyde	4-methylbenzaldehyde	Aldehyde	5	4
Decanal	Hexadecanal	Aldehyde	2	2
Nonanal		Aldehyde	2	2
Octanone	2-octanone 3-octanone	Ketone	2	2
Hexanone	2,6-di-t-butyl-4-hydroxymethylene-2,3,5,6-detetrahydrocyclohexanone Cyclohexanone	Ketone	2	1
Heptanone	3-heptanone 4-heptanone	Ketone	2	2
Butanone	2-butanone	Ketone	1	1
Nonanone	2-nonanone	Ketone	1	1
Pentanal		Aldehyde	1	1

**Appendix D. Multivariate correlation of entire data set with three sample types
(Healthy, Negative Screen and PAC)**

	Average CI	PSA	Prostate Size	Tobacco Use	Tumor Type	BMI	Age
Average CI	1.000 0.6314	0.0329 0.6314	-0.3066 0.0273	-0.0465 0.7193	0.2505 0.0409	-0.0408 0.7428	0.0101 0.9352
PSA	0.0329 0.6314	1.000	0.2747 0.0403	-0.0635 0.4516	0.6016 <0.0001	-0.0805 0.5473	-0.0207 0.6676
Prostate Size	-0.3066 0.0273	0.2747 0.0403	1.000	0.0391 0.9977	-0.0654 0.5295	-0.0167 0.9470	-0.0478 0.5214
Tobacco Use	-0.0465 0.7193	-0.0635 0.4516	0.0391 0.9977	1.000	0.0835 0.5012	-0.1911 0.1226	-0.0682 0.5764
Tumor Type	0.2505 0.0409	0.6016 <0.0001	-0.0654 0.5295	0.0835 0.5012	1.000	-0.0325 0.7937	-0.0323 0.7955
BMI	-0.0408 0.7428	-0.0805 0.5473	-0.0167 0.9470	-0.1911 0.1226	-0.0325 0.7937	1.000	-0.1076 0.3862
Age	0.0101 0.9352	-0.0207 0.6676	-0.0478 0.5214	-0.0682 0.5764	-0.0323 0.7955	-0.1076 0.3862	1.000

Appendix E. Multivariate correlation of PAC patient data

	Average CI	PSA	Prostate Size	Tobacco Use	Tumor Type	BMI	Age
Average CI	1.0000 0.9562	0.0128 0.9562	-0.3034 0.1929	0.1087 0.6391	-0.081 0.7271	-0.2688 0.2387	0.1510 0.5136
PSA	0.0128 0.9562	1.0000	-0.0781 0.9314	-0.1059 0.6479	0.2948 0.1946	-0.2050 0.3728	0.1476 0.5231
Prostate Size	-0.3034 0.1929	-0.0781 0.9314	1.0000	0.2294 0.5325	-0.5021 0.0629	0.1518 0.9308	0.0205 0.9944
Tobacco Use	0.1087 0.6391	-0.1059 0.6479	0.2294 0.5325	1.0000	0.1132 0.6251	-0.1859 0.4197	-0.1694 0.4629
Tumor Type	-0.081 0.7271	0.2948 0.1946	-0.5021 0.0629	0.1132 0.6251	1.0000	-0.2803 0.2184	-0.1684 0.4657
BMI	-0.2688 0.2387	-0.2050 0.3728	0.1518 0.9308	-0.1859 0.4197	-0.2803 0.2184	1.0000	-0.1519 0.5110
Age	0.1510 0.5136	0.1476 0.5231	0.0205 0.9944	-0.1694 0.4629	-0.1684 0.4657	-0.1519 0.5110	1.0000

Appendix F. Multivariate correlation of healthy patient data

	Average CI	PSA	Prostate Size	Tobacco Use	Tumor Type	BMI	Age
Average CI	1.0000	-0.2947	-0.2081 0.2434	0.0774 0.7528	0.0000	0.1214 0.6204	0.1241 0.6127
PSA	-0.2947 0.2207	1.0000	0.5447 0.0227	0.0242 0.9217	0.0000	-0.2143 0.3782	0.1736 0.5057
Prostate Size	-0.2081 0.2434	0.5447 0.0227	1.0000	-0.1931 0.5904	0.0000	0.0242 0.8257	-0.0543 0.6903
Tobacco Use	0.0774 0.7528	0.0242 0.9217	-0.1931 0.5904	1.0000	0.0000	-0.0656 0.7898	-0.1248 0.6106
Tumor Type	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000
BMI	0.1214 0.6204	-0.2143 0.3782	0.0242 0.8257	-0.0656 0.7898	0.0000	1.0000	-0.1627 0.5057
Age	0.1241 0.6127	0.1736 0.5057	-0.0543 0.6903	-0.1248 0.6106	0.0000	-0.1627 0.5057	1.0000

Appendix G. Multivariate correlation of negative screen patient data

	Average CI	PSA	Prostate Size	Tobacco Use	Tumor Type	BMI	Age
Average CI	1.0000 0.6040	-0.0474 0.6040	-0.4481 0.4657	-0.3087 0.1590	-0.0853 0.6721	0.0074 0.9710	-0.2295 0.2494
PSA	-0.0474 0.6040	1.0000	0.5035 0.0473	-0.1975 0.4760	0.5390 0.0618	0.1719 0.7726	-0.2869 0.2436
Prostate Size	-0.4481 0.4657	0.5035 0.0473	1.0000	0.2673 0.5483	0.3294 0.3276	-0.2113 0.9330	0.2511 0.9235
Tobacco Use	-0.3087 0.1590	-0.1975 0.4760	0.2673 0.5483	1.0000	0.2822 0.3046	-0.2865 0.1674	0.0148 0.9734
Tumor Type	-0.0853 0.6721	0.5390 0.0618	0.3294 0.3276	0.2822 0.3046	1.0000	0.1620 0.4194	-0.0554 0.7839
BMI	0.0074 0.9710	0.1719 0.7726	-0.2113 0.9330	-0.2865 0.1674	0.1620 0.4194	1.0000	-0.0242 0.9045
Age	-0.2295 0.2494	-0.2869 0.2436	0.2511 0.9235	0.0148 0.9734	-0.0554 0.7839	-0.0242 0.9045	1.0000