University of Dundee

DOCTOR OF PHILOSOPHY

Urinary Volatile Organic Compounds (VOCs) in the Diagnosis of Urothelial Bladder Cancer

Zhu, Simian

Award date:
2019

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 28. Jan. 2023

Urinary Volatile Organic Compounds (VOCs) in the Diagnosis of Urothelial Bladder Cancer

A thesis presented for the degree of

Doctor of Philosophy

to the University of Dundee by

Simian Zhu M.Sc. B.Sc.

Division of Imaging Science and Technology

School of Medicine

Ninewells Hospital and Medical School

University of Dundee

June/2019

Table of Contents

Table of Contents ..... **I**

List of Tables ..... **VI**

List of Figures ..... **VIII**

List of Abbreviations ..... **XIII**

Acknowledgements ..... **XV**

Declaration ..... **XVII**

Abstract **1**

1. Introduction **3**

2. Literature Review **6**

2.1. Bladder and Bladder Cancer 6

2.1.1. Bladder anatomy 6

2.1.2. Incidence of bladder cancer 6

2.1.3. Symptom of bladder cancer 7

2.1.4. Grading and staging of bladder cancer 7

2.1.5. Economics of bladder cancer 14

2.1.6. Conventional Diagnosis of Bladder Cancer 14

2.1.7. Non-invasive Diagnosis of Bladder Cancer 15

2.2. Volatile Organic Compounds and Bladder Cancer 22

2.2.1. What are VOCs? 22

2.2.2. Why are Biogenic VOCs so important? 22

2.2.3. Why does cancer patients' urine have different VOCs species/ levels? 23

2.2.4. How do biogenic VOCs generated by cancer cells get into urine? 37

2.2.5. How do biogenic VOCs cancer biomarkers help cancer grading and staging? 39

3. Development of Fluorescence Cross-Response Sensor Array System 41

3.1. Introduction 41

3.1.1. Gouterman Four Orbital Model of Porphyrins 42

3.1.2. Interactions between VOC and Fluorescence Sensitive Materials 49

3.2. Materials & Methods 53

3.2.1. System Overview 53

3.2.2. Sample processing module 54

3.2.3. LAKK-M System 57

3.2.4. Sensory film manufacture 59

3.2.5. Reaction chamber design 62

3.2.6. Control and communication components 65

3.2.7. Software and devices 67

3.3. Results & Discussion 67

3.3.1. LAKK-M 67

3.3.2. Influence of manufacturing material 68

3.3.3. Selection of sensitive materials based on the single-element test 77

3.3.4. Parameters 82

3.3.5. Film reproducibility 83

3.3.6. Sensory array layout 88

3.3.7. Fluorescence decays 88

3.4. Conclusion 90

4. Distinguish Test of Urinary VOC Biomarkers for Bladder Cancer 92

4.1. Introduction 92

4.1.1. Sensing technology 93

4.1.2. Signal processing and Pattern recognition 96

4.2. Materials & Methods 97

4.2.1. Materials 97

4.2.2. Sample processing module 97

4.2.3. Reaction chamber & sensory film 98

4.2.4. Experiment workflow 98

4.2.5. Data processing 99

4.2.6. Differential spectrum 101

4.2.7. Discriminate analysis 101

4.3. Results & Discussion 103

4.3.1. Target VOC biomarkers 103

4.3.2. Sensors responses 105

4.3.3. Data processing 109

4.3.4. Discriminate analysis 112

4.3.5. Classification results 116

4.3.6. Cost efficiency 124

4.4. Conclusion 125

5. Clinical Tests of Human Urine Samples 126

5.1. Introduction 126

5.2. Materials & Method 130

5.2.1. Sample processing module 130

5.2.2. Reaction chamber design 131

5.2.3. Patient Recruitment 132

5.2.4. Sample collecting protocol 132

5.2.5. System workflow 135

5.3. Results & Discussion 135

5.3.1. Patient demographic 136

5.3.2. Urine sample preparation 137

5.3.3. Effect of storage and freezing-thawing 140

5.3.4. Discrimination analysis 143

5.3.5. Comparison with other similar studies 155

5.4. Conclusion 155

6. Conclusion and Future Works 158

7. References 166

Appendix 1 – Publications 180

Appendix 2 – Patient Recruitment Documents 181

List of Tables

Table 2-1 World Health Organization (WHO) histological classification of tumours of the urinary tract. Reproduced with permission from the WHO[14]. 9

Table 2-2 Definitions of TNM staging system of bladder cancer, reproduced with permission from AJCC cancer staging manual, Eighth edition [20].....12

Table 2-3 AJCC Prognostic stage groups, reproduced with permission from AJCC cancer staging manual, Eighth edition [20].....13

Table 2-4 Summary of diagnostic accuracy of FDA approved bladder cancer urinary biomarkers. Reproduced with permission from Chou et al.[33].....21

Table 3-1 Fluorescence dye candidates.....60

Table 3-2 Comparison of PLA and ABS in 3D printing material properties 69

Table 3-3 Structure of fluorescence dye candidates.....74

Table 4.4-1 Urinary bladder cancer VOC biomarkers tested. 104

Table 4-2 Leave-one-out cross-validation results of the VOCs test PLS-DA classification model, latent variables = 3, R2 = 0.97, Q2 = 0.83. 118

Table 4-3 Leave-one-out cross-validation results of the VOCs test FS-LDA classification model, R2 = 1.00, Q2 = 0.77.....119

Table 4-4 Leave-one-out cross-validation results of the VOCs test PCA-LDA classification model, principle components = 4 (explanation>90%), R2 = 0.79, Q2 = 0.77 (Feature extraction: PCA) 121

Table 4-5 Comparison of the classification performance of PLSDA, PCA-LDA, and

FS-LDA. The sensitivities and specificities are generated from leave-one-out cross validation.....122

Table 5-1 Participant demographics of original recruitment.....136

Table 5-2 Updated participant demographic and cancer group members pathological diagnosis results 137

Table 5-3 Leave-one-out cross-validation results of FS-LDA model 145

Table 5-4 Leave-one-out cross-validation results of PCA-LDA classification model (principle components = 5) 146

Table 5-5 Leave-one-out cross-validation results of PLSDA classification model (4 latent variables) 147

Table 5-6 Confusion matrix of 20% random Mote Carlo cross-validation of PLSDA model, latent variables = 4 149

Table 5-7 Leave-one-out cross-validation results of PLSDA grading model (10 latent variables) 151

Table 5-8 Confusion matrix of 20% random Mote Carlo cross-validation of PLSDA grading model, latent variables = 10, 3996 out of 8000 samples failed to assigning groups. 151

Table 5-9 Leave-one-out cross-validation results of PCA-LDA grading model (3 principle components) 153

Table 5-10 Leave-one-out cross-validation results of PLSDA staging model (4 latent variables) 154

Table 5-11 Comparison of other studies of urine-based diagnosis of bladder cancer. N/A: Not available; CV: cross-validation 155

List of Figures

Figure 2-1 VOCs like aldehydes, ketones, dicarbonyls, furans, and hydrocarbons can be produced by lipid peroxidation as secondary products. (courtesy T.Shibamoto [69], reproduced with permission from Elsevier) 28

Figure 2-2 Classic lipid peroxidation and the end products (courtesy KM. Schaich [74], reproduced with permission from Elsevier). 31

Figure 2-3 Bladder cancer biomarkers and possible metabolism pathways, red highlighted potential biomarkers are upregulated, green highlighted are downregulated. (courtesy of D. Rodrigues et al. [96], reproduced with permission from Elsevier). 36

Figure 3-1 Porphine, the simplest porphyrin, showing meso- and β- carbons, which are most common locations for substituents. 43

Figure 3-2 Gouterman four-orbital model and HOMO/LUMO of porphyrin. a) Two HOMOs (a_{2u} and a_{1u}) and LUMOs (e_{gx} and e_{gy}) in the four orbitals model. b) Left: Four electronic transitions of porphyrin molecular orbitals: $a_{2u}\rightarrow e_{gy}$, $a_{1u}\rightarrow e_{gy}$, $a_{2u}\rightarrow e_{gx}$ and $a_{1u} \rightarrow e_{gx}$. Right: States and transitions diagram of porphyrins, and their correspondences with UV/Vis abs bands. (Reproduced with permission from Namuangruk et al. [129], with permission from The Royal Society of Chemistry) .. 44

Figure 3-3 Typical Porphyrin UV/Vis absorption spectrum: one Soret band and four Q bands 46

Figure 3-4 Jablonski diagram 47

Figure 3-5 UV/Vis absorption and emission (λ_{ex} = 418 nm) spectrum of TPP in toluene (courtesy of Mahesh et al. [131], reproduced with permission from Elsevier) 49

Figure 3-6 Intermolecular interactions on a semiquantitative energy scale. (Reproduced with permission from Janzen et al. [132]. Copyright (2006) American Chemical Society). 49

Figure 3-7 TCPP can bond to β-maltose but not α-lactose 51

Figure 3-8 Two structures of Methyl orange.....52

Figure 3-9 Reichardt's dye and light induced charge shift.....53

Figure 3-10 Block diagram of fluorescence sensor array-based VOCs detection system 54

Figure 3-11 VOCs sample processing module and Bronkhorst system sketch figure. LFM: Liquid Flow Metre. CEM: Controlled Evaporator and Mixer. MFC: Mass Flow Controller 55

Figure 3-12 Picture of VOCs sample processing module and Bronkhorst system, the reaction chamber used is Mark. II, detailed sketch was shown in Figure 3-16. 55

Figure 3-13 Urine sample processing module sketch figure.....57

Figure 3-14 Optic fibre layout of LAKK-M probe tip.....58

Figure 3-15 Example user interface of LAKK-M PC terminal.....59

Figure 3-16 Reaction chamber design Mark. I for single element tests.....63

Figure 3-17 Reaction chamber design Mark. II for multi-element sensor array tests, red arrow shows the direction of the film rack movement during the reading process. 64

Figure 3-18 Reaction chamber design Mark. III for urine tests.....65

Figure 3-19 Control and communication component for urine tests sketch figure....66

Figure 3-20 Controller PC terminal user interface.....66

Figure 3-21 Block diagram of PC terminal program in LabVIEW 67

Figure 3-22 Changes over time of optic power of blue light source in LAKK-M.....68

Figure 3-23 3D printing techniques. Left: Fused Deposition Modelling (FDM), Right: Stereolithography (SLA). 71

Figure 3-24 Fluorescence background of 3D printing materials 72

Figure 3-25 Fluorescence background of substrate materials.....76

Figure 3-26 Single-element test 78

Figure 3-27 Original spectrum of all 16 sensitive material candidates under blue, green, red, and uv laser 80

Figure 3-28 Deposition of Zn-PpIX (left) and ZnTPP (right) THF solution on the PVDF substrate, the Zn-PpIX are crystalized on the film, form unevenly died area.81

Figure 3-29 Original fluorescence spectrum of Eosin Y under 532nm excitation.....85

Figure 3-30 Normalized fluorescence spectrum of Eosin Y under 532nm excitation85

Figure 3-31 Normalized coefficient of variation of 10 repeats of Eosin Y fluorescence spectrum under 532nm excitation 86

Figure 3-32 nCV of 8 selected sensitive materials, measurement number = 10.....88

Figure 3-33 24-elements sensory array layout and sensitive material composition ..88

Figure 3-34 emission intensity decays of prophyrin sensitive materials element point P2 (TPP) and P6 (TFPP) on PVDF film. 90

Figure 4-1 Analogy of the biological and the electronic noses (courtesy Ghasemi-Varnamkhasti et al. [141], reproduced with permission from Elsevier). 93

Figure 4-2 Flow diagram of data processing 100

Figure 4-3 Differential Spectrum and feature selection.....102

Figure 4-4 Differential Spectrum of each sensor point in response to ethylbenzene 106

Figure 4-5 Differential Spectrum of each sensor point in response to hexanal 107

Figure 4-6 Differential Spectrum of each sensor point in response to Lauric aldehyde 108

Figure 4-7 Differential Spectrum of each sensor point in response to nonanoyl chloride 109

Figure 4-8 Dataset structure of raw csv file exported from LAKK-M program.....110

Figure 4-9 Sensor point responses to VOCs under excitation by 450nm, 532nm, and 365nm laser. The differential spectrum has been denoised using Savitzky-Golay filtering [123]. 117

Figure 4-10 FS-LDA plot of urinary VOC bladder cancer biomarkers test (Feature selection) 119

Figure 4-11 PCA-LDA plot of urinary VOC bladder cancer biomarkers test (Feature extraction: 4 components PCA) 121

Figure 4-12 a. PLS-DA score plot of various concentration tests of nonanoyl chloride, vapour concentration: 200 ppt ~ 200ppb, reaction time: 120s, carrier gas: nitrogen. b. Differential spectrum of element 1 under 450nm laser excitation when exposed to various concentrations of nonanoyl chloride. c. Differential spectrum of element 5 under 532nm laser excitation when exposed to various concentrations of nonanoyl chloride [123]. 123

Figure 5-1 Reaction chamber Mark. III 131

Figure 5.5-2 Recruitment flowchart for each case. WLC: white light cystoscopy; PIS: patient information sheet; CF: consent form; TURBT: transurethral resection of

bladder tumour. 134

Figure 5-3 Vapour pressure curves of different Azeotrope solution compared to ideal Raoult's law[162]. 139

Figure 5-4 Comparison of signal profile of fresh and frozen-thawed urine samples 143

Figure 5-5 Canonical score plot of FS-LDA classification model.....144

Figure 5-6 Score plot PCA-LDA classification model, principle components = 5..146

Figure 5-7 Score plot of PLSDA classification model, latent variables = 4 147

Figure 5-8 Score plot of PLSDA grading model, latent variables = 10 150

Figure 5-9 Canonical score plot of PCA-LDA grading model, principle components = 3.....152

Figure 6-1 Potential techniques to minimize the size of the devices: a. ultrasonic automizer. b. finger-tip sized spectrometers made by HAMAMATSU™ 163

Figure 6-2 Example smart phone user interface of future mobile urinary VOCs diagnosis device. Left to right: patient information enrolment page; fluorescence sensor data collection page; discriminant analysis score demonstration page.....164

List of Abbreviations

Acrylonitrile butadiene styrene (ABS)

Adenosine triphosphate (ATP)

Alcohol dehydrogenase (ADH)

Aldehyde dehydrogenase (ALDH)

American joint committee on cancer (AJCC)

Artificial neural networks (ANN)

Bladder tumor antigen (BTA)

Cancer stem cells (CSCs)

Conducting polymer (CP)

Computerised tomography (CT)

Cytochrome P450 (CYP)

Dimethylformamide (DMF)

Electron transport chain (ETC)

Enzyme-linked immunosorbent assay (ELISA)

Feature extraction (FS)

Flexible cystoscopy (CSC)

Fluorescence in situ hybridization (FISH)

Food and drug administration (FDA)

Fused deposition modelling (FDM)

Gas chromatography (GC)

Highest-energy occupied molecular orbital (HOMO)

Intravenous urogram (IVU)

k-nearest-neighbours (k-NN)

Limit of detection (LOD)

Linear discriminant analysis (LDA)

Liquid chromatography (LC)

Lowest-energy unoccupied molecular orbital (LUMO)

Magnetic resonance imaging (MRI)

Malondialdehyde (MDA)

Mass spectroscopy (MS)

Metal oxide semiconductor (MOS)

Metal oxide semiconductor field effect transistor (MOSFET)

Multidisciplinary team (MDT)

Muscle-invasive bladder cancer

National health service (NHS)

Nicotinamide adenine dinucleotide (NADH)

Nicotinamide adenine dinucleotide phosphate (NADPH)

Non-muscle-invasive bladder cancer (NMIBC)

Nuclear magnetic resonance (NMR)

Nuclear matrix protein 22 (NMP22)

Oxidative phosphorylation (OXPHOS)

Papillary urothelial neoplasm of low malignant potential (PUNLMP)

Paris system for reporting urine cytology (PSRUC)

Partial least squares discriminant analysis (PLS-DA)

Participant information sheet (PIS)

Photodynamic diagnosis (PDD)

Photosensitizing (PS)

Polylactic acid (PLA)

Polyunsaturated fatty acid (PUFA)

Polyvinylidene fluoride (PVDF)

(MIBC)

Positron emission tomography (PET)

Principle components (PCs)

Principal component analysis (PCA)

Quartz microbalances (QMB or QCM)

Reaction oxygen species (ROS)

Reduced flavin adenine dinucleotide (FADH2)

Selected ion flow tube mass spectrometry (SIFT-MS)

Stereolithography (SLA)

Superoxide dismutase (SOD)

Support vector machine (SVM)

Surface acoustic wave (SAW)

Tetrahydrofuran (THF)

Tetraphenyl porphyrin (TPP)

Transurethral resection of bladder tumours (TURBT)

Tricarboxylic acid cycle (TCA cycle)

Volatile organic compounds (VOCs)

White light cystoscopy (WLC)

World Health Organization (WHO)

Acknowledgements

I would like to dedicate this thesis to my grandmother, Xixiu Zhou, who passed away last year during my thesis writing. In the last few years of her life she has been fighting against chronic kidney disease, I hope my humble works could make a step forward for people to have better understand of those painful diseases.

I would like to give my sincerely thanks to my supervisors, Prof. Ghulam Nabi and Prof. Zhihong Huang, for their guidance, support and encouragement during past years. I will keep all your guidance and speeches deeply in my heart.

I would like to thank Dr. Chunhui Li for helping me in both academic and life, I would also like to thank her for introducing me the signal processing and data analysis algorithms.

I would like to thank all clinical staff for their kindly helps and supports during the patient recruitment of the study: Kate, Allison, Cristine, Mr. Nadwani, and Mr. Kata from NHS Tayside and Ninewells hospital.

I would like to thank Dr. Nikola Krstajic for his help in proofreading and comments of my paper and help in analysing fluorescence data.

I would like to offer my thanks to my colleges in urology group: Dr Cheng Wei, Dr. Scott Palmer and Dr. Naief Dahran for their kindly helps and instructive discussions for the past years, it is my honour to work with you. Thanks also goes to Dr. Yuting Ling, Dr. Yunwei Xu, Dr. Guangying Shangguan, Dr. Xiaowei Zhou, Dr. Tingyi Jiang and Mr. Kanheng Zhou for their helps and encouragement through my PhD.

Finally, I would like to thank my family for their understanding, encouragement and endless support in my life, I love you all forever.

Declaration

I, Simian Zhu, certify that this thesis has been written by me as a record of work carried out by me, that it has not been submitted in any previous application for a higher degree and that all references cited have been consulted by me.

Signature of candidate……. Date…….

I confirm that Simian Zhu has completed the minimum period of registration for full-time study at the University of Dundee and has fulfilled the conditions of the University of Dundee, thereby qualifying him to submit his thesis in application for the degree of Doctor of Philosophy.

Signature of supervisor…….. Date……..

Abstract

Bladder cancer is the seventh most common cancer in the UK (2011), fourth most common cancer in men and thirteenth most common in women. The total annual cost of bladder cancers in the UK was £55.39 million. The cost per patient was £8349 per year, including diagnosing, treating and following up, as well as indirect costs like loss of earnings etc. Cystoscopy with biopsy is the current gold standard for diagnosing bladder cancer. However, it is expensive and invasive, and has difficulties in diagnosing early stage cancer or *carcinoma in situ* (flat tumour). Non-invasive diagnosis of bladder cancer like urine cytology and urinary biomarker tests have poor specificity therefore they are used as supportive tests rather than definitive tests. The aim of this study is to develop a detecting system for bladder cancer urinary volatile organic compounds (VOCs) biomarkers and explore its application in non-invasive diagnosis of bladder cancer. Urinary VOCs are a class of chemical compounds found in urine that generated from biological activities of cells, cancer cells have special metabolic alternations compare to normal cells, therefore having different VOCs profiles that can be used for discrimination. Compare to conventional mass-spectroscopy and e-nose techniques for VOCs detection, the fluorescence sensor array has good detection accuracy and much lower cost of use, it is especially ideal for developing non-invasive point-of-care device for bladder cancer diagnosis and surveillance purposes. This study successfully developed a fluorescence cross-response sensor array system for diagnosing bladder cancer by detecting the urinary VOCs. On this system, a distinguishing test of four urinary VOC biomarkers: ethylbenzene, hexanal, lauric aldehyde, and nonanoyl chloride, was undertaken and achieved a sensitivity of 77.75% and a specificity of 93.25%. In a proof-of-principle clinical trial involving 79 participants (38 bladder cancer patients and 41 healthy controls), this system using a PLSDA model successfully identified over 80% of urine

samples (86.08% with leave-one-out, 81.76% with Mote Carlo cross-validation) with 77.42%-84.21% sensitivity and 85.82%-87.80% specificity. This study revealed the possibility of using low-cost optical sensor system for medical diagnosis purposes, and this could inform larger scale multi-centre trails of biogenic VOCs diagnosis and other possible clinical applications in the future.

1. Introduction

Bladder cancer is the seventh most common cancer in the UK (2011), fourth most common cancer in men and thirteenth most common in women. In recent years, its incidence rates in general are decreasing (probably due to classification changes, reductions in smoking and exposure to occupational carcinogens), but it still has the highest incidence rates in older men and women [1-4].

Bladder cancer also has been considered as one of the most expensive cancer among all cancer types, this is mainly because bladder cancer has very high recurrence rate and lacks inexpensive surveillance tools. According to National Institute for Health and Care Excellence (NICE) guidance of bladder cancer management, each year the treatment for each newly diagnosed bladder cancer patient costs £1480 at beginning, with £248 for each follow-up procedure for each low-risk non-muscle-invasive bladder cancer patient and the number of follow-up appointments for those patients are 7 per year for five years after confirmed diagnosis[5].

Cystoscopy with biopsy is the current gold standard for diagnosis of bladder cancer. However, as summarized by NICE guidance, this procedure is expensive and uncomfortable, and it is an invasive test. The current approaches for improving the bladder cancer diagnosis are mainly focusing on two directions: one is to consistently improve the diagnostic accuracy of cystoscopy by applying novel medical imaging techniques in order to better distinguish between cancerous tissue and non-cancer but easily confused tissue; another approach is to use the blood or urine biomarkers as supplementary means for conventional diagnosis methods and to lower the cost and number of follow-up appointments for bladder cancer surveillance.

The latter approach now had successfully delivered several commercial products on the market: qualitative (NMP22, Alere) and quantitative (BladderChek, Alere) nuclear

matrix protein 22 (NMP22) test; qualitative (BTA stat, Polymedco) and quantitative (BTA TRAK, Polymedco) bladder tumor antigen (BTA) test; fluorescence in situ hybridization (FISH) (UroVysion, Abbott Molecular); and fluorescence immunohistochemistry (ImmunoCyt, Scimedx) are the six FDA approved urinary biomarker tests for bladder cancer diagnosis and surveillance. Those urine-based assay have sensitivity ranged 0.58-0.78, and specificity ranged 0.74-0.88. Compare to conventional cystoscopy and cytology, those urine-based diagnosis and surveillance assay has similar sensitivities and specificities, with much lower cost per patient per test. However, all of six tests are laboratory based, which means they require trained specialists to perform those tests, therefore limited their applications. The point-of-care diagnosis and surveillance tool for bladder cancer is still lacking.

The aims of this project are:

a) Develop a fluorescence diagnosis tool that can generate unique responses to various VOC vapours.
b) Test the aforementioned tool with urinary volatile organic compound biomarkers for bladder cancer.
c) Explore the application by test the tool with clinical samples from bladder cancer patients, for example, human urine.

To achieve this target, first a strong theoretical background of the mechanism of urinary VOCs and its relationship with presences of bladder tumour must be established. Chapter 2 is the literature review about bladder and bladder cancer, biogenic VOCs and their pathways, metabolism alternation of cancer and generation of VOC biomarkers, and correlations between VOC biomarkers and cancer grading and staging. Chapter 3 further introduced the physiochemistry background of fluorescence gas sensitive materials and the principle of interactions between VOC and such materials, then this chapter introduced a custom-built detecting system for

VOC samples and a sensor array made from fluorescence gas sensitive materials, and discussed the properties of the detecting system and the sensor array film. Chapter 4 introduced and compared the common gas sensing techniques and signal processing and pattern recognition algorithm, then tested the detecting system from chapter 3 with four purified samples of urinary VOC biomarkers for bladder cancer. Chapter 5 reviewed and compared recent studies about application of urinary VOC biomarkers in diagnosis of bladder cancer and discussed the results of the clinical study about application of aforementioned fluorescence urinary VOC detecting system in diagnosis, grading and staging of bladder cancer using human urine samples. Chapter 6 and 7 concluded the works done so far and prospected the future works.

2. Literature Review

This literature review covered three topics: bladder cancer and current diagnosis methods, theoretical background of bladder cancer volatile organic compounds (VOCs) biomarkers, and photophysical mechanisms of fluorescence VOCs sensors. The aim of this chapter is to provide a comprehensive background knowledge of urinary VOCs diagnosis of bladder cancer from biological, medical, and engineering perspectives.

2.1. Bladder and Bladder Cancer

2.1.1. Bladder anatomy

The bladder lies within the pelvic cavity, behind the symphysis pubis. It can expand superiorly into the abdominal cavity when full and is entirely situated in the pelvic cavity when empty. The empty bladder is a three-side pyramid shaped organ with an apex, a base, a superior surface, and two inferolateral surfaces. The apex of the bladder is connected to the anterior abdominal by the median umbilical ligament. The base of the bladder is an inverted triangle shaped surface and faces posteroinferiorly; the urethra drains inferiorly from the under vertex of the triangle, while two ureters enter the bladder from the two upper base vertexes. Unlike other parts of the bladder mucosa, the mucosal lining of the area between two ureters openings and urethra opening is firmly attached to the smooth muscle wall to prevent mucosa fold, keeping a smooth area. This smooth triangle area also called Trigon [6].

2.1.2. Incidence of bladder cancer

Bladder cancer is the seventh most common cancer in UK (2011), fourth most common cancer in men and thirteenth most common in women. In recent years, its incidence rates in general are decreasing (probably due to classification changes, reductions in smoking and exposure to occupational carcinogens), but it still has the highest incidence rates in older men and women [1-4]. The crude mortality rate of bladder

cancer is 8.2 deaths per 100,000 individuals in the UK (2012). Similar to incidence rates, mortality is strongly related to age, with an average of 68% of bladder cancer deaths in the UK between 2010 and 2012 occurring in those aged 75 or more [1-4]. The incidence and mortality ratios between male and female are around 35:10 and 2:1, respectively. The cause of such gender disparity is still unclear but tobacco smoking and the presence of androgen receptors are considered possible reasons [7].

2.1.3. Symptom of bladder cancer

The most common symptom of bladder cancer is micro- or macro-haematuria (blood in urine), occurring in 13.7% and 78.3% of patients, respectively [8]. Although haematuria is a common early sign of bladder cancer, it is still easily missed because: it is painless; it is transient; microscopic haematuria is invisible to the naked eye. Visible haematuria is usually linked with higher grading and more advanced cancer. However, having visible haematuria does not mean it must be cancer; only 10% of patients with visible haematuria are diagnosed with bladder cancer [9, 10]. Considering 2-7% of men and 3-15% of women in the general adult population have microscopic haematuria [11], the cost-effectiveness of screening using dip-stick tests is too small, even in high-risk groups like heavy smokers and those with high occupational carcinogen exposure [12].

2.1.4. Grading and staging of bladder cancer

Histologically, the most common type of bladder cancer is urothelial carcinoma, constituting around 75%-90% of all cases [13]. As shown in Table 2-1[14], there are differentiations and variants of urothelial carcinoma, squamous and glandular being the most common, accounting for 60% and 6% of cases, respectively [13]. The variants are linked with pathologic stage and aggressiveness. For example, the micropapillary urothelial carcinoma is highly aggressive, has lymphovascular invasiveness, and is

easily metastasised to local lymph nodes and distant sites, which behave differently than other high-grade but locally advanced bladder tumours. Thus, the treatment plans concerning early cystectomy or neoadjuvant therapy are controversial [14]. Another example is plasmacytoid urothelial carcinoma, this variant is rare but aggressive and has unique discohesive plasmacytoid morphology, showing a strong predisposition for peritoneal spread. It therefore has poor prognosis [15].

Urothelial tumours
Infiltrating urothelial carcinoma
- with squamous differentiation
- with glandular differentiation
- with trophoblastic differentiation
- Nested
- Micropapillary
- Lymphoepithelioma-like
- Lymphoma-like
- Plasmacytoid
- Sarcomatoid
- Giant cell
- Undifferentiated

Non-invasive urothelial neoplasias
- Urothelial carcinoma in situ
- Non-invasive papillary urothelial carcinoma, high grade
- Non-invasive papillary urothelial carcinoma, low grade
- Non-invasive papillary urothelial neoplasm of low malignant potential
- Urothelial papilloma
- Inverted urothelial papilloma

Squamous neoplasms
- Squamous cell carcinoma
- Verrucous carcinoma
- Squamous cell papilloma

Glandular neoplasms
Adenocarcinoma
- Enteric
- Mucinous
- Signet-ring cell
- Clear cell

Villous adenoma
Neuroendocrine tumours
- Small cell carcinoma
- Carcinoid
- Paraganglioma

Melanocytic tumours
- Malignant melanoma
- Nevus

Mesenchymal tumours
Rhabdomyosarcoma
Leiomyosarcoma
Angiosarcoma
Osteosarcoma
Malignant fibrous histiocytoma
Leiomyoma
Haemangioma
Other

Haematopoietic and lymphoid tumours
Lymphoma
Plasmacytoma

Miscellaneous tumours
Carcinoma of Skene, Cowper and Little glands
Metastatic tumours and tumours extending from other organs

Table 2-1 World Health Organization (WHO) histological classification of tumours of the urinary tract.

Reproduced with permission from the WHO[14].

From a histopathological point of view, there are non-muscle-invasive bladder cancer (NMIBC) and muscle-invasive bladder cancer (MIBC), the former constitutes 75% of patients, and the latter constitutes 25% [16]. Two concepts, **Grading** and **Staging,** are the most important factors in bladder cancer prognosis and management.

The term "**Grading**" is based on the differentiation of the cells; higher differentiation means the cells are closer to normal tissue, thus the cancer has a lower grade and degree of malignancy. Meanwhile, poorer differentiation means the cells are reversed from differentiation (anaplasia), have lost the mature morphological characteristics and leads to excessive growth (neoplasm), thus the cancer has a higher grade and degree of malignancy. "**Staging**" is based on the depth of tumour invasion and metastasis and relies on clinical examinations like cystoscopy and transurethral resection biopsies. Although it cannot be exact, the histological staging is still the gold standard, and provides important information for evaluation of the risk of tumour prognostication [15].

Both concepts are important for evaluating the prognosis and management of cancer, but for the two types (NMIBC and MIBC), each concept differs in importance. For NMIBC, grading is more important, because the low-grade tumour is less aggressive than a high-grade tumour, therefore is considered as "low-risk cancer". While for MIBC, the more important prognostic factor is staging. This is because the aggressiveness of the tumour has been confirmed by its invasiveness and how far the tumour penetrates the bladder wall; whether it invades the surrounding tissue and nodes becomes the major concern for prognosis and subsequent management.

In 1973, the World Health Organization introduced the numeric grading system for urothelial carcinoma. This system has four grades, starting from papilloma, and then grade 1 to 3. In 2004, the WHO updated the grading system of urothelial carcinoma and included the papillary urothelial neoplasm of low malignant potential (PUNLMP)

as a new category, replacing the grade 1 category of the 1973 system [17]. The most significant change is the low grade, formerly grade 2. Almost 40% of grade 2 cases in the 1973 system are categorized higher in the 2004 system however studies have found that both systems are significant in describing the disease progression and recurrence [15]. Such a grading system is widely accepted and inherited by the latest 2016 WHO grading system [18].

The widely accepted staging system for bladder cancer is the American Joint Committee on Cancer (AJCC) TNM staging system. In this system, the letters T, N, and M stand for primary tumour, regional lymph node metastasis and distant metastasis, respectively. Each category represents the relevant clinical and pathological classification of the patient and, combined together, provide prognostic factors for stage grouping [19].

a. Definition of Primary Tumour (T)

T Category	T Criteria
TX	Primary tumour cannot be assessed
T0	No evidence of primary tumour
Ta	Non-invasive papillary carcinoma
Tis	Urothelial carcinoma *in situ*: "flat tumour"
T1	Tumour invades lamina propria (subepithelial connective tissue)
T2	Tumour invades muscularis propria
pT2a	Tumour invades superficial muscularis propria (inner half)
pT2b	Tumour invades deep muscularis propria (outer half)
T3	Tumour invades perivesical soft tissue
pT3a	Tumour invades perivesical soft tissue microscopically
pT3b	Tumour invades perivesical soft tissue macroscopically (extravesical mass)
T4	Extravesical tumour directly invades any of the following: prostatic stroma, seminal vesicles, uterus, vagina, pelvic wall, abdominal wall
T4a	Extravesical tumour invades directly into prostatic stroma, seminal vesicles, uterus, vagina
T4b	Extravesical tumour invades pelvic wall, abdominal wall

b. definition of Regional Lymph Node (N)

N Category	N Criteria
NX	Lymph nodes cannot be assessed
N0	No lymph node metastasis
N1	Single regional lymph nodes metastasis in the true pelvis (perivesical, obturator, internal and external iliac, or sacral lymph node)
N2	Multiple regional lymph nodes metastasis in the true pelvis (perivesical, obturator, internal and external iliac, or sacral lymph node metastasis)
N3	Lymph node metastasis to the common iliac lymph nodes

c. Definition of Distant Metastasis (M)

M Category	M Criteria
cM0	No distant metastasis
cM1	Distant metastasis
cM1a	Distant metastasis limited to lymph nodes beyond the common iliac
cM1b	Non-lymph-node distant metastasis
pM1	Distant metastasis, microscopically confirmed
pM1a	Distant metastasis limited to lymph nodes beyond the common iliac, microscopically confirmed
pM1b	Non-lymph-node distant metastasis, microscopically confirmed

Table 2-2 Definitions of TNM staging system of bladder cancer, reproduced with permission from AJCC cancer staging manual, Eighth edition [20]

Table 2-2 shows the criteria for classification for each category in bladder cancer staging. Specifically, the category T is defined by the size and contiguous extension of the primary tumour. As previously mentioned, 75% of all cases are NMIBC and about 70% of those NMIBC cases are non-invasive papillary carcinoma (Ta), 20% are T1 and 10% are Tis [21]. The category N is defined by the absence/presence of cancer in regional lymph nodes and the number of positive lymph nodes. The most frequent lymph node metastasis sites are perivesical, internal and external iliac and obturator basins [20]. The category M is defined by the absence/presence of distant metastasis. The prefixes of cM and pM represent the time point each stage is assigned: clinical (before treatment) and pathological (after surgical removal and being reviewed by specimens), respectively. Retroperitoneal lymph nodes, lung, bone, and liver are the most common distant metastasis sites of bladder cancer [20]. Table 2-3 shows the AJCC stage group assignment by combination of T, N, and M categories.

When T is…	When N is…	When M is…	Then the stage group is…
Ta	N0	M0	0a
Tis	N0	M0	0is
T1	N0	M0	I
T2a	N0	M0	II
T2b	N0	M0	II
T3a, T3b, T4a	N0	M0	IIIA
T1-T4a	N1	M0	IIIA
T1-T4a	N2, N3	M0	IIIB
T4b	Any N	M0	IVA
Any T	Any N	M1a	IVA
Any T	Any N	M1b	IVB

Table 2-3 AJCC Prognostic stage groups, reproduced with permission from AJCC cancer staging manual, Eighth edition [20].

2.1.5. Economics of bladder cancer

Surveillance is very important in the management and follow-up of bladder cancer patients. This is because of both the high recurrence rate of NMIBC and MIBC and the short progression time of metastatic disease. The surveillance strategies are based on the combination of grading, staging, and other pathological characteristics. In general, low-risk patients need over 5 years of surveillance after initial surgery, intermediate-risk patients are under lifelong surveillance with adjuvant immunotherapy and high-risk patients need multiple surgeries with extending adjuvant immunotherapy. From an economical perspective, bladder cancer is a very “expensive” disease. This is due to not only direct costs from long-term surveillance, but also indirect costs from loss of earnings, time costs, caregiving, and physical and social functioning loss with the patients [22]. In a study published in 2004, the total annual cost of bladder cancers in the UK was £55.39 million. The cost per patient was £8349 per year, including diagnosing, treating and following up, as well as indirect costs like loss of earnings. The cost of low-grade bladder cancer was £35.25 million, about £33 million of that was direct costs, resulting in a significantly heavy burden on National Health Service (NHS) resources [23].

2.1.6. Conventional Diagnosis of Bladder Cancer

Most patients with early stage bladder cancer start with haematuria, sometimes accompanied by frequent or urgent urination and dysuria. Patients are asked to provide urine samples for urinary cytology test. Through microscopic observation of the shedding epithelium cells from the urine tract and bladder, urinary cytology can detect highly abnormal cells, making it ideal for diagnosing high-grade tumours. For the low-grade tumours, cell morphology changes may not be that obvious, therefore diagnosis relies on the inspector’s experience. However, inspector’s experience should be treated

as a supportive test rather than a definitive test. Cystoscopy with biopsy is the current gold standard for diagnosing bladder cancer. A flexible cystoscope is usually used for cystoscopy observation only; it has a flexible thin tube mounted with a light source and camera and only requires local anaesthetic. The light source can illuminate the area within the bladder, allowing for the detection of visible morphological changes and the extent of the tumour. A rigid cystoscope has thick and rigid tubes, which allows larger surgical instruments to enter the bladder though the urethra. For example, instruments for biopsy using transurethral resection of bladder tumours (TURBT) [24]. In addition to conventional white light cystoscopy (WLC), various cystoscopy techniques have been developed. Photodynamic diagnosis (PDD) is used for diagnosing bladder cancer because it can distinguish the tumour from healthy tissue under the specific light wavelength. Photosensitizing (PS) agent can be absorbed by cancer cells and causes the cells to have a different fluorescence characterisation compared to normal cells. Recently, PS-free diagnosis systems based on endogenous tissue signals (like Raman spectroscopy, autofluorescence spectroscopy, and narrow band imaging) were developed in order to countervail some drawbacks of PS based PDD.

Most of the imaging techniques have relatively low tissue penetration (<500μm), which limits the application in diagnosing the invasiveness and metastasis of cancer. Computerised tomography (CT) and intravenous urogram (IVU) can help to gain depth and whole images of the bladder and urine tract. Magnetic resonance imaging (MRI) and Positron emission tomography (PET) can provide the information about metastasis and conditions of nearby lymph nodes and soft tissue.

2.1.7. Non-invasive Diagnosis of Bladder Cancer

2.1.7.1. Urine Cytology

Urine cytology is a pathological interpretation of the morphological features of shed

cells in stained cellular smears from urine [25]. Urine cytology has a reported sensitivity of 44% (38% ~ 51%) and specificity of 96% (94% ~ 98%) [26]. Such statistics perfectly reflect the nature of the stand-alone urine cytology in bladder cancer diagnosis. Urine cytology has very good specificity, which means abnormal cells found through this method are highly likely to be cancer cells and can be considered as strong evidence of urothelial malignancies. However, the relatively low sensitivity means some cases may be missed, especially those with low-grade malignancies, which has minor morphological changes [27].

There are several factors that affect the accuracy of urine cytology: specimen collection; variation in specimen processing; subjective interpretation; and lack of reporting standards [27]. Besides from positive and negative, there are other terminologies used in the urine cytology report, like suspicious and atypical; the classification of those groups remains controversial and may affect the sensitivity of the diagnosis. To address this problem, The Paris System for Reporting Urine Cytology (PSRUC) was formed in 2016 to establish a universal standard for urine cytology reports. The aim of this system is to improve the communication between pathologist and clinician by standardising the terms used in the urine cytology reports. It is surprising that, until recent years, pathologists around the world did not have standardised language. Now, the first generic terminology for urine cytology is in use and several studies have shown optimistic findings as a result. For example: lower rates of atypical urothelial cells diagnosis [28]; better characterization of atypical urothelial cells, low-grade urothelial neoplasm, or suspicious for high-grade urothelial carcinoma [29]; and increased correlation between surgical biopsy diagnosis and urinary cytology diagnosis [30]. Even though the Paris system has been used in a limited number of institutions for only one year, the preliminary results have shown positive improvement in urine cytology practices [31].

2.1.7.2. Urinary Biomarkers

Over the past 30 years, several urinary biomarkers have been discovered; some of them have received U.S. Food and Drug Administration (FDA) approval and are commercially available for bladder cancer diagnosis and surveillance. In general, urinary biomarkers can be classified into several types by their detection targets: tumour-associated antigens; blood group antigens; growth factors; cell cycle/apoptosis; and extracellular matrix proteins. There are six commercially available bladder cancer urinary biomarker tests approved by the FDA: qualitative (NMP22, Alere) and quantitative (BladderChek, Alere) nuclear matrix protein 22 (NMP22) test; qualitative (BTA stat, Polymedco) and quantitative (BTA TRAK, Polymedco) bladder tumor antigen (BTA) test; fluorescence in situ hybridization (FISH) (UroVysion, Abbott Molecular); and fluorescence immunohistochemistry (ImmunoCyt, Scimedx).

a) Nuclear matrix protein 22 (NMP22)

Nuclear matrix protein 22 is commonly found as part of the mitotic apparatus released from urothelial nuclei upon cellular apoptosis. The NMP22 has been found to have a significantly higher concentration in the urine of cancer patients than the urine of healthy controls [32], and has been used as a bladder cancer urinary biomarkers for more than 20 years. The NMP22 qualitative test (BladderChek) is an immunochromatographic assay, only a few drops of urine are needed on the testing card, and the colour change indicates the result.

The NMP22 quantitative test is an enzyme-linked immunosorbent assay (ELISA) that is able to measure the concentration of the NMP22 protein in the urine specimen. The qualitative test has been approved as a point-of-care test for diagnosis and surveillance, while the quantitative test is approved for surveillance and can only be carried out in laboratory.

Mowatt et al. analysed 41 studies consisting of 13,885 participants and reported the

performance of NMP22. It was found that NMP22 achieved a sensitivity of 0.68 (95% CI 0.62-0.74) and specificity of 0.79 (95% CI 0.74-0.84). NMP22 was more sensitive than urine cytology (0.44, 95% CI 0.38-0.51), but less specific (0.96, 95% CI 0.94-0.98) [26]. Another study carried out by Chou et al. directly compared all six commercially available urinary biomarker tests, including the NMP22[33]. By summarizing 19 quantitative NMP22 studies (consisting of 3,555 participants) and 4 qualitative NMP22 studies (consisting of 2,039 participants), the authors found quantitative NMP22 tests had higher sensitivity than qualitative NMP22 tests (0.69 [95% CI 0.62-0.75] vs. 0.58 [95% CI 0.39-0.75]). In contrast, specificity is higher in qualitative NMP22 tests (0.77 [95% CI 0.70-0.83] vs. 0.88 [95% CI 0.78-0.94]).

There are several interference factors that affect the accuracy of NMP22 tests. NMP22 is released during cellular apoptosis, therefore the benign urologic changes (involving cell-turnover and apoptosis, like infection, stone, haematuria and instrumentation) may increase NMP22 release and cause false-positive results [34].

b) Bladder tumor-associated antigen (BTA)

There are two commercially available urinary biomarker tests based on detection of human complement factor H (cFH)-related protein. Factor H can interrupt the complement cascade and bind and interrupt the activation of C3b, thus evade the attack from the immune system and gain selective growth advantage to cancer cells [35]. Like NMP22, BTA tests also have qualitative point-of-care immunochromatographic and quantitative lab-based ELISA assay: branded BTA stat and BTA TRAK, respectively. Both BTA stat and BTA TRAK received FDA approval for diagnosis and management of bladder cancer, in conjunction with cystoscopy.

In the same meta-analysis, Chou et al. found the sensitivities of BTA stat and BTA TRAK were 0.64 (95% CI 0.58-0.69) and 0.65 (95% CI 0.54-0.75), while the specificities were 0.77 (95% CI 0.73-0.81) and 0.74 (95% CI 0.64-0.82),

respectively[33]. This study included 21 studies for qualitative BTA stat with 894 participants, and 4 studies for quantitative BTA TRAK with 125 participants. The relative low sensitivity and specificity of BTA tests are the main reason that prevent it from being used alone, instead of as supplementary method for cystoscopy.

Like NMP22, BTA assays also have interference issues, because Kupffer cells, hepatocytes, vascular endothelial cells, and platelets can produce and secrete factor H-related proteins into the serum as well as bladder cancer cells. Therefore, urinary tract infections, stones, installing catheters or stents and any conditions that may cause haematuria will bring significant impact of false-negative results [36, 37].

c) Fluorescence in situ hybridization (FISH)

UroVysion/uFISH is a multi-target, multi-colour fluorescence, in situ hybridization assay test that detects aneuploidy in chromosomes 3, 7, and 17, and the loss of the 9p21 locus in urothelial cells [38]. FISH has been approved by FDA for bladder cancer diagnosis and surveillance in adjunct to cystoscopy and urine cytology. Similar to urine cytology, FISH is reported dichotomously (positive/negative) based on cell chromosomal and morphological changes, and it also requires specialized equipment and trained personnel.

Mowatt et al. analysed the data from 14 studies consisting of 3321 participants tested with FISH and found the overall sensitivity and specificity were 0.76 (95% CI 0.65-0.84) and 0.85 (95% CI 0.78-0.92), respectively[26]. Chou et al. found 0.63 (95% CI 0.50-0.75) sensitivity and 0.87 (95% CI 0.79-0.93) specificity from 11 studies with 416 participants[33]. Compared to urine cytology, FISH has higher sensitivity but lower specificity; compared to other non-invasive urinary tests, FISH has highest specificity and competitive sensitivity [33].

The major disadvantages of FISH include the requirement of specialized equipment and trained personnel, difficulty of interpreting the atypical or non-significant

chromosomal and morphologic changes, and lack of universal standard definitions of positive results. As mentioned in AUA guideline, FISH was recommended for aiding treatment decisions for patients with equivocal cytology and/or cystoscopy[39].

d) Immunohistochemistry

The ImmunoCyt test is a fluorescence immunohistochemistry test for urine specimen. Developed by Fradet and Lockhart, the ImmunoCyt combines three fluorescence labelled antibodies (M344, LDQ10 and 19A211) against two commonly observed antigens (mucins and a high molecular weight form of carcinoembryonic antigen, which is over-expressed in most of the bladder cancer cells but not in normal transitional epithelium cells) [40]. Under the fluorescence microscope, positive shed cells are green or red while the negative cells show no fluorescence. Comploj et al. reported over 7000 ImmunoCyt and urine cytology tests between January 2002 and March 2011 [41]. This study found that the ImmunoCyt test has higher sensitivity than urine cytology (0.681 vs. 0.345) but lower specificity (0.723 vs. 0.979). The authors then reported that the ImmunoCyt test combined with conventional urine cytology can offset their respective defects and achieve better overall sensitivity (0.728) and comparable specificity (0.719). The authors confirmed the value of ImmunoCyt and cytology analyses in the follow-up of patients with NMIBC. Another meta-analysis, analysing 14 studies published between 1999 and 2006, also supported this conclusion, finding that almost all ImmunoCyt tests have better sensitivities than urine cytology alone, and specificity is improved by at least 15% when combing the two [42].

2.1.7.3. Summary

As shown in Table 2-4, among the six FDA approved bladder cancer urinary diagnosis methods, the immunohistochemistry test achieved highest sensitivity (0.84[26]/0.78[33]/0.681[41]), while the FISH achieved the highest specificity (0.85[26]/0.87[33]). Compared to urine cytology, all six biomarkers have better

performance in overall sensitivity, especially in detection of low-grade tumours. However, the false-positive rate reduces their performance in terms of specificity. Therefore, at present, using biomarkers in conjunction with cystoscopy or cytology is recommended to offset the drawbacks of the conventional diagnosis methods.

Biomarker	Sensitivity (95% CI)	Specificity (95% CI)
Quantitative NMP22	0.69 (0.62–0.75)	0.77 (0.70–0.83)
BladderChek	0.58 (0.39–0.75)	0.88 (0.78–0.94)
BTA stat	0.64 (0.58–0.69)	0.77 (0.73–0.81)
BTA TRAK	0.65 (0.54–0.75)	0.74 (0.64–0.82)
UroVysion	0.63 (0.50–0.75)	0.87 (0.79–0.93)
ImmunoCyt	0.78 (0.68–0.85)	0.78 (0.72–0.82)

Table 2-4 Summary of diagnostic accuracy of FDA approved bladder cancer urinary biomarkers. Reproduced with permission from Chou et al.[33].

In recent years, besides from the aforementioned six FDA-approved commercial urinary biomarkers, more than 100 novel urinary biomarkers have been found and tested for diagnosis and surveillance purposes in bladder cancer. A systematic review published recently summarized 115 studies of novel bladder cancer urinary biomarkers, reported between January 2013 and July 2017 [43]. Of those studies, 105 single biomarker studies achieved 2-94% sensitivity and 46-100% specificity, 10 multiple biomarker combination studies achieved 24-100% sensitivity and 48-100% specificity. 14 out of 115 studies reached ≥80% sensitivity and specificity. Of those, Orosomucoid 1 (ORM1), serine protease HtrA1, and Survivin performed the best and have the greatest potential in the application of novel clinical diagnosis and surveillance tools.

In the UK, particularly the NHS, the mean cost per test for PDD was £1371, WLC £937, flexible cystoscopy (CSC) £441, urine cytology £92, NMP22 £39, ImmunoCyt

£54 and FISH £55 [26]. Obviously, the cost of urinary biomarker tests is much cheaper than conventional invasive tests and remains competitive when compared with the cost of urine cytology. However, none of the urinary biomarker tests beat the urine cytology in head-to-head comparisons of specificity. Therefore, urinary biomarker tests are still considered supplementary methods to urine cytology in diagnosis and management of bladder cancer.

2.2. Volatile Organic Compounds and Bladder Cancer

2.2.1. What are VOCs?

As the name suggests, **volatile organic compounds (VOCs)** are a class of organic compounds with high vapor pressure at room temperature, which makes them volatile. VOCs produced from biological activities are called **biogenic VOCs** in this thesis, unless otherwise stated. Biogenic VOCs means the VOC is produced from a human cells' biological activities; the VOCs produced by microorganisms within the human body are not included in the definition. Usually, biogenic VOCs are produced from cell activities and are present in the blood stream, released by exhaled breath, skin, saliva, milk, and urine. Common biogenic VOCs include hydrocarbons, alcohols, aldehydes, ketones, esters, and aromatic compounds.

2.2.2. Why are Biogenic VOCs so important?

Biogenic VOCs contain information about nutrients concerning human health. In ancient times, Chinese and Greek doctors already knew that certain smells indicated some diseases such as corrupt breaths for digestive problems, fruity flavours from diabetes, and ammonia odours from patients with kidney or liver diseases. More recently, news reports have stated that pet dogs warned their masters by sniffing and pointing at some part of body and eventually found out there was melanoma [44] and a breast tumour [45].

Willis et al. (2004) published a study demonstrating that six dogs, trained with human urinary odour, could successfully identify 22 out of 54 bladder cancer patients, compared to a 14% accuracy rate by chance alone [46]. This is the first published clinical study of the potential of animal olfactory systems in cancer diagnosis. Since then, more urinary odour animal olfactory detection studies of lung [47, 48], breast [47] and prostate [49-51] cancer were published and showed varying accuracy, from no better than chance to higher than 99%. The animal olfactory detections of cancer are highly dependent on training and stimulus, which limits the application of such methods [52].

However, animal olfactory detection studies revealed that there are differences between the odours of cancer patients and that of healthy people. Those differences were confirmed by analytical chemistry methods like gas chromatography (GC), mass spectroscopy (MS), nuclear magnetic resonance (NMR) spectroscopy, thus establishing the link between biogenic VOCs and odour differences. So far, more than 200 urinary VOCs with a correlation with cancer have been found, in both species and concentrations levels. Detailed discussion of the biogenic VOCs cancer biomarkers will be in Chapter 5: Urine tests.

2.2.3. Why does cancer patients' urine have different VOCs species/ levels?

The different levels or species of VOCs between cancer cells and healthy cells may be induced by various metabolism and nutritional changes during tumour growth [53-55]. This topic is still in fast development; people are continuously discovering new metabolism pathways and expanding knowledge of cancer cell metabolism and its consequences. This part will summarize some the most widely accepted theories and mechanisms of the origin of biogenic VOCs cancer biomarkers.

In general, from a metabolic point of view, there are four major changes that can affect

the biogenic VOCs generation: **oxidative stress**, **cytochrome p450 detoxication**, **carbohydrates metabolism** (glycolysis/gluconeogenesis pathways), and **lipid peroxidation** [56]. Briefly, cytochrome p450 detoxication activates for cells clearance of toxins. Oxidative stress is a commonly found phenomenon in cancer cells causing lipid peroxidation. Another commonly observed change in cancer cells is special energetic metabolism pathways, for fulfilling their requirements for rapid growth and proliferation. Those biological processes will generate unique metabolites and subsequently turn into signature VOCs.

2.2.3.1. Carbohydrates metabolism

Warburg effect, a phenomenon of abnormal upregulated **aerobic glycolysis** in cancer cells, was first observed by Otto Warburg in 1920s [57]. In normal cells, adenosine triphosphate (ATP) is produced mainly from oxidative phosphorylation (OXPHOS) to provide the energy needed for cell activities. In normal cells, glucose first converts into pyruvate through glycolysis in the cytoplasm, then the pyruvate turns into acetyl-CoA at the mitochondria and participates in the tricarboxylic acid cycle (TCA cycle). Then, it produces reduced nicotinamide adenine dinucleotide (NADH) and reduced flavin adenine dinucleotide ($FADH_2$) and finally transfers the energy to ATP through electron transport chain [58].

While in many tumour cells the OXPHOS is inhibited as compensation, glycolysis is promoted as an altered metabolism pathway for producing ATP. The excessive pyruvate is turned to lactate and discharged, rather than entering the TCA cycle. From the perspective of efficient energy generation, one glucose can net produce 2 ATP through glycolysis whereas through normal OXPHOS it can net produce 36 ATP. However, the aerobic glycolysis is actually necessary for tumour cells. There are intermediate products of glycolysis and truncated TCA cycle; citric acid being transported out from mitochondria turns into malate in the cytoplasm and then

continues the TCA cycle, skipping major energy production steps. These products can be used as materials for synthesis of nucleotides, lipids and proteins, to satisfy the conditions of rapid tumour growth. Despite less ATP production, glycolysis is faster than OXPHOS, which is also an adaptation of that facilitates fast tumour growth. Lactate, the final product of glycolysis, will secrete to the extracellular environment, it will acidify the tumour microenvironment and promote tumour growth and metastasis [59].

Tali et.al studied the concentration of VOCs released by lung cancer cells *in vitro* and found that some VOCs masses changed after admission of 2DG and 3BrPA glycolysis inhibitors [60]. Glycolysis inhibitors are able to block the hyperglycolysis of cancer cells, therefore the authors infer that they can also reduce the VOCs production in relation to glycolytic and subsequent metabolic pathways. They found the concentration of methanethiol, one of the by-products of glucose and lactate metabolism, significantly decreased after admission of 3BrPA. Similarly, 2-methyl-1,3-butadiene and carbonic acid, one of the VOCs related to hypoxia condition, decreased after 3BrPA, indicating the lower pH of microenvironments of cancer cells. This study brings more understanding of the mechanism of the Warburg effect and evidences the origins of the biogenic VOCs cancer biomarkers (aerobic glycolysis).

Yamagishi et al. explained one possible mechanism of cancer cells producing odorous sulfur-containing compounds and its relationship with glycolysis [61]. They found glucose, usually largely accumulated by tumour tissue, can have Maillard reaction with methionine and produce sulfur-containing compounds like hydrogen sulfide and methanethiol (also the culprit of smelly urine after eating asparagus). Further study found lactic acid, the end product of glycolysis, can also react with sulfur-containing amino acid and produce hydrogen sulfide and methanethiol. Those VOCs were found significantly higher in flatus from colon cancer patients and exhaled breath from

patients with lung cancer compared to healthy people. Furthermore, those sulfur-containing compounds can further bind to iron atoms in haemoglobin, thereby further intensifying the histotoxic hypoxia in tumour tissue. This study is also strong evidence of how aerobic glycolysis can promote specific biogenic VOCs production, and such VOCs are not only indicators but also accomplices of growing tumours.

Another energetic metabolism alternation that happens in cancer cells is **glutaminolysis**. Similar to aerobic glycolysis, the glutaminolysis is upregulated in cancer cells to compensate for energy generation, to supplement the intermediate products for TCA cycle, and to generate nicotinamide adenine dinucleotide phosphate (NADPH) for biosynthesis and protecting against the toxicity of ROS [62]. Although the direct products of glutaminolysis are not volatile, it can still influence the biogenic VOCs production. This is reflected in: glutamate and aspartate, as precursors of nuclei acid biosynthesis; NADPH and subsequent biosynthesis; pyruvate, as in Warburg effect.

2.2.3.2. Lipid metabolism

The main functions of lipid in living cells include storing energy, acting as structural components of membranes and signalling. **Lipid metabolism** is considered one of the origins of biogenic VOCs. Like all other metabolisms, there are two subcategories of lipid metabolism: **lipid biosynthesis** (anabolism) and **lipid catabolism**. An increase in de novo lipid synthesis has been observed in many cancer types and is closely related to the rapid growth and proliferation needs of cancer cells [63]. While the lipid catabolism, especially the **lipid peroxidation**, is one of the major origins of cancer VOCs biomarkers, the mechanism is related to **reaction oxygen species (ROS)** and its damage to lipid membranes, ether by exogenous or endogenous ROS.

From the point of view of the origins, there are two classes of biogenic VOCs: endogenous and exogenous. From the patients' perspective of finding cancer

biomarkers, people prefer endogenous VOCs more because endogenous VOCs reflect the metabolic reactions within the cells, and they are also more controllable and predictable. However, the roles of exogenous VOCs cannot be ignored; those chemicals can be metabolised and eliminated in urine, but the time of clearance is varied from weeks to months, and could be longer if a duct is formed [64]. More importantly, the intake exogenous VOCs can participate in the production of ROS, thus affecting the cell metabolism and endogenous VOCs production.

ROS are a family of oxygen containing chemically reactive chemicals. Commonly found ROS in cells include peroxides (H_2O_2 and others), superoxide (O_2^-), hydroxyl radical ($^{\cdot}OH$), and singlet oxygen ($^1[O_2]$). The ROS can lead to cell damage by attacking DNA or RNA, polyunsaturated fatty acid (PUFA), amino acid in proteins, and deactivating some enzymes by oxidation of co-factors [65]. Carcinogens and mutagens, like high-energy radiation, pollutants, tobaccos, and toxins, can increase the production of ROS. The cytotoxicity and genotoxicity of ROS is considered one of the main mechanisms of their carcinogenicity and mutagenicity [66].

Besides from exogenous, cells produce ROS continuously in mitochondria as by-products of aerobic respiration. Its level is controlled by antioxidant enzymes like superoxide dismutase (SOD) and glutathione peroxidase. In normal cells, the antioxidant enzymes maintain the dynamic equilibrium between generation and deactivation of ROS. During **oxidative stress**, the balance is broken; the generation of ROS exceeds the clearance by SOD or the repair of the damage it caused, and may lead to cell damage, apoptosis or even necrosis depending the severity of oxidative stress [67].

The ROS can, on the one hand, activate the body's detoxification process of **cytochrome p450** and catalyse the exogenous VOCs into more water-soluble compounds and excrete them to the blood. On the other hand, it also can attack lipid

on the cell membranes and break it into smaller molecules. This process is done by **free radical chain reaction**. During this reaction, fatty acid lose hydrogen atoms to ROS and become fatty acid radicals, then fatty acid radicals react with molecular oxygen and produce a peroxyl-fatty acid radical. The peroxyl-fatty acid radicals react with another free fatty acid and turn it into a new fatty acid radical, thereby continuing the cycle. The peroxyl-fatty acid radical receives a hydrogen atom and becomes lipid hydroperoxide; it may further turn in to another lipid radical, other reactive products, or aldehydes [66]. This phenomenon is called **lipid peroxidation**, and it can cause cell membrane and distal tissue damages. The VOCs produced from lipid peroxidation include saturated hydrocarbons like ethane and pentane, as well as reactive aldehydes like malondialdehyde (MDA) and 4-hydroxynonenal (HNE), which has been used as a breath indicator for evaluating the level of lipid peroxidation of human body [68].

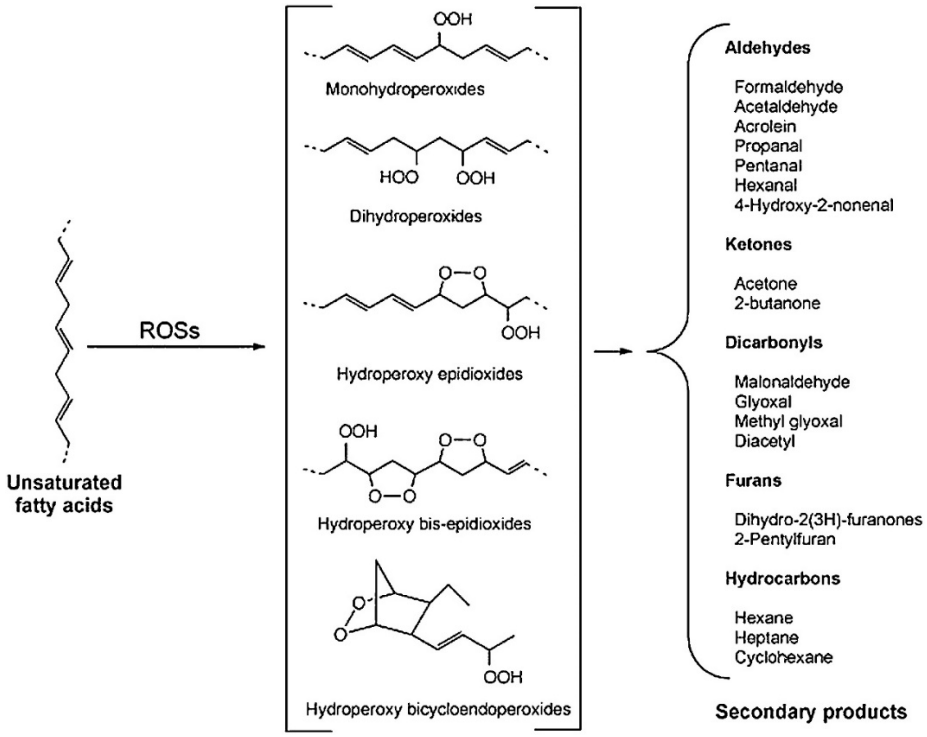


Figure 2-1 VOCs like aldehydes, ketones, dicarbonyls, furans, and hydrocarbons can be produced by lipid peroxidation as secondary products. (courtesy T.Shibamoto [69], reproduced with permission from Elsevier)

Figure 2-1 summarises a general process of production of aldehydes, ketones, dicarbonyls, furans and hydrocarbons by lipid peroxidation of unsaturated fatty acids, and some examples of intermediate products [69]. As the ROS and lipid peroxidation chain reaction is an important concept of understanding the origins of biogenic VOCs, it is necessary to explain the mechanism and product pathway in detail:

Lipid peroxidation has selective targets of fatty acids to attack. The PUFA is more sensitive than the saturated one because the $\mathrm{C-H}$ bond of the methylene group is weakened by the double bond adjacent to it, this specific site on PUFA is called the activated methylene bridge (RH). The chain reaction initialized by the activated methylene bridge (RH) causes PUFA to lose a hydrogen atom to a free radical and become an alkyl radical ($\mathrm{R^{\cdot}}$). It soon combines with a molecular oxygen and forms a lipid peroxyl radical ($\mathrm{ROO^{\cdot}}$). The lipid peroxyl radical is capable of capturing a hydrogen atom from other PUFA, thereby propagating the chain reaction [70]. In summary, this process can be expressed as:

$$\mathrm{RH + {}^{\cdot}OH \rightarrow R^{\cdot} + H_2O} \qquad (1-1)$$

$$\mathrm{R^{\cdot} + O_2 \rightarrow ROO^{\cdot}} \qquad (1-2)$$

$$\mathrm{ROO^{\cdot} + RH \rightarrow ROOH + R^{\cdot}} \qquad (1-3)$$

The lipid hydroperoxide (ROOH) is the first stable product of lipid peroxidation. However, in the presence of metal ion, there are:

$$\mathrm{Fe^{2+} + ROOH \rightarrow RO^{\cdot} + OH^{-} + Fe^{3+}} \qquad (1-4)$$

$$\mathrm{Fe^{3+} + ROOH \rightarrow ROO^{\cdot} + H^{+} + Fe^{2+}} \qquad (1-5)$$

$$\mathrm{Cu^{+} + ROOH \rightarrow RO^{\cdot} + OH^{-} + Cu^{2+}} \qquad (1-6)$$

$$\mathrm{Cu^{2+} + ROOH \rightarrow ROO^{\cdot} + H^{+} + Cu^{+}} \qquad (1-7)$$

This products of ROOH-dependant lipid peroxidation reactions are lipid peroxyl and

alkoxyl radicals ($RO^{\cdot}$), which are able to capture hydrogen atoms from other PUFA and thus re-initialise the chain reactions [71].

At the termination step, small molecule secondary products, such as those shown in Figure 2-1, are formed, along with the deactivation of free radicals, thus ending the chain reaction. When the concentration of previously formed radicals is high enough, with suitable environmental conditions (temperature, oxygen pressure, solvent etc.), the lipid free radicals can recombine and may generate stable products, thus terminating the chain reactions. For example, two peroxyl radicals ($ROO^{\cdot}$) can recombine and form an unstable tetroxide intermediate (ROOOOR) and soon break into an alcohol and a ketone. However, there are many different decay pathways of peroxyl radicals, therefore the end products are varied, too. Carbonyl groups, alcohols, unsaturated carbon bonds or epoxides, or even new radicals to propagate the radical reaction chain, are all possible [72]. Another example of diversity of end products is alkoxyl radicals ($RO^{\cdot}$): it can recombine with an alkyl radical ($R^{\cdot}$) and form an ether, ketone and alkane, or ketone and alcohol, depending on the site of radical group. Or it can recombine with another alkoxyl radical to form a lipid peroxide (R_1OOR_2). Under suitable conditions, alkoxyl radicals can have α-scission or β-scission and produce smaller aldehydes and alkyl radicals and propagate the radical reaction chain [73]. Figure 2-2 summarises the common reactions during the classical lipid peroxidation chain and the products; the end products may have further oxidations and may generate more secondary products [74].

It is worth noting that “termination” does not mean the radical reaction chain is fully stopped. As seen in aforementioned examples, there are always new free radicals being generated. Along with ROS from cellular aerobic breathing, the chain reactions will only slow down when clearance of free radicals exceeds the rate of new chain production.

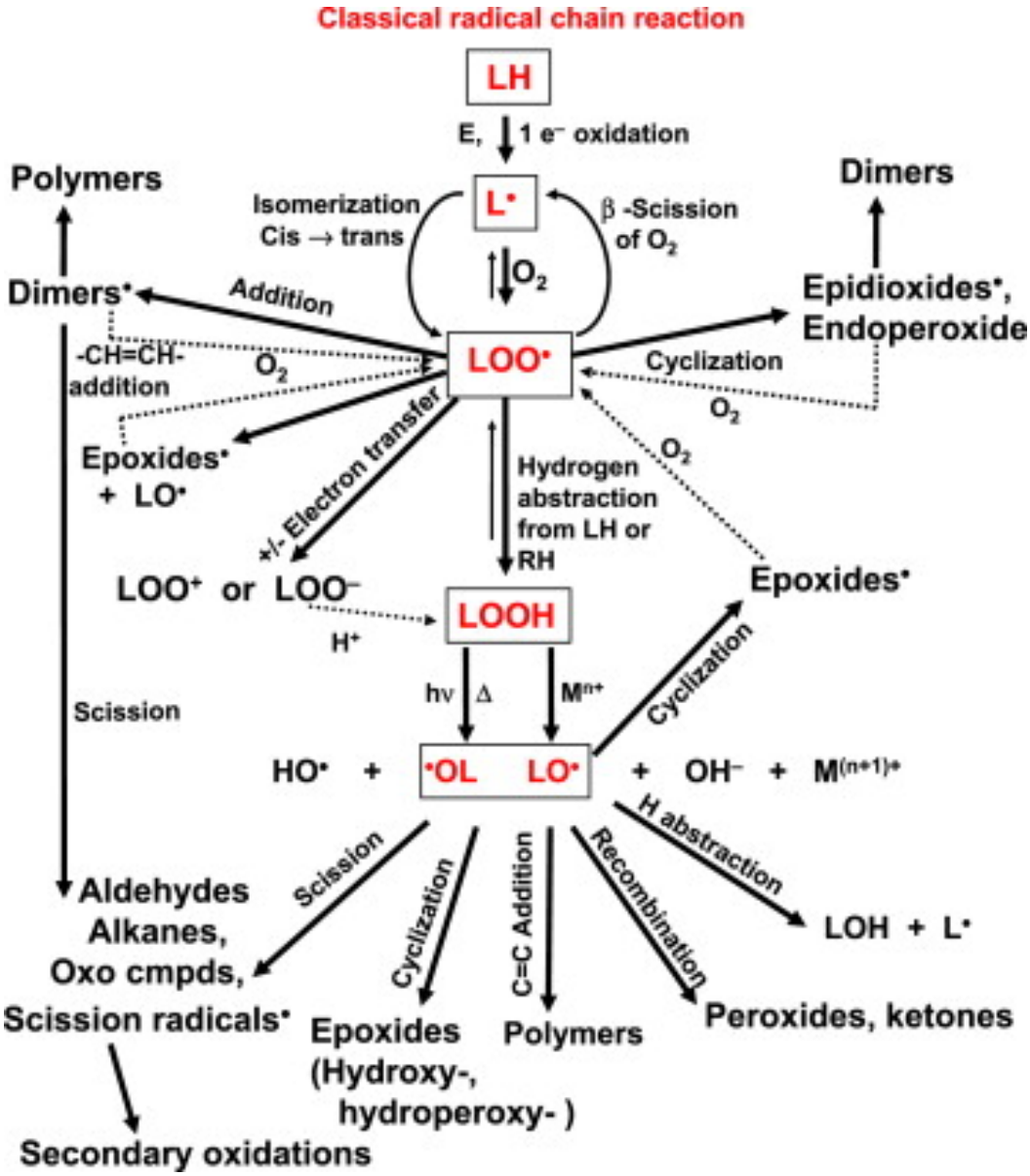


Figure 2-2 Classic lipid peroxidation and the end products (courtesy KM. Schaich [74], reproduced with permission from Elsevier).

So, what is the relationship between ROS and cancer cells? In almost all cancer cell types, the intercellular ROS level is higher than normal cells [75]. However, cancer cells can also express higher levels of antioxidant proteins to maintain the ROS homeostasis and maintain their functionality [76]. In normal cells, ROSs are mainly generated by OXPHOS as by-products; molecular oxygen reacts with electrons

released from electron transport chain (ETC) and turns into superoxide radicals (O_2^-), and releases to the cytoplasm through the mitochondrial permeability transition pore [77]. In the cytoplasm the superoxide can be dismutated into hydroperoxide (H_2O_2) by SOD and further degraded into water by glutathione peroxidase (GPx). It can also react with another hydroperoxide and produce a more active hydroxyl radical ($^{\cdot}OH$) and cause damage to lipids, DNAs, and proteins [75].

The main reason for higher ROS levels in cancer cells than normal cells is their higher metabolic level. The rapid proliferation and growth of cancer cells requires more energy production than normal cells. For this reason, more NADH are generated from enhanced energetic metabolism, meaning more reduced cites in the ETC complex and higher chances of electron leakages during the ETC process, resulting in more ROS productions [78, 79]. In mitochondrial theory of cancer, mutation of nuclear or mitochondrial genes encoding components of the mitochondrial ETC can lead to partial inhibition of electron transfer [80]. This will cause the electrons to accumulate at a place and make them easier to be captured by molecular oxygen and form superoxide radicals, resulting in the high cellular ROS level in cancer cells.

2.2.3.3. Cytochrome P450

Another important source of endogenous ROS is through Cytochrome P450 (CYP)-dependent microsomal electron transport system. The CYP-dependent microsomal electron transport system can catalyse oxidation of a substrate to a monooxygenated substrate [81]. During the CYP monooxygenase cycle, two electrons will move from NADPH to the CYP heme prosthetic group. Later, if the transfer of an oxygen atom to a substrate is not tightly coupled to NADPH utilisation, the electrons will be deployed to CYP–oxygen complexes instead of the substrate, then the complexes dissociate and release ROSs like O_2^-, H_2O_2, and $^{\cdot}OH$ [82]. In many cancer types, including bladder cancer, CYP genes are overexpressed [81, 83]. One possible explanation is that they

are activated by environmental toxins, however the detoxication may lose control due to the mutation of CYP genes and produce more ROS. Moreover, the CYP enzymes are involved in the catalysation of bioactivation of chemical procarcinogens into reactive carcinogens, thus promoted the carcinogenesis [81].

Cytochrome p450, along with Alcohol dehydrogenase (ADH) and aldehyde dehydrogenase (ALDH), are able to catalyse alkanes into alcohols, aldehydes, and carboxylic acids, as consequent process of turning secondary products of lipid peroxidation into a wide variety of VOCs [84]:

$$\text{Alkane} \xrightarrow{\text{Cytochrome p450}} \text{Alcohol} \xrightarrow{\text{ADH}} \text{Aldehyde} \xrightarrow{\text{ALDH}} \text{Carboxylic Acid} \qquad (1-8)$$

2.2.3.4. Oxidative stress

Besides from endogenous source, exogenous sources like UV light, ionizing radiation, inflammatory cytokines and carcinogens can also stimulate ROS production [85]. DNA damage by ROS attacks are widely accepted as a major cause of cancer [85]. Although there are DNA repair mechanisms to recover the damages, persistent DNA damage may lead to replication errors, activation of oncogenes, inactivation of tumour suppressor genes and genomic instability, thus inducing carcinogenesis [81]. When the ROS production exceeds the clearance of antioxidant, more and more ROS will accumulate within the cells. Such persistent DNA damage is easier to happen, this status is so-called **Oxidative Stress**.

Oxidative stress is a double-edged sword. In normal cells, it can lead to cell damage and carcinogenesis, but it can also stimulate cell activities and gene expression when at a low level, known as hormesis [86]. Meanwhile, in cancer cells, oxidative stress guides metabolic reprogramming and participates in activation of various signal cascades. Many studies have found that higher than normal levels of oxidative stress may contribute to the survival and proliferation of cancer cells [76, 87-90]. However,

contrastingly, ROS-induced apoptosis is a key mechanism of many cancer chemotherapy methods; cancer cells are more vulnerable than normal cells to toxicity induced by oxidative stress. With the addition of exogenous oxidative agents, the higher ROS basal level of cancer cells makes them closer to reach the threshold of cytotoxicity. This results in cancer cell cycle arrest, senescence, and apoptosis [81, 88]. The mutation of genes coding antioxidant enzymes in cancer cells may reduce anti-oxidation effectiveness, which also contributes to cancer cells intolerance of oxidative stress induced by chemotherapy [88, 91].

2.2.3.5. Other mechanisms

The rapid growing cancer cells require biosynthesis of a large amount of proteins, lipid, and nucleic acid. As mentioned previously, upregulated aerobic glycolysis and glutaminolysis can provide materials for those biosynthesises, thereby apparently reducing the concentration of other materials required for such biosynthesis. For example, pyrimidine biosynthesis is largely promoted for providing the materials for RNA and DNA synthesis [92, 93], while its intermediate products like ureidosuccinic acid are largely consumed [94], therefore showing decreased concentration. A similar phenomenon was observed on the purine biosynthesis, whereby the concentration of synthesis product (hypoxanthine) increased and that of the breakdown product (uric acid) decreased [95].

Figure 2-3 shows the systematic sketch of possible pathways of bladder cancer metabolic biomarkers [96]. The red highlighted metabolites indicate a decrease relative to healthy controls while green indicates an increase. Besides from aforementioned shortened TCA cycles, glycolysis, purine and pyrimidine metabolism, and oxidative stress related pathway (Taurine, GSH etc.), there are large groups of upregulated metabolisms of fatty acid, amino acid, and glycerophospholipids. Such upregulated metabolisms contribute to energy alternation, rapid growth and

proliferation, and membrane formation of cancer cells. Notably, some of the metabolites' changes are observed in vitro cell culture and those observed in urine are not volatile therefore are not considered as VOCs biomarkers.

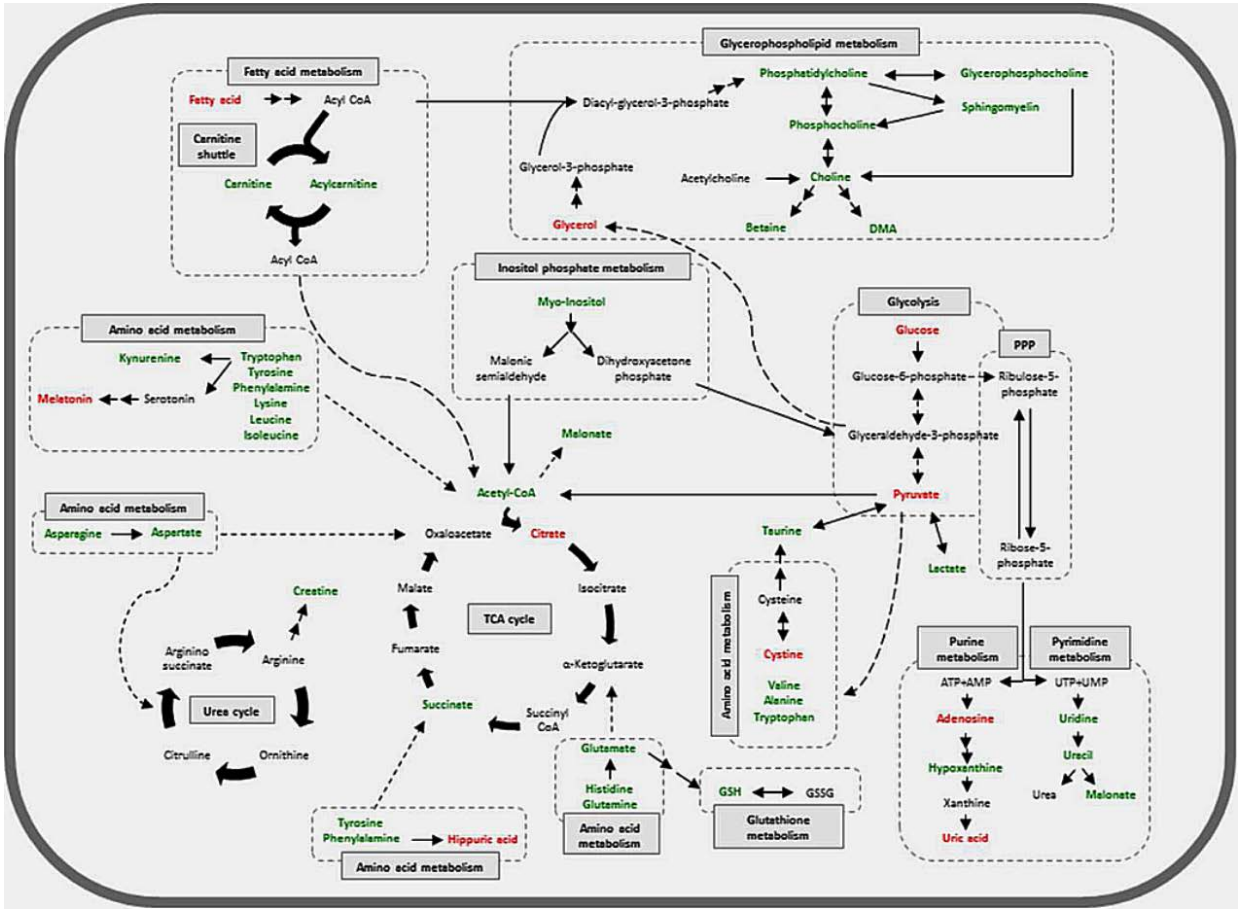


Figure 2-3 Bladder cancer biomarkers and possible metabolism pathways, red highlighted potential biomarkers are upregulated, green highlighted are downregulated. (courtesy of D. Rodrigues et al. [96], reproduced with permission from Elsevier).

Lee et al. revealed the metabolic pathways of biogenic VOCs in lung cancer cells using ^{13}C isotope labelling [97]. By measuring the ^{13}C enrichment ratio during isotopic flux, they found that 2-pentadecanone, one of the most studied cancer VOC biomarkers found in breast, lung, colon and pancreatic cancer cells, is produced from the metabolic pathway of glucose to fatty acid synthesis. This pathway is upregulated and possibly used for meeting the requirements for producing a membrane for the rapid proliferation cancer cells. This finding confirmed that increased fatty acid synthesis is one of the reasons that cancer cells produce abnormal levels of VOCs, linking the shortening of the TCA cycle and the lipid biosynthesis with the production of those VOCs.

2.2.4. How do biogenic VOCs generated by cancer cells get into urine?

For bladder cancer, the biogenic VOCs generated from cell activities have two possible ways of entering urine: direct excretion into urine or entering blood and subsequently urine through the kidneys. Because the kidney's filtration process allows molecules with a weight of up to 7k free to pass, basically all VOCs in plasma could eventually appear in urine. The real questions are: how many biogenic VOCs generated from cancer cells can enter the urine and how many of them will release to the urine headspace?

These could be answered by designing an experiment comparing the VOC profiles of bladder cancer tumour tissue, blood headspace and urine headspace of the same patient. However, no such experimental results have been published yet. The closest approach was carried out on mice models of lung cancer in 2012; Hanai et al. analysed the headspace of the culture medium of A549 lung cancer cell line and urine of A549 implanted mice and found seven VOCs increased in both culture medium and implanted mice urine compared to controls [98]. It is reasonable that all seven VOCs

they found had higher concentrations in urine than in culture medium, as urine comes from ultrafiltration and the concentration of plasma. However, there is still no experimental support for VOCs biomarkers distribution between bladder cancer cell line and biofluid.

From a thermodynamic point of view, the distribution of volatile solute between liquid solvent and its headspace follows Henry's law:

$$k_H = {}^{C}/_{P_i} \qquad (1-9)$$

Where k_H (mol L^{-1} atm^{-1}) is the Henry's law constant, or distribution coefficient, C (mol L^{-1}) is the concentration of the solute in the solution, and the P_i (atm) is the partial pressure of solute in the headspace.

The Henry's law constant is temperature dependent; it can be determined by a static headspace method by measuring the concentration of solute in the headspace of known concentration solution. Although common VOCs' k_H in water is available to check on the NIST online database and several studies had gathered experimental data for some chemicals' k_H in blood, a database for VOCs' k_H in urine is still lacking.

Wilson et al. measured the Henry's law constant of o-xylene and trichloroethylene in water, blood and urine [99]. Both o-xylene and trichloroethylene have low solubility in aqueous solvents and show lower k_H in urine than in plasma and whole blood but show slightly higher k_H than in water. The authors hypothesised that the chemicals may have reactions with contents in blood, thus increasing the apparent solubilities. Such results suggest that VOCs solubilities in urine may be closer to those in water than in blood. Again, this is only about two VOCs' solubilities in urine, more experimental data is urgently needed for the development of urinary VOCs studies.

In summary, the VOCs generated from bladder cancer cells can enter urine directly,

while those entering the blood stream may further degenerate in the liver by Cytochrome p450 and turn into different molecules. Those VOCs that are smaller than 7k MW can pass through the glomerular filtration freely and mix with those directly excreted by cancer cells. The solubilities of VOCs in urine may be lower than in blood, more volatile molecules can be released from urine thus increasing the detectability.

2.2.5. How do biogenic VOCs cancer biomarkers help cancer grading and staging?

As tumour growth is a dynamic process, the oxygen level within the tumours change dynamically as well. During the rapid growth of the tumour, the tumour tissue outgrows the blood supply, therefore resulting in tissue hypoxia. As mentioned previously, the hypoxia may cause the Warburg effect and produce characteristic VOCs. Meanwhile, the cancer cells can also produce various cell factors to promote the new blood vessel growth (angiogenesis). The newly grown blood vessels are chaotic and irregular therefore, at this stage, there are hypoxic and anoxic regions existing at the same time in the tumour. As we known, the reperfusion after myocardial infarction or cerebral ischaemia can cause the generation of ROS. Therefore, we can also hypothesise that reperfusion of the tumour tissue can also produce ROS and its characteristic VOCs.

Different studies have shown that cancer stem cells (CSCs) display molecular and functional heterogeneity, even at the early stage within the primary tumour. Only a few cell clones present the ability of becoming metastasis-initiating cells. Therefore, the functional heterogeneity of CSCs might require them to use different types of metabolism. For example, Pascual et al. found high expression of lipid metabolism genes of human oral carcinomas cells, these are especially able to initiate metastasis [100]. Further studies showed those cells are upregulated in lipid uptake and transport

fatty acid β- and α-oxidation, lipid biosynthesis and intracellular lipid storage genes and are able to increase the lipid intake from the extracellular environment, allowing cells to produce extra ATP through lipid β-oxidation.

Different metabolic phenotypes may indicate the outcomes of the cancer, and such differences could be detected by urinary metabolites profiles. Alberice et al. studied 48 urine samples from bladder cancer patients and linked the recurrence rate with the existence of some urinary metabolites [95]. They found a total of 27 metabolite features have significant differences between patient groups. Among the 27 metabolites, tryptophan and N-acetyltryptophan were found to be significantly increased in low-risk patients. Meanwhile, leucine, methylated derivatives of lysine and histidine were observed to be increased in patients with recurrence and in high-risk patients relative to low-risk.

In summary, tumour growth and metastasis can alter the tumour metabolism. Therefore, as an indicator of cell metabolism, the VOCs also have potential uses in tumour staging and grading.

3. Development of Fluorescence Cross-Response Sensor Array System

The aim of this part of work is to develop a system that is able to:

1. Vaporise VOCs from liquid sample
2. Guide the vapours to interact with a fluorescence sensory array
3. Detect the differences in the fluorescence signal of the sensory array before and after such interaction

Afterwards, the differences of fluorescence signal will be used to build a discrimination model to classify the testing samples and to predict unknown samples' classification. Successful achievement of those goals can provide a potential platform for a non-invasive urinary diagnosis tool for bladder cancer.

In this chapter, the development of both hardware and software of the system will be discussed. As a key component of the detection method, the manufacturing methods of fluorescence VOCs sensor array are also described.

3.1. Introduction

According to the EU's definition, VOCs are "any organic compound, having 293.15 K and a vapour pressure of 0.01 kPa or more" [101]. VOCs are present everywhere in our daily life, bringing us both pleasant and foul fragrances. In recent years, "VOCs" has been a hot key word in environmental topics. People are raising concerns about VOCs with regards to indoor air qualities, atmospheric pollution, volcano eruptions in Indonesia or Iceland, and related healthcare and medical effects. As a potential pathogenic factor, VOCs have already been linked with respiratory diseases, allergy, and irritations [102]. More efforts are committed to find out the long-term health risks of human expose to indoor VOCs.

Recently, using biogenic VOCs as a cancer diagnostic marker has drawn widespread attention in academia. VOC analysis has been proved to have potential in the diagnosis of a wide range of cancers, including breast [103-105], thyroid [106], colorectal [107-112], esophagogastric [113], gastric [114, 115], lung [116-121], pancreatic [122], bladder [123-126], and prostate [127, 128] cancers. This is from a variety of sources including exhaled breath [103-108, 113-120, 127], urine [111, 112, 122-126, 128], sweat [121], and stool [109, 110].

As discussed in Chapter 2, the sources of biogenic VOCs include: lipid peroxidation; Warburg effect (cancer cells prefer accelerated glycolysis rather than oxidative phosphorylation); xenobiotic metabolism (drugs, smoking, diet, inhaled air pollutants); other metabolic alterations like increasing fatty acid oxidation and glutaminolysis.

In this part, we will introduce the basic principle of the optic VOCs sensory technique, how the sensitive material emits fluorescence, how it will behave when exposed to VOCs vapours, and how those responses are linked with a unique 'fingerprint' of specific VOCs.

3.1.1. Gouterman Four Orbital Model of Porphyrins

Porphine is a macrocyclic compound that comprises four pyrroles and four methane groups. The four carbons between each pyrrole ring are called *meso* carbon, while the other eight available substitution sites on the pyrrole rings are called peripheral carbon (β-carbon), as shown in Figure 3-1:

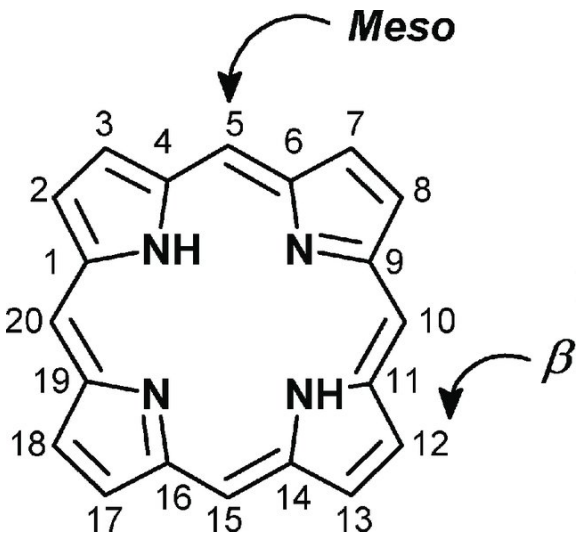


Figure 3-1 Porphine, the simplest porphyrin, showing meso- and β- carbons, which are most common locations for substituents.

All the substitution products of porphine are collectively referred to as porphyrin. The carbon and nitrogen on the porphine ring are in sp^2 hybridization. Since the porphine ring is a plain structure, the remaining un-hybridized p orbitals is distributed in parallel on each side of porphine plane and forms a 24-atom circular delocalized π bond. This conjugated system increases the stability of the macrocyclic compound and gives it colour.

Gouterman's four-orbital model was first proposed by Martin Gouterman in the 1960s to explain the absorption spectra of porphyrins. In molecular orbital theory, the electrons are distributed around the whole molecule, while the bonding between atoms is determined by the energy level of molecular orbitals, which is always a pair of the lowest energy bonding orbital and a high energy antibonding orbital. When porphyrin molecule absorbs light, the electron from π-bonding orbital receives energy and be excited to an antibonding π* orbital, this is called $\pi \rightarrow \pi^*$ electronic transitions.

More specifically, the electronic transition only happens when the energy of a photon matches the energy gap between the orbitals. The lower energy photons will simply

transmit without absorption. Therefore, the smallest energy gap between the orbitals determines the least energy requirement for a photon to be absorbed. Such orbitals are called “frontier orbitals”, because they are the frontier of electronic transition occurrences. The frontier orbitals are the **highest-energy occupied molecular orbital (HOMO)** and the **lowest-energy unoccupied molecular orbital (LUMO)**, with and without electrons, respectively. Notably, there is usually more than one HOMO and LUMO, therefore the transitions are not unique. This is also why there are multiple absorption peaks in the spectra.

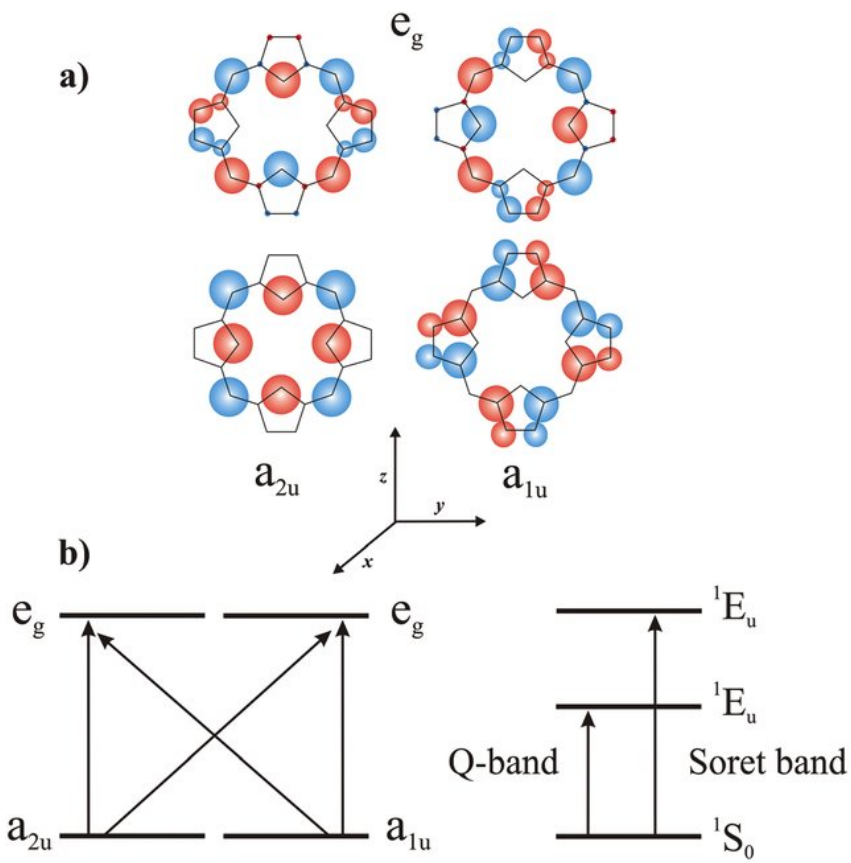


Figure 3-2 Gouterman four-orbital model and HOMO/LUMO of porphyrin. a) Two HOMOs (a_{2u} and a_{1u}) and LUMOs (e_{gx} and e_{gy}) in the four orbitals model. b) Left: Four electronic transitions of porphyrin molecular orbitals: $a_{2u}\rightarrow e_{gy}$, $a_{1u}\rightarrow e_{gy}$, $a_{2u}\rightarrow e_{gx}$ and $a_{1u}\rightarrow e_{gx}$. Right: States and transitions diagram of porphyrins, and their correspondences with UV/Vis abs bands. (Reproduced with permission from Namuangruk et al. [129], with permission from The Royal Society of Chemistry)

In Gouterman's four-orbital model, there are two HOMOs and LUMOs (see Figure 3-2a), namely a_{2u} (HOMO-1), a_{1u} (HOMO), e_{gx} (LUMO), e_{gy} (LUMO+1). Hence, four combinations of transitions may occur: two from HOMOs to highest LUMO+1 ($a_{2u}\rightarrow e_{gy}$ and $a_{1u}\rightarrow e_{gy}$), giving rise to the absorption band located at the lower wavelength: the Soret band. Also, two from HOMOs to lower LUMO ($a_{2u}\rightarrow e_{gx}$ and $a_{1u}\rightarrow e_{gx}$), giving rise to the weaker absorption band located at the longer wavelength: the Q bands (Figure 3-2.b).

Orbitals are used for describing the location and energy of individual electrons, while states are used for describing the total energy and overall electron configuration of the whole molecule. The **ground state** $^{1}\boldsymbol{S_0}$ is the lowest-energy electron configuration of the molecule (which in our case is when the electrons are occupying the HOMOs). All other electron configurations, when electrons are on the different orbitals (for example, when one electron gets excited and transitions to a LUMO), are called **excited states** $^{1}\boldsymbol{E_u}$.

Figure 3-3 shows the typical UV/Vis absorption spectrum of porphyrin: one strong absorption band at the lower wavelength region called Soret band and four relatively weaker bands at longer wavelength region called Q bands. The reason for four split Q bands is related to porphyrins' vibrational excitation states, and a symmetry break by two N-H protons.

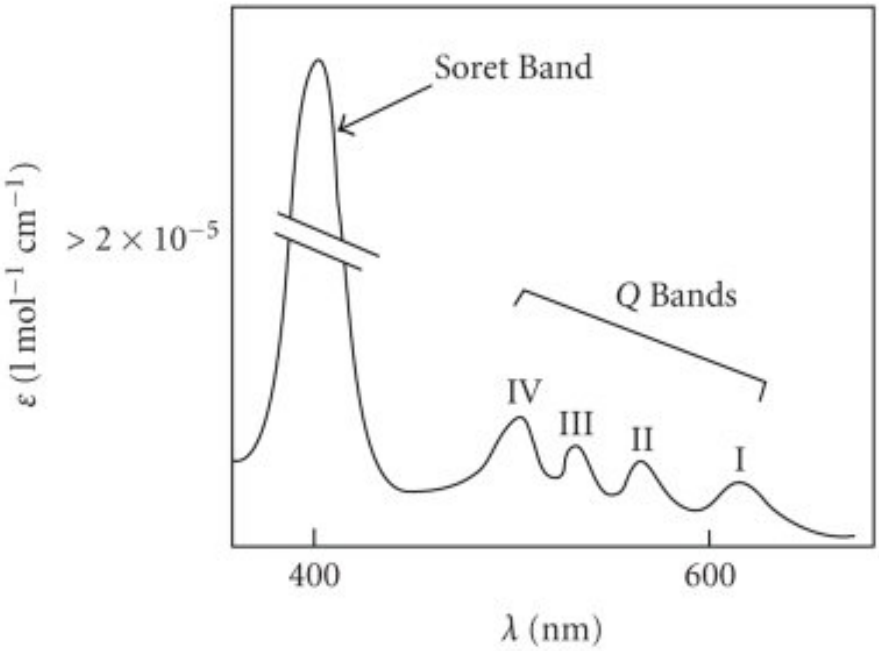


Figure 3-3 Typical Porphyrin UV/Vis absorption spectrum: one Soret band and four Q bands

The excited states of molecules tend to return to lower energy ground states, and there are several ways to “deexcite” excited states. **Non-radiative decay** can turn the photonic energy into heat through molecular vibration and collisions with solvent molecules. Another way is **radiative decay**, including **fluorescence** and **phosphorescence**, which are two ways of emitting light from excited molecules. Both fluorescence and phosphorescence usually emit a longer wavelength light than the excitation light; the major difference is that fluorescence has a faster rate (with shorter lifetime) and higher energy. This difference is because the phosphorescence decay experiences an intermediate state, called triplet excited state, thus leading to slower and longer wavelength emissions of light. Besides from the aforementioned ways, excited molecules can transfer the excitation energy to another ground-state acceptor molecule through **energy transfer**. This is not related to the main topic of this thesis, but it has unique characteristics in cancer diagnosis and treatment, known as photodynamic therapy and fluorescence resonance energy transfer assay.

In general, there are three steps of a molecule to return from excited state to ground state by fluorescence emission. Jablonski diagram (Figure 3-4) illustrates this process.

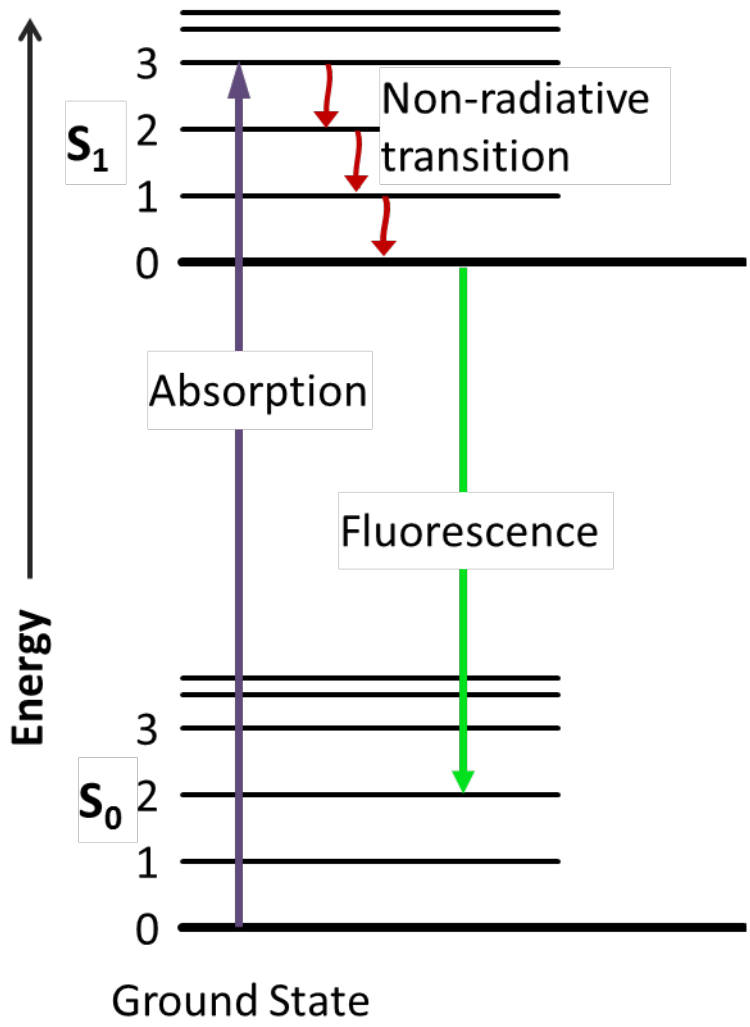


Figure 3-4 Jablonski diagram

First, the excited molecules have a quick (~picoseconds) non-radiative relaxation to a lower energy excited state called the equilibrium excited state. During this process, the molecules at the various vibrational and rotational energy sub levels $S_1(1,2,3\ldots)$ are returned to the lowest vibrational sub-state $S_1(0)$, leading to a portion of energy loss.

Then, the electrons in the high energy orbital return to the lower energy orbital and emit a photon. On the molecular energy level, the molecule quickly returns from equilibrium excited state to the ground state and fluoresce. Similar to the excited state, the ground state also splits into variance vibrational and rotational sub-states, therefore these transitions have multiple destinations, expressed as $S_1(0) \rightarrow S_0(1,2,3\ldots)$.

Finally, the molecules in mixing sub-states $S_0(1,2,3\ldots)$ undergo non-radiative relaxations and return to the lowest energy equilibrium ground state $S_1(0)$.

From that we can see the method of energy loss in each step, leading to a lower energy level of fluorescence than the initial absorption. On the spectrum, this is reflected by a longer wavelength fluorescence band than the longest absorption band.

Specific to tetraphenyl porphyrin (TPP or H_2TPP; also called meso-tetraphenyl porphyrin), the phenyl substituted porphyrin occurs in the meso carbon position (i.e. the 5-, 10-, 15-, 20- carbons in Figure 3-1). The meso substituents may influence the density of the electron cloud of the macrocyclic structure, therefore affecting the energy level of the frontier orbitals of porphyrin. However, usually, the stronger the electron donating substituent, the higher energy level of LUMO and the smaller the HOMO/LUMO gap [130]. Since the phenyl group is a weak electron donating group, the TPP has a slightly redshifted maximum absorption and emission compared to porphin's (See Figure 3-5).

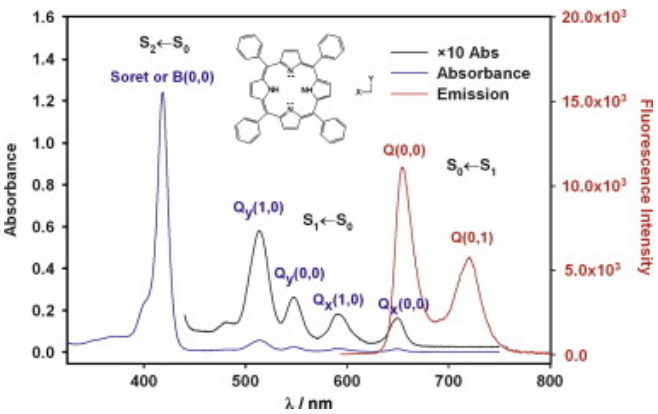


Figure 3-5 UV/Vis absorption and emission (λ_{ex} = 418 nm) spectrum of TPP in toluene (courtesy of Mahesh et al. [131], reproduced with permission from Elsevier)

3.1.2. Interactions between VOC and Fluorescence Sensitive Materials

Previous electronic noses are based on the cross-reactive sensor arrays that depend on the changes of their properties when exposed to the analytes, mainly mass (polymer coated surface acoustic wave/quartz crystal microbalance sensors) and conductivity (conductive polymer, metal oxide sensors). On the level of molecular forces, those sensors depend on the van der Waals interactions between VOC molecules with their sensitive materials (see Figure 3-6).

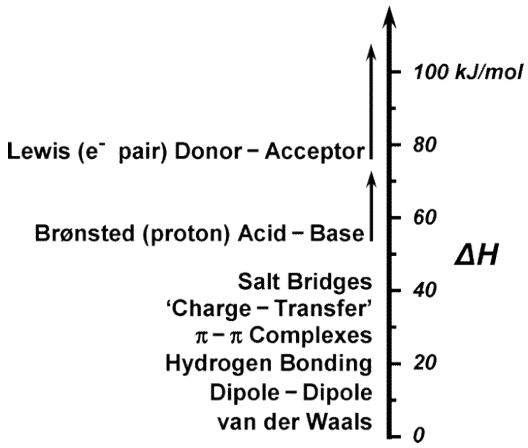


Figure 3-6 Intermolecular interactions on a semiquantitative energy scale. (Reproduced with permission from Janzen et al. [132]. Copyright (2006) American Chemical Society).

For photonics VOC sensor, there are two fundamental requirements: each sensitive material must have a centre to interact strongly with the VOCs and each interaction centre must be coupled to an intense fluorophore. The first requirement means the interactions could be not only the physical absorption and van der Waals interaction like other e-nose sensors, but also π-π interactions and Bronsted and Lewis acid/base

interactions. The second requirement means the sensitive materials should be fluorophores that have large conjugate structures that allow for interaction with the VOC molecules.

In general, three classes of dye fulfil the abovementioned requirements: Lewis acid/base dyes, such as porphyrins and metalloporphyrins; dyes responding to Bronsted acidic or basic pH indicators; dyes with large permanent dipoles, such as zwitterionic solvatochromic dyes.

3.1.2.1. Lewis acidic/basic dyes

When Lewis acid and basic adducts, the electron from Lewis basic HOMO and the LUMO of Lewis acid have very closed energy levels. The interaction makes a pair of new low-energy bonding and high-energy anti-bonding. The non-binding electron pair from Lewis basic then fills into the low-energy bonding and forms the coordinate covalent bond and is therefore in the porphyrin complex. The metal ion is Lewis acid, while the porphyrin ligand is Lewis basic; the donor atoms in metalloporphyrin are the inner nitrogen, usually there are four.

For porphyrins, the large molecular surface area allows enough space to interact with the analyte molecule. For example, in the *meso*-5,10,15,20-tetrakis(4-carboxyphenyl) porphyrin (TCPP, or H_2TCPP), the distance between the nitrogen atom within the porphin ring to the oxygen atom in the carboxyl group is between 0.841nm (not bonded to hydrogen) and 0.858nm (bonded to hydrogen) [133]. Therefore, it has good interaction with molecules that have two hydrogen bond sites spacing roughly 0.85nm, such as β-maltose (the distance between two ends hydroxyl groups is 0.859nm [134]). Although it has the same molecular weight and a similar distance between the two ends of hydroxyl groups (0.837nm), the α-lactose are not able to form stable bonding to TCPP, because the hydroxymethyl group on galactose is on the same side of the two hydroxyl groups, resulting large steric hindrance and preventing bonding.

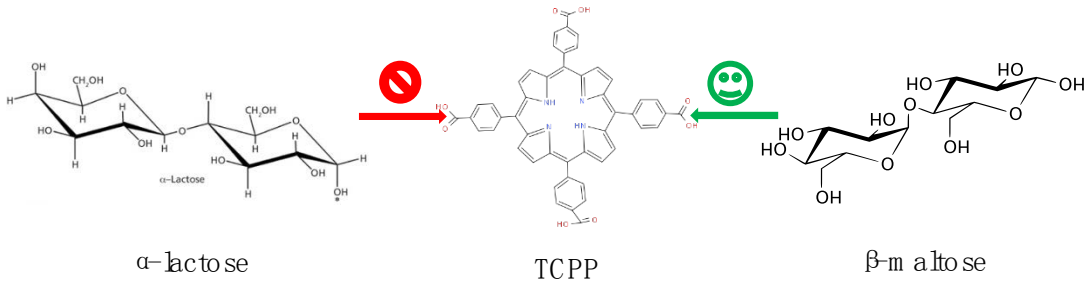


Figure 3-7 TCPP can bond to β-maltose but not α-lactose

For the metalloporphyrin, the metal ion in the centre has the ability to form further coordination with other molecules. Therefore, it has specific bonding to strong ligands. In this situation, the metalloporphyrin is acting as a Lewis acid and it receives an electron pair from the ligand, while the ligand is Lewis base and it donates an electron pair and forms a metalloporphyrin-analyte coordination bond.

3.1.2.2. Bronsted acidic/basic dyes

In Bronsted acidic/basic theory, a chemical that can give a proton is acidic, and chemicals that receive protons are alkaline. The pH indicators are usually weak acids or bases that have two forms of structure. The structures depend on environmental proton strength, for example, methyl orange, a commonly used pH indicator in titrations, has two forms of structure: azo and quinoid structure (Figure 3-8). In basic conditions, methyl orange is in azo structure form, and it is yellow, whereas with a hydrogen ion bond with one of the nitrogens in the nitrogen-nitrogen double bond, it takes a quinoid structure is red due to the fact that a higher degree of delocalisation can lead to longer wavelengths absorption.

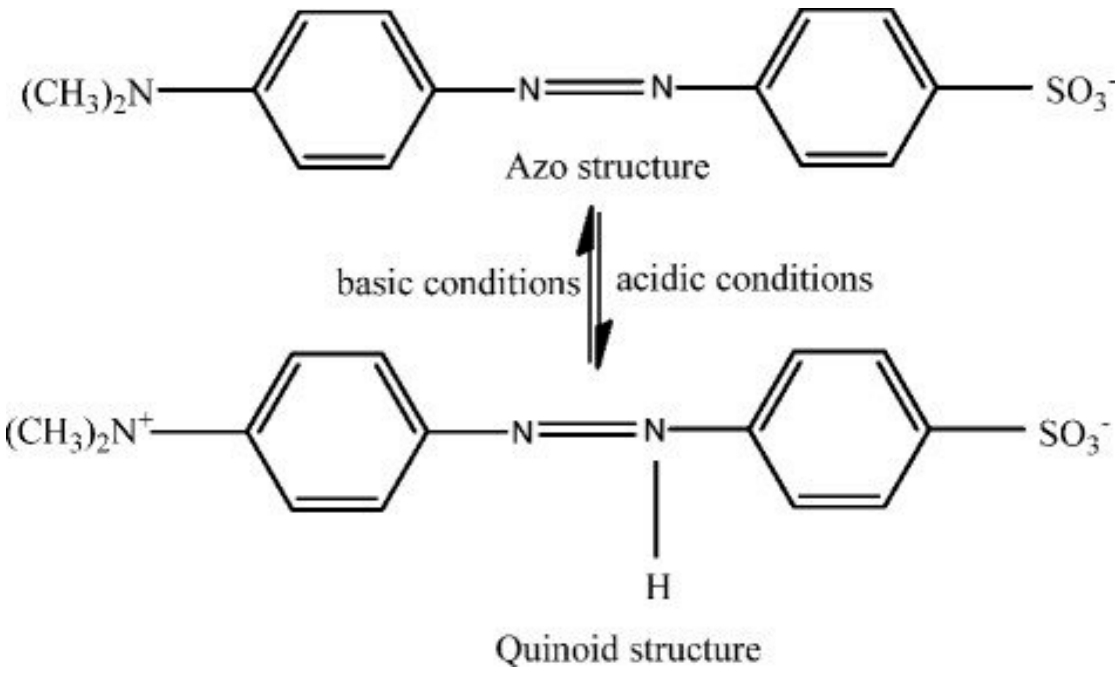


Figure 3-8 Two structures of Methyl orange

The pH indicator can respond to the general acidity/basicity of the environment, it also can respond to the overall polarity of the detecting vapours.

3.1.2.3. Solvatochromic dyes

Solvatochromic dyes will change colour based on the environmental polarity, usually from solvent polarity. Take Reichardt's dye as example, the positive and negative charges in this molecule are separated at the nitrogen end and oxygen end (see Figure 3-9, left), but the whole molecule is neutral. Such molecules are called zwitterions. When the molecule absorbs light, an electron will shift from the HOMO to the LUMO (see Figure 3-9, right), from the oxygen end to the nitrogen end. Therefore, the polar ground state of Reichardt's dye is more stable in the polar solvent, while the nonpolar excited state is more stable in the nonpolar or less polar solvents. In other words, for excitation state the energy in nonpolar solvent needs to be lower than in polar solvent, hence showing redshift in spectrum.

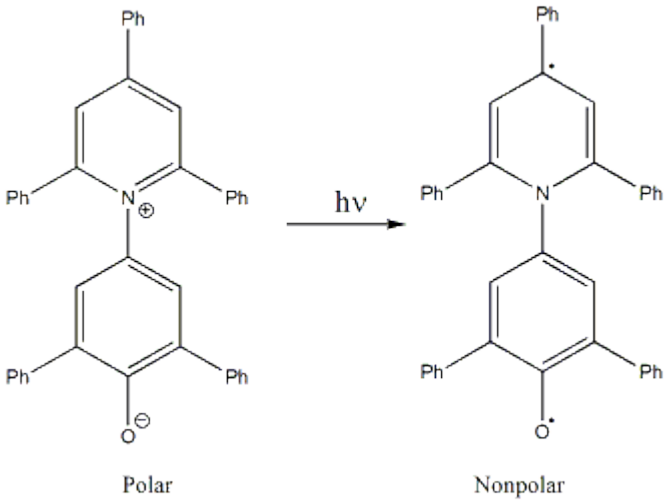


Figure 3-9 Reichardt's dye and light induced charge shift

Different bonding molecules can largely affect the stability of the ground state and excited state of solvatochromic dyes, thus influencing the HOMO-LUMO gap and electronic transitions, and finally expressed as shifts on the spectrum. Comparing to porphyrins, solvatochromic dyes have better affinity to molecules like aromatics and ketones, which will increase the selectivity and sensitivity of the sensory array.

3.2. Materials & Methods

3.2.1. System Overview

In general, the fluorescence sensor array system is made of three modules: **sample processing module**, **responding module**, and **analysing module**. In this chapter, we will only focus on the first two modules, and the third module will be discussed along with the experimental data in the following chapters.

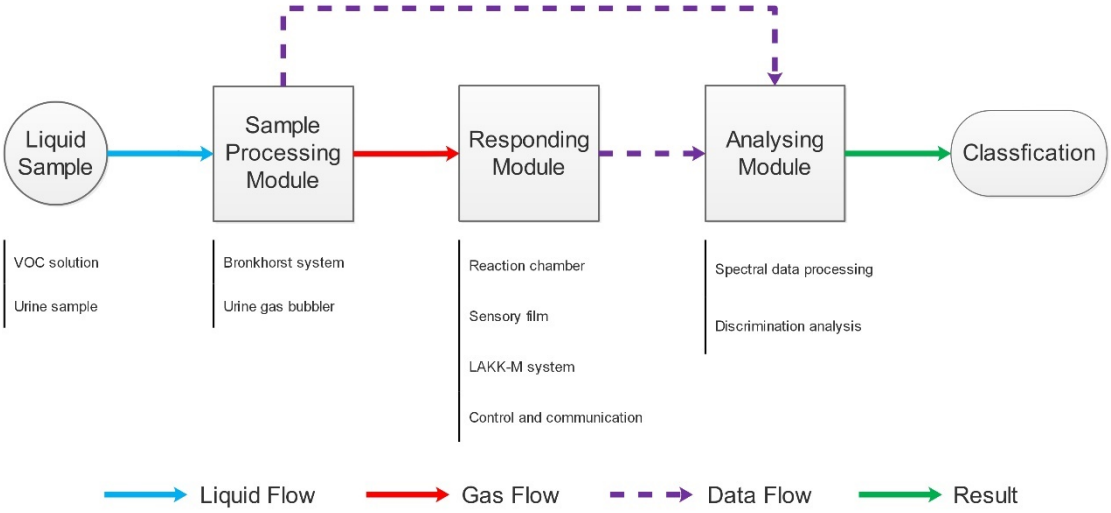


Figure 3-10 Block diagram of fluorescence sensor array-based VOCs detection system

As shown in Figure 3-10, the sample processing module receives the liquid state sample and transforms it into gaseous state. The responding module receives the gas from the sample processing module and generates the spectrum data based on the sensory array's response. Finally, the analysing module receives all the data flows from previous modules and generates a classification result. The specific components for each module may vary depending on different applications, but the system structure as a whole remains the same, as shown in Figure 3-10.

3.2.2. Sample processing module

3.2.2.1. For VOCs tests

The sample processing module (Bronkhorst UK Ltd., UK) for the VOC vapour test consists of three parts: a liquid flowmeter for controlling of the VOC liquid flow; a gas flowmeter for controlling of the carrier gas flow; and the mixer for vaporising the VOC liquid and mixing it with the carrier gas for specific concentration. The sketch of the system is shown in Figure 3-11. A central controller (E-8000, Bronkhorst UK Ltd.,

UK) was used for community and control. On the PC terminal, FlowView and FlowDDE were used for monitoring and controlling the working parameters of the system.

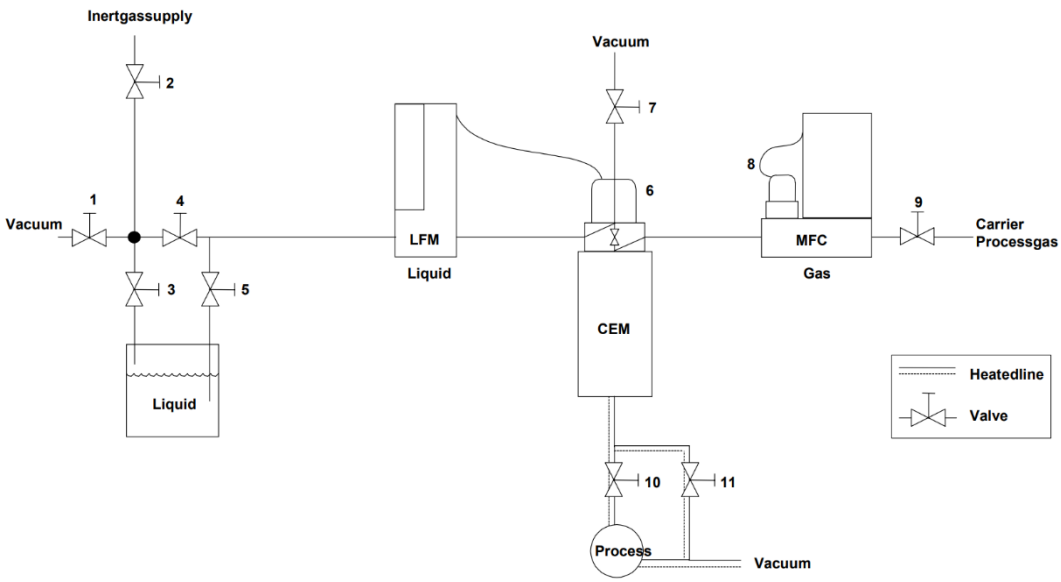


Figure 3-11 VOCs sample processing module and Bronkhorst system sketch figure. LFM: Liquid Flow Metre. CEM: Controlled Evaporator and Mixer. MFC: Mass Flow Controller

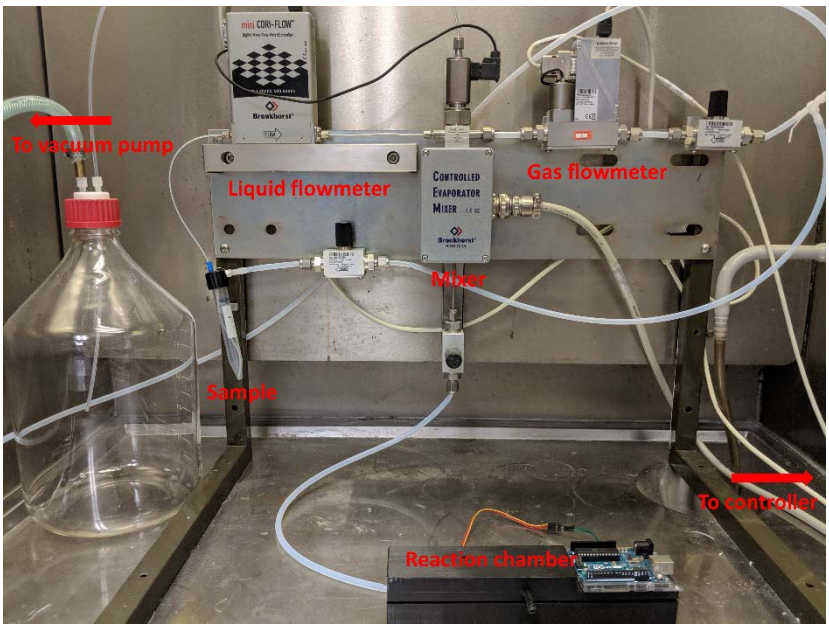


Figure 3-12 Picture of VOCs sample processing module and Bronkhorst system, the reaction chamber

used is Mark. II, detailed sketch was shown in Figure 3-17.

3.2.2.2. For urine tests

The sample processing module for urine testing was developed on the Arduino microcontroller platform. A RS Pro D250 micro diaphragm gas pump (RS Components Ltd, UK) was used for pumping the gas flow; its gas flow speed is adjustable by simply changing the working voltage, the default working voltage is 5V and in this circumstance the flow speed is 380mL/min.

As shown in Figure 3-13, two 15mL centrifuge tubes (Fisher, UK) work as a gas bubbler-safety bottle. The tubing outlet of the pump is under the liquid surface of the urine in a gas bubbler; when the pump is on, the air will be pushed through the urine and increase the gas-liquid contact area and therefore speedup the process. In some cases, the protein from blood and other secretions in the urine is not removed completely. The bubbles generated from the bubbling process are hard to break and may stack up and flow into the reaction chamber. Therefore, to prevent contamination, the safety bottle is used. The safety bottle can provide extra volume to accommodate the urine bubbles, thereby avoiding the bubbles going directly into the reaction chamber. Three-way-valves (1 and 2 in Figure 3-13) are used to control the gas flow in the system. When both valves are switched to the middle, the gas will circulate between the two 15mL tubes and avoid the reaction chamber. The flow speed of the gas pump can be just by supply voltage, under the maximum flow speed (5mL/s), the system can achieve a gas-liquid dynamic balance of the urinary vapor under room-temperature in less than 2 minutes. Then, valves 1 and 2 are switched to a down position, allowing the gas to flow into the reaction chamber and circulate between the two 15mL tubes and the reaction chamber. There are two reasons for doing so: first, it gives the operator time to adjust the output power of the gas pump without facing the risk of contaminating the reaction chamber; second, it allows the gas to be prepared

before contact with the sensory film. The latter is very important to maintain the reproducibility of the experiment by reducing the concentration fluctuations at the beginning of the gas flow.

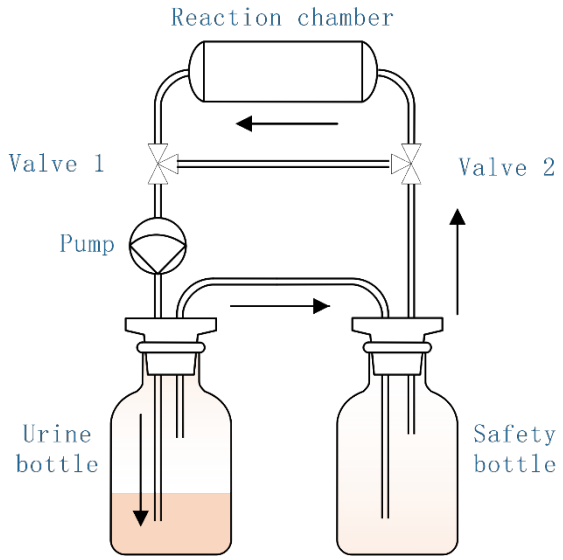


Figure 3-13 Urine sample processing module sketch figure

3.2.3. LAKK-M System

The LAKK-M system (SPE LAZMA, Russia) was not designed for fluorescence sensors detection. However, the LAKK-M has full functionality for a fluorescence spectrometer with 4 fluorescence channels (UV: 365nm, Blue: 450nm, Green: 532nm, and Red: 633nm), and its compact design allows peripheral equipment to cooperate with it freely. The laser sources are coupled with the detection fibre in a 2.5mm diameter optic fibre probe; each tip is separated by around 1mm, allowing the light source to illuminate a 12mm^2 circular area with a 1.5mm distance above the surface of the sensory film. The emitted fluorescence signal is transferred by a receiver optic fibre through the corresponding requisite filter to the CCD spectrophotometer and converted to the electronic signal (Figure 3-14). The spectrophotometer has measurement range of 342nm to 815nm, with a spectral resolution of 0.22nm. LDF 3.1.1 software (SPE LAZMA, Russia) is used for recording and analysing the data. A typical user interface

of LDF software is shown in Figure 3-15. A typical measurement process includes:

1. Initialising the spectrometer by pressing the “setup” button on the fluorescence panel and waiting for the message to turn green
2. Turning on the light source by pressing the physical button on the front panel of the device
3. Waiting until the prompt message “measuring” disappears and the spectra shows up
4. Pressing the “pause” then “save” buttons to save the spectra
5. Turning off the light source by pressing the physical button on the device.

By the default setting, one measurement will take about 30 seconds to complete. Changing the acquisition numbers and exposure time can help to shorten the measurement time but may reduce the signal-to-noise ratio.

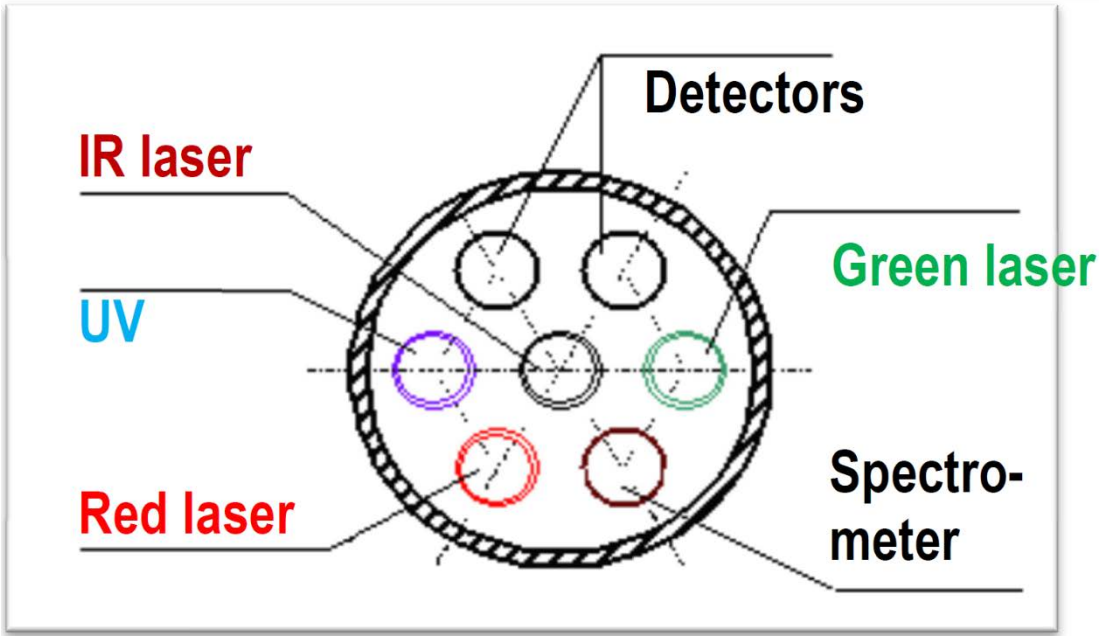


Figure 3-14 Optic fibre layout of LAKK-M probe tip

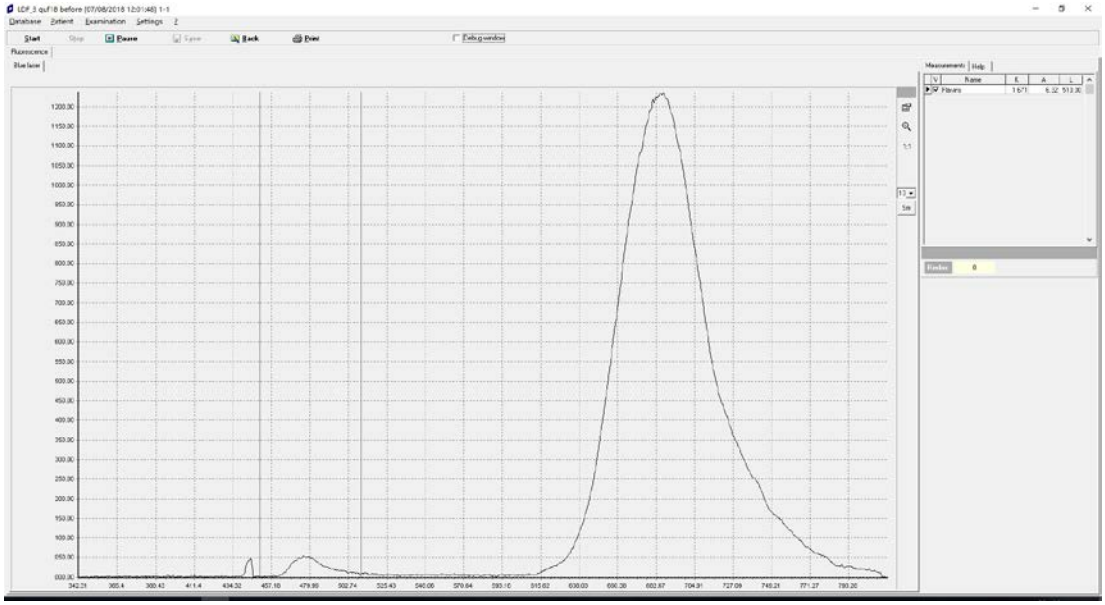


Figure 3-15 Example user interface of LAKK-M PC terminal

3.2.4. Sensory film manufacture

In the literature, there are various manufacturing methods used for making the fluorescence and colorimetric VOC sensor array [132, 135, 136]. In general, those methods are focused on how to deposit the sensitive material onto a solid surface, with chemically and physically stable fixation. At the very beginning of this study, spin coating, as a fast and inexpensive method, was tested.

No.	CAS	Name
(1)	14609-54-2	4,4′,4″,4‴-(Porphine-5,10,15,20-tetrayl)tetrakis(benzoic acid)
(2)	14172-91-9	5,10,15,20-Tetraphenyl-21H,23H-porphine copper(II)
(3)	14074-80-7	5,10,15,20-Tetraphenyl-21H,23H-porphine zinc
(4)	90587-86-3	5,10,15,20-Tetrakis(4-sulfonatophenyl)-21H,23H-porphine manganese(III) chloride
(5)	35218-75-8	4,4′,4″,4‴-(Porphine-5,10,15,20-tetrayl)tetrakis(benzenesulfonic acid)
(6)	917-23-7	meso-Tetraphenylporphyrin
(7)	15086-94-9	Eosin Y
(8)	81-88-9	Rhodamine B
(9)	14320-04-8	Zinc phthalocyanine
(10)	25440-14-6	5,10,15,20-Tetrakis(pentafluorophenyl)porphyrin
(11)	553-12-8	Protoporphyrin IX
(12)	15442-64-5	Protoporphyrin IX zinc(II)
(13)	143-74-8	PhenolRed
(14)	493-52-7	MethylRed
(15)	7385-67-3	Nile Red
(16)	10081-39-7	Reichardt's dye

Table 3-1 Fluorescence dye candidates

Sixteen dye candidates, as shown in Table 3-1, were purchased from Sigma (Sigma-Aldrich, UK). Solvents tetrahydrofuran (THF, CAS: 109-99-9) and dimethylformamide (DMF, CAS: 02/12/1968), film substrate polyvinylidene fluoride (PVDF) membrane (Amersham™ Hybond®) and 10μL pipette tips (Corning® DeckWorks™) were purchased from Sigma (Sigma-Aldrich, UK).

3.2.4.1. Spin coating

An Ossila Spin Coater (Ossila Ltd, UK) was used for preparing the spin-coated sensory film. Cover glass, quartz glass, and silicon wafer were used as substrates. 10mM Eosin-Y THF solution was used as a coating solution.

Before starting, the substrate was cut into a square with a side length of 25mm and cleaned using detergent, deionized water and nitrogen purge. This process was intended to clean up the grease and dust on the surface of the substrate. Then, the substrate was fixed at the centre of the spinner by the claw fixing mechanism. One drop of the coating solution was spotted on the middle of the substrate, then the lid was closed, and the spinner started spinning at 6000 rpm. Most of the coating solution was flung off the side. With the airflow, the solvent volatilised and left the solute deposited on the top surface of the substrate to form an evenly covered film.

The thickness of the film depends on the spin speed, concentration and solvent evaporate rate, as well as solution viscosity, temperature, vapour pressure, and humidity. Roughly speaking, 6000 rpm spinning speed should form a thin film thickness at nanometre level.

The finished spin-coating sensory films were then stored in sealed plastic boxes wrapped with aluminium foil. The fluorescence signal measurements were carried out following the same standard procedure as other measurements in this study.

3.2.4.2. Polyvinylidene fluoride (PVDF) sensory film

Polyvinylidene fluoride (PVDF) film was also used for making the sensory film. PVDF low-fluorescence membranes were purchased from Sigma (Sigma-Aldrich, UK) and cut into small pieces before use. For single-element test, the films were 10mm in diameter, allowing one spot of sensitive material to deposit on. For the multi-element array, the films were cut into 25mm×25mm square size, allowing up to 25 spots deposit on.

All the sensitive materials were prepared in 1mM THF solution. For the single-element test, each time 10 μL of solution was spotted it made a circle around 6mm in diameter

that contained 1×10^{-8} mol of solute. For the multi-element array, a stainless-steel mould was made to assist the precise spotting of the solution. Before spotting, 50mm size PVDF film was mounted on the back of the mould and fixed with aluminium foil. The small through-hole only allowed the pipette tip to insert no more than 2mm to ensure the exact position. 1.5 μL of the solution was the amount used for each spotting, making a circular shape spot around 3mm in diameter. After spotting, the mould was dissembled and cleaned using acetone, water and dried in a fume cupboard overnight. The film reproducibility will be discussed in later chapter, briefly, the doping volume and solubility of the sensitive material can largely affect the final products therefore the manufacture process must be carefully controlled, the reproducibility of the film was measured by normalized coefficient of variance (nCV) which also will be discussed detailly in later chapters.

3.2.5. Reaction chamber design

The reaction chamber is the core component of the system; it is the place where the VOC vapour interacts with the sensory film. Such interaction is indirectly observed through the changes of the fluorescence signal of the sensory elements. Any influencing factors that affect either the VOC-sensor interactions or the process of sensory elements' fluorescence occurrence will affect the final output of the system. Therefore, the reaction chamber must provide a stable physical, chemical and optical environment for the VOC-sensor interaction and its measurement.

3.2.5.1. For VOCs tests

The first design of the reaction chamber (Mark. I) aims to validate the feasibility of the sensory film and the VOC vapour interactions, therefore only single element is allowed

in this design. The setup is simple and consists of: a 1mm diameter meshed ring holder, which can fit tightly on the neck of a 5mL glass volumetric flask (BRAND, Sigma-Aldrich, UK); a corresponding screw cap with a 10 mm through hole to seal the chamber and allow the insertion of the LAKK-M probe. During the experiment, a flask filled with 5mL purified VOC solution is mounted to the bottom of the chamber and each time a 5mm diameter sensory film with single sensory point is placed within the chamber and sealed with the screw cap. By calculation, the distance of the tip of the probe to the surface of the film is set to 6 mm, which allows the light to illuminate a $4mm^2$ circle area on the film.

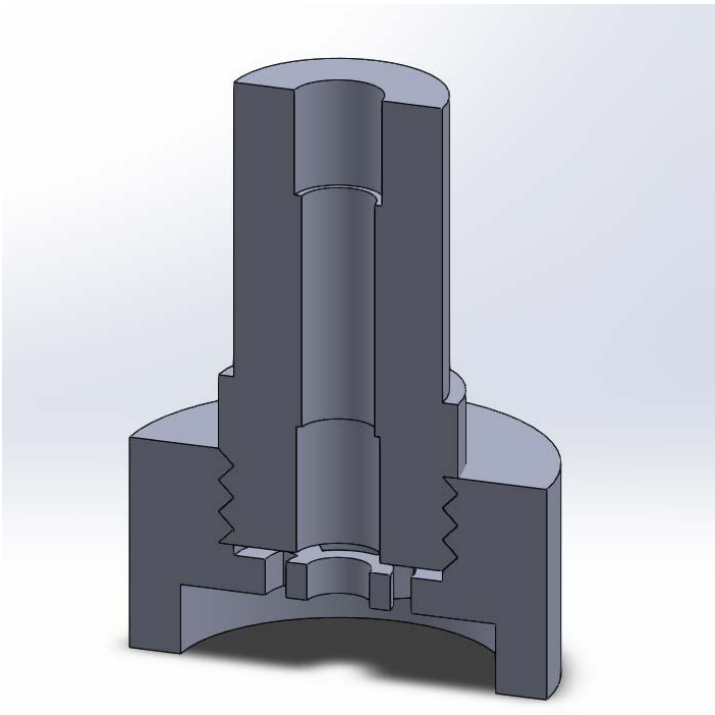


Figure 3-16 Reaction chamber design Mark. I for single element tests

This design has some defects and it is only for proof of principle purposes. Therefore, a better design of a multi element chamber is necessary. Using the same black PLA 3D printing material, the Mark. II has a bigger reaction chamber to accommodate larger multi element film and mechanic components for switching the elements. The Mark.

II has a cube-shaped chamber with an inlet and outlet for the gas flow. Within the chamber, there is a movable rack and corresponding gear that are driven by a servo motor. During the experiment, the rack is first placed on 'initial' position. At this position, the sensory film fixed on the rack is directly exposed to the gas within the chamber. After a certain time, the rack moves to the 'reading' position and detaches the gas chamber. The sensory points on the film then pass through the reading window one after another, which allows the light probe of LAKK-M to read the fluorescence spectrum of each point separately (Figure 3-17). This chamber design was intended to be used in purified urinary VOC biomarkers tests; the results will be discussed in chapter 4.

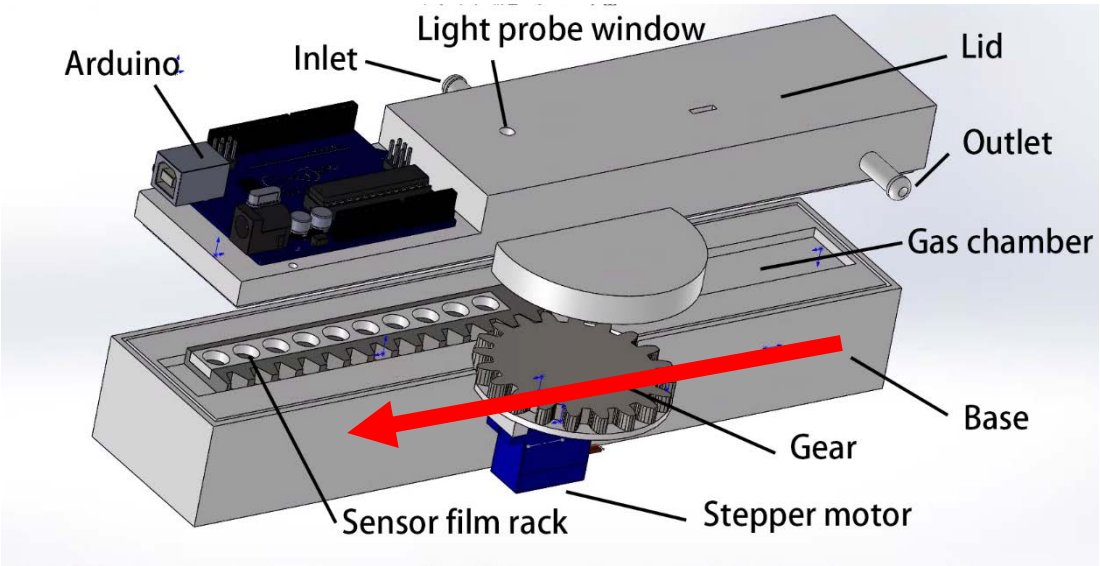


Figure 3-17 Reaction chamber design Mark. II for multi-element sensor array tests, red arrow shows the direction of the film rack movement during the reading process.

3.2.5.2. For urine tests

For the urine test, a smaller and more compact reaction chamber is designed (Mark. III) to adapt to the nature of the low concentration and small amount of urine vapour. In addition to the smaller box-shaped gas chamber, two semi-cylindrical chambers are attached to the inlet and outlet of the gas chamber. It is mainly used for pre-mixing the

urine vapour at the very beginning of each flow to balance the vapour concentration around the film surface and to work as a buffer to prevent the film from contamination from urine spillage. In this design, the sensory film is restricted to a 25mm×25mm square and allowed a maximum of 25 elements as small as 1mm diameter each.

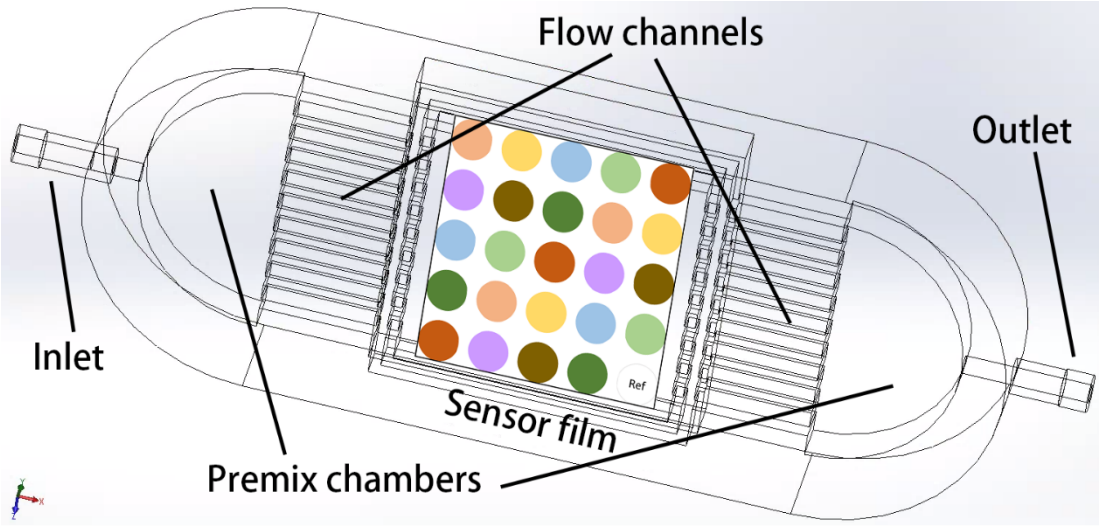


Figure 3-18 Reaction chamber design Mark. III for urine tests

3.2.6. Control and communication components

3.2.6.1. For VOCs test

One Arduino micro-controller was used for controlling the movement of the film platform during multi-element tests. Single element tests reaction chamber does not have mechanical movement parts.

3.2.6.2. For urine test

As shown in Figure 3-19, two Arduino micro-controllers work as the main hosts in the control and communication components. The Arduino 1 controls the motion components. In this experiment, a servo motor (a) was used for the movement of the laser probe on x-axis, while the servo motor (b) was used for the movement of film platform on y-axis. This provides a 2D movement at the point of laser focus, with a resolution of 0.06mm. The temperature and humidity sensor in the reaction chamber,

together with the micro diaphragm gas pump on the main pipeline, were controlled by Arduino 2. This Arduino collected the real time flow speed, humidity, and temperature information through the sensors and transported them to the host PC. The PC terminal (Figure 3-20) was developed using LabVIEW 2016 (National Instruments Corporation, US) with open-source LINX toolkit. This program allows the user to select any of the sensory points to focus on or perform a full array reading and it also can display the flow speed, humidity, and temperature on the screen and save this in a log file.

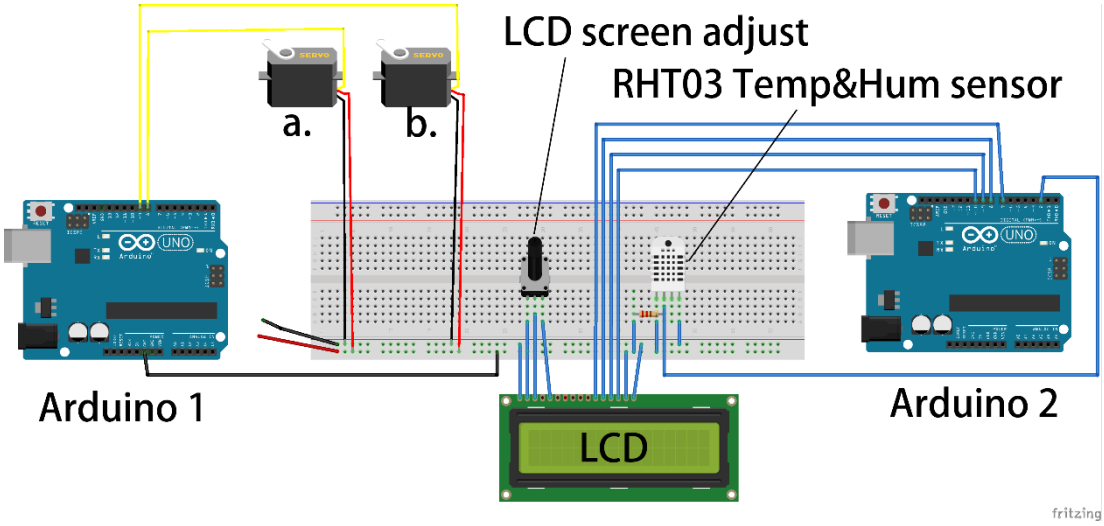


Figure 3-19 Control and communication component for urine tests sketch figure

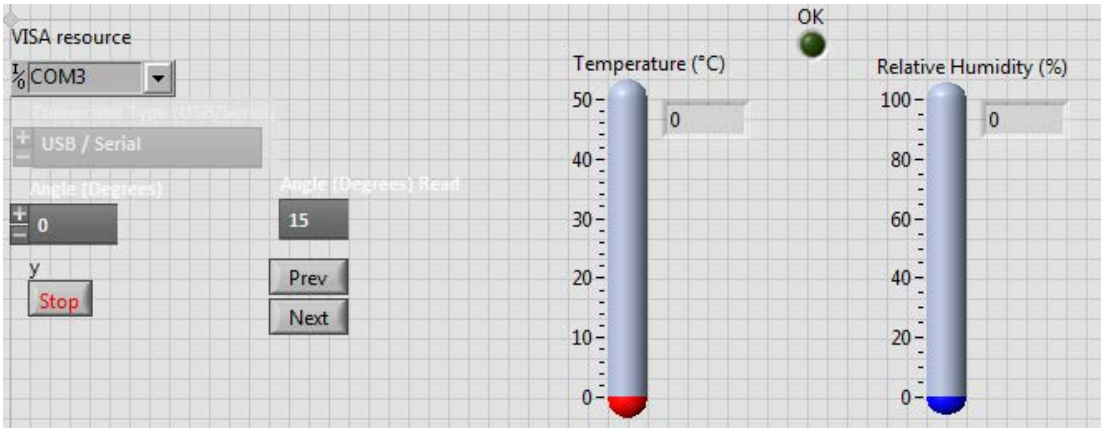


Figure 3-20 Controller PC terminal user interface

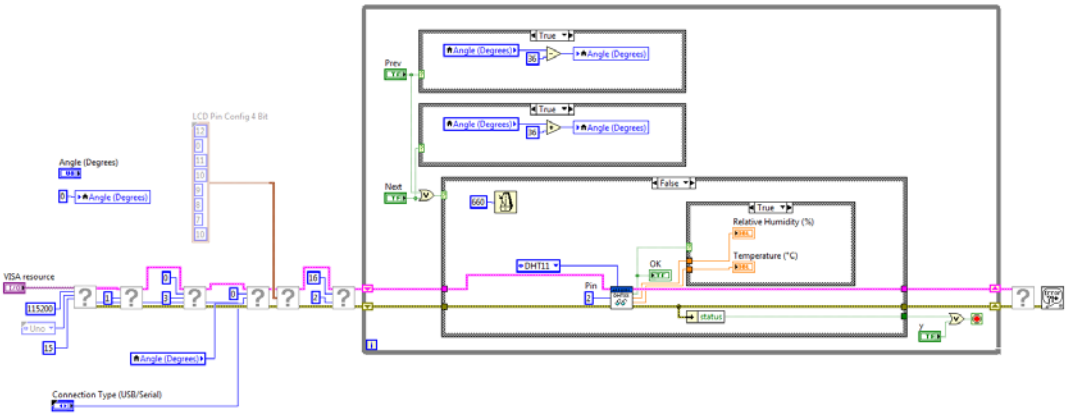


Figure 3-21 Block diagram of PC terminal program in LabVIEW

3.2.7. Software and devices

All the sketching and modelling works were done by SolidWorks 2014 (Dassault Systèmes SE, France). 3D printer modelling of slicing and parameter setup was done by Cura 3.1 (Ultimaker, Netherlands) and PreForm 2 (Formlabs, Inc., US). The 3D printers used were Ultimaker 2 (Ultimaker, Netherlands) and Form 2 (Formlabs, Inc., US). Signal processing and data analysis algorithms were developed using MATLAB R2018b (The MathWorks, Inc., US).

3.3. Results & Discussion

3.3.1. LAKK-M

There are four built-in light sources in LAKK-M device, with slightly different output powers: 1.5mW for UV; 3.5mW for blue; 4.5mW for green; and 5.5mW for red. The optic power of one light source is not stable, it will change over time. Figure 3-22 shows the optic power of the blue light source. At the first 5 seconds, the optic power steeply rises, followed by a gentle decline. After 1500 seconds, the optic power reaches a stable level with about 0.2mW fluctuations.

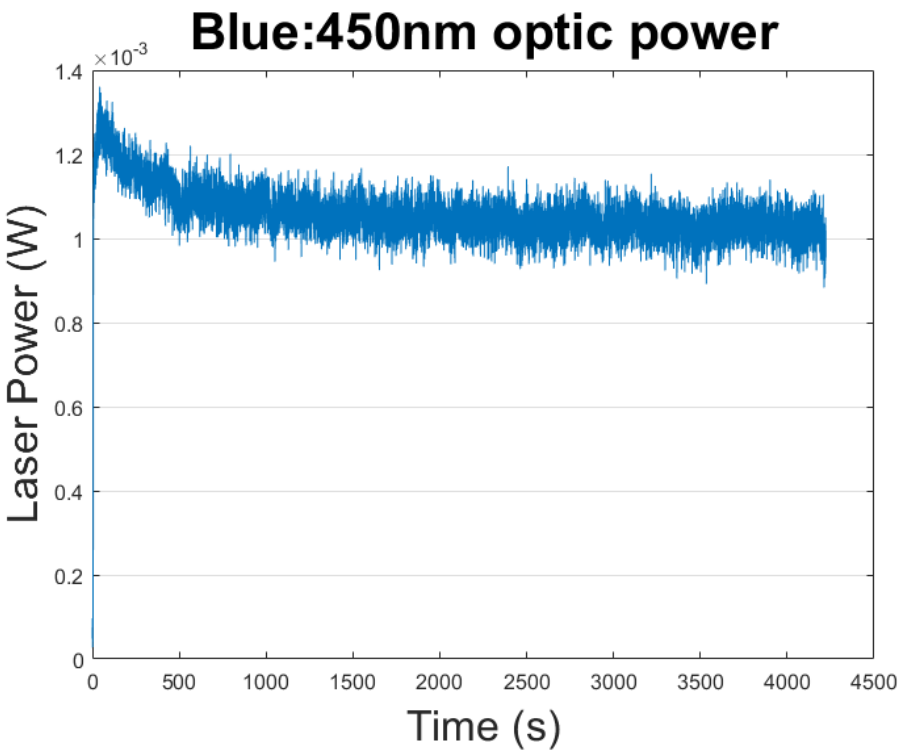


Figure 3-22 Changes over time of optic power of blue light source in LAKK-M

Because of that, the influence of the optic power of the light source is not negligible. When we consider the nature of spectrum, there is some information that links to the optic power of the light source. Firstly, the backscattering peak, which is the excitation light that is returned to the spectrometer due to the backscattering. The intensity of the backscattering light is affected by both the optic power of the incident light and the absorption and scattering coefficient of the illuminated object. The higher the incident light power, the higher backscattering light. Meanwhile, higher absorption or scattering coefficients can lead to lower backscattering light.

3.3.2. Influence of manufacturing material

3.3.2.1. 3D printing materials

For the fused deposition modelling (FDM) 3D printer, the most commonly used

printing materials are acrylonitrile butadiene styrene (ABS) and polylactic acid (PLA). Both ABS and PLA have good mechanical strength and processing performance, the main difference being the temperature.

Name	PLA	ABS
Chemical composition	Polylactic Acid	Acrylonitrile-Butadiene-Styrene copolymer
Nozzle temperature	180-200℃	210-240℃
Glass transition temperature	60-65℃	~105℃
Mechanical property	Higher strength	Better elasticity
Heat plate	Not required	Required (>80℃)
Fumes	None - light	Heavy
Biodegradable	Yes	No
Block nozzle	Easy	Hard

Table 3-2 Comparison of PLA and ABS in 3D printing material properties

During FDM 3D printing, the printing material filament is pushed through a hot (usually over 180°C) metal nozzle where the plastic melts (also called glass transition) and becomes soft and partly fluid-like. The soft plastic is squeezed and stretch to a fine line in the 1/10-millimetre level. The nozzle travels through a pre-set route over the printing platform, while the melted plastics are deposited and stacked layer-by-layer and form the shape of the desired object.

Comparing to PLA, ABS has a higher glass transition temperature. Therefore, ABS require a higher printing temperature but also has better thermal stability. However, for this study, a good thermal stability was not that necessary; the temperature of VOC vapours leaving the mixer was below 50°C, and even consider the device heating the temperature are safe for both materials.

However, because the monomers of the ABS are toxic and there are pungent fumes during the printing, thermal decomposition products may potentially release during the printing. This can cause health risks and, most importantly, remains on the finished

objects and interacts with the sensitive materials, bringing interference to the results. Although there is no evidence of residues, the release of thermal decomposition products of ABS during printing has been confirmed by experiments [137]. Moreover, the shrinkage of the ABS is much higher than PLA, which easily causes deformation and lifting on the edge of the object.

Based on aforementioned factors, PLA is more suitable for our study. Even though PLA has no fluorescence background, pure PLA is nearly transparent, therefore its printing product cannot be light-shielding. After repeat tests with different coloured filaments, a dark black PLA filament was chosen, and the background fluorescence was minimized.

For the stereolithography (SLA) 3D printer, the printing material was pre-mixed resin with photosensitizer. The standard resin we used for Formlabs 2 is made from methacrylated oligomer (75~90%), methacrylated monomer (25~50%), and diphenyl(2,4,6-trimethylbenzoyl) phosphineoxide (1~3%). During SLA printing, a thin layer of resin was used to cover the surface of the printing platform. A 405nm violet laser beam focussed on this layer and incurred photopolymerization, whereby the photosensitizer was activated by the light and released radicals to polymerize the resin monomers and form the desired solid layer. Similar to FDM, the SLA also prints the object in layer-by-layer style, the difference is the SLA machine we used (Formlabs 2) prints upside down (Figure 3-23).

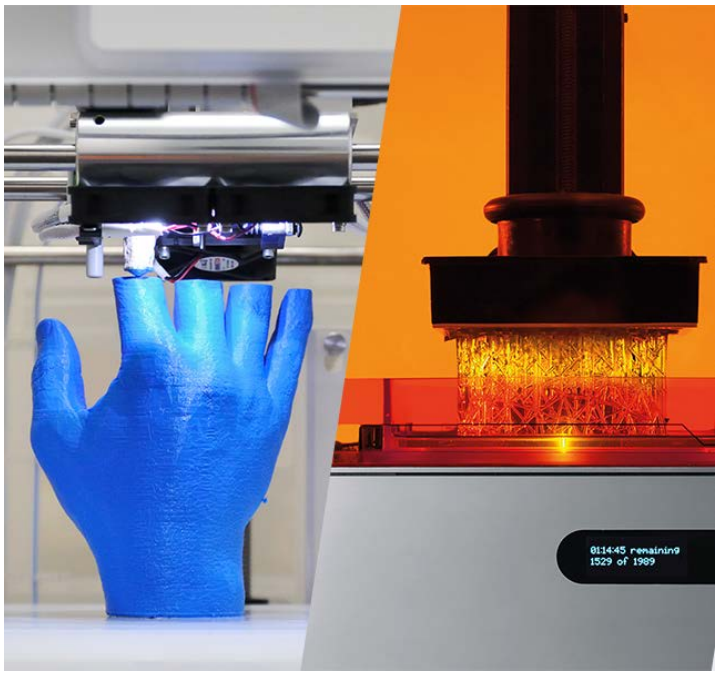


Figure 3-23 3D printing techniques. Left: Fused Deposition Modelling (FDM), Right: Stereolithography (SLA).

Compared to FDM, the biggest advantage of SLA 3D printing is its way higher precision (0.2mm vs. 0.05mm layer thickness), but the costs of the machine and consumables are also 4-5 times higher. Additionally, the objects printed by SLA are solid, whilst most of the FDM objects are printed hollow inside to prevent the wrapping. Therefore, the SLA object has better bending resistance. For the impact resistance, cured resin is hard but brittle. Compared to ABS and PLA, the resin object is slightly more fragile. Similar to other 3D printing materials, the colour agents added in resin have large fluorescence backgrounds. However, because isopropanol was used for washing the residues of resin, this process will bring residues of isopropanol to the finished object. Therefore, it is not suitable for making the reaction chamber. Every finished object was further washed with a large amount of running water and dried at least 48 hours before use.

Due to the above reasons, the SLA machine was used for making the mechanical parts

as well as the parts that do not have contact with the sensory film directly. Meanwhile, the FDM machine was used for making the reaction chamber and the parts that can be illuminated by the laser beam. Figure 3-24 shows the fluorescence background of PLA (black), PLA (white) and Methacrylic resin 3D printing materials. As clearly shown in the figure, the black PLA has a much lower background than other materials.

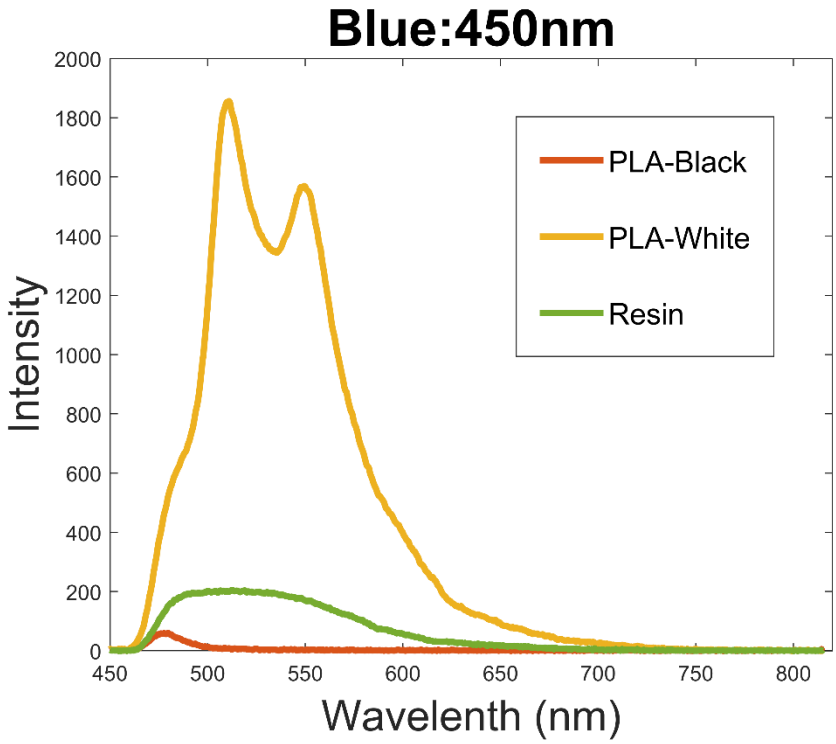


Figure 3-24 Fluorescence background of 3D printing materials

3.3.2.2. Sensitive materials

As described previously, there are three types of fluorescence dyes that suitable for VOCs sensitive material: dyes that respond to Lewis acidity/basicity (e.g., porphyrins and metalloporphyrins), pH indicators that respond to Bronsted acidity/basicity, and dyes with large permanent dipoles that respond to polar VOCs.

Based on this principle and wider literature, 16 fluorescence dyes were chosen for the candidates of the sensitive materials. Chemical structures of the candidates in Table 3-1 are shown in Table 3-3. From the perspective of classification: TCPP (1), TPPS

(5), TPP (6), TFPP (10), and PpIX (11) are free-based porphyrins; CuTPP (2), ZnTPP (3), $(Mg-Cl)TPPS$ (4), and ZnppIX (12) are metalloporphyrins; ZnPc (9) is phthalocyanine metal complex; eosin Y (7), rhodamine B (8), phenol red (13) and methyl red (14) are pH indicators; Reichardt's dye (16) and Nile red (15) are solvatochromic dyes.

From a photochemistry point of view, porphyrin free bases are Lewis basic, and can be used for detection of Lewis acid. Metalloporphyrins and phthalocyanine metal complex are Lewis acids and can be used for detection of Lewis basic. pH indicators can be used for detection of Bronsted acid/basic. Solvatochromic dyes can be used for detection of local polarity changes.

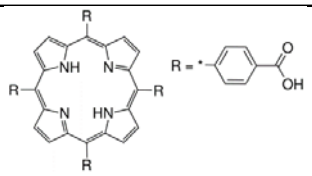
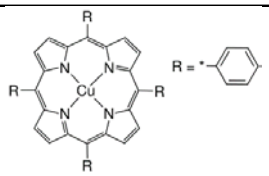
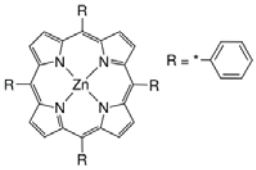
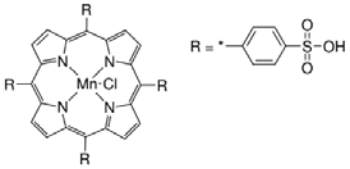
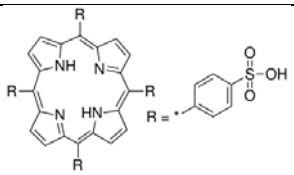
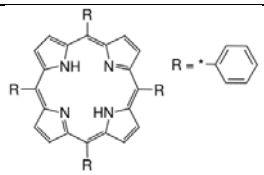
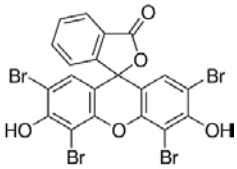
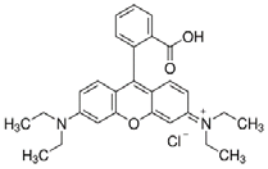
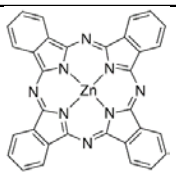
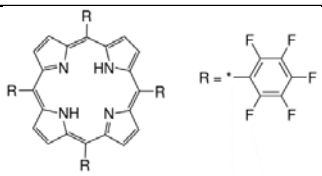
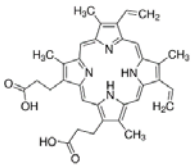
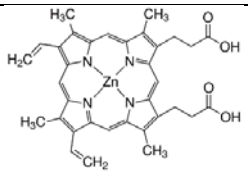
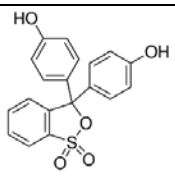
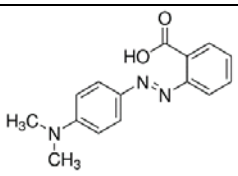
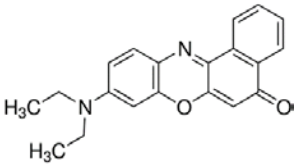
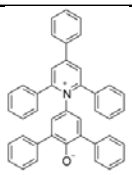
 (1)	 (2)
 (3)	 (4)
 (5)	 (6)
 (7)	 (8)
 (9)	 (10)
 (11)	 (12)
 (13)	 (14)
 (15)	 (16)

Table 3-3 Structure of fluorescence dye candidates

3.3.2.3. Sensory film substrate materials & manufacturing methods

The sensitive materials, especially porphyrin and their derivatives, have higher solvability in the aprotic solvents like Tetrahydrofuran (THF), Dimethylformamide (DMF), and Chloroform. Those solvents have very low viscosity at room temperature: THF is 0.48 cP; DMF is 0.92 cP; and Chloroform is 0.56 cP (as a reference, the viscosity of water is 1 cP). This characteristic brings about difficulties in making the sensor films using the spin-coating technique. Usually, to make spin-coated film from such thin solution, viscous additive is required to reduce the fluidity of the solution. However, this will also thicken the sensory film and affect the contact between the analyte gas and the sensitive material. Both glass and silicon wafer have good affinity to THF, however the viscosity of solution makes it hard to adhere to the substrate and results in defects like bubbles, stipes and pinholes of the film.

There are more methods to create thin solid deposition film like vacuum evaporation, Molecular-beam epitaxy (MBE), and Langmuir–Blodgett (LB) film [138]. These methods can usually create solid films that have thin to low micrometre to nanometre levels, which are ideal for making gas sensor components in mass production.

In this study, the sensory film was made by simply putting spot sensitive material solution on the polymer film. Polyvinylidene fluoride (PVDF) film was chosen because of its low fluorescence background and high affinity to most of the hydrophobic sensitive materials used in this study.

Different sizes of the film were made based on the application and device design. For VOC vapour tests, 10 µL of 1mM sensitive material solution was spotted to make each sensitive point of 10-mm in diameter. For urine tests, each point was made from 1.5 µL of 1mM sensitive material solution, covering 3-mm in diameter in a 5×5 array layout on a 25mm×25mm PVDF film.

PVDF film has been widely used in Western Blot as the transfer membrane for holding proteins separate by gel electrophoresis. For fluorescence visualization Western Blot, the low fluorescence backgrounds PVDF film can help biologists to get a clear, dark background. For our study, such a property is also very helpful to prevent the interference from autofluorescence of the background.

Figure 3-25 shows the fluorescence backgrounds of different substrate materials under blue laser: glass, PVDF and silicon wafer. The spectrum has been normalized with the total area under curve to remove the influence of power variance of the light source. As shown in the figure, PVDF has the lowest background at the main peak region but has some signals around 500~550 nm region.

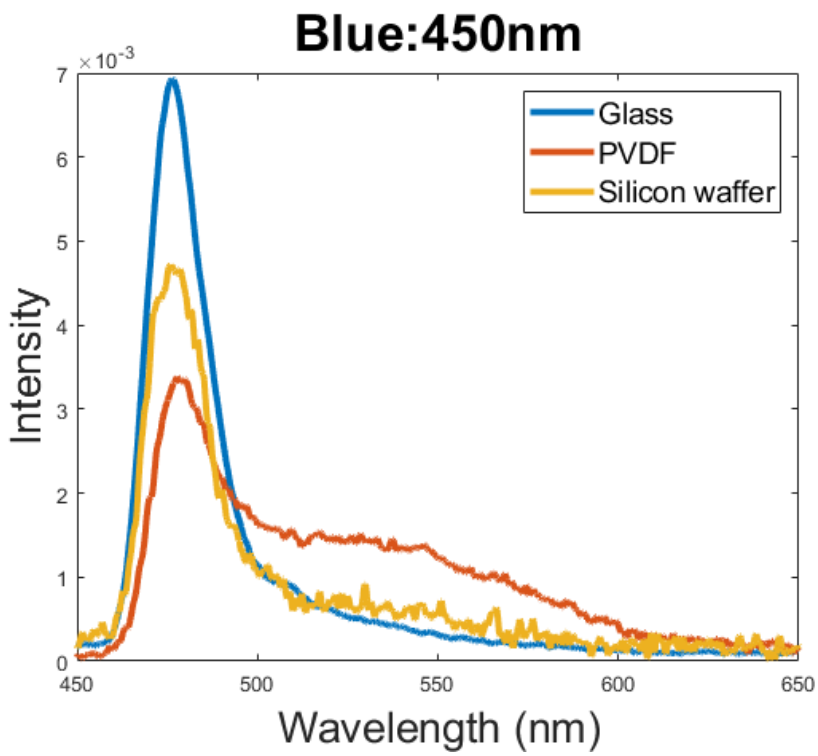


Figure 3-25 Fluorescence background of substrate materials

3.3.3. Selection of sensitive materials based on the single-element test

For single-element tests, there are no actual control components. Most of the operations were performed manually by human operators, with the only data needing to be transported being the spectrum data, which was manually exported from the LAKK-M software. There are some advantages and disadvantages to this. Positively, it is very easy to build and minimise enough so as to only focus on the interactions between VOC vapour and sensory film. However, because there is no ventilation, the VOC vapour can only rely on natural diffusion to enter the reaction chamber, which is inaccurate in both time and concentration. To address this issue, exposure of the vapour during the single-element test was performed for a longer time to ensure the sensitive materials interacted fully and completely with saturated VOC vapour. Safety is another concern. Every time a sensory film is changed, the whole chamber needs to be disassembled. Unavoidably, VOC vapour will release to the surrounding environment, therefore operators must wear full safety equipment including organic vapour respirator, gloves, safety glasses and chemical resistant lab coats, and all operations must be performed in a vented cupboard.

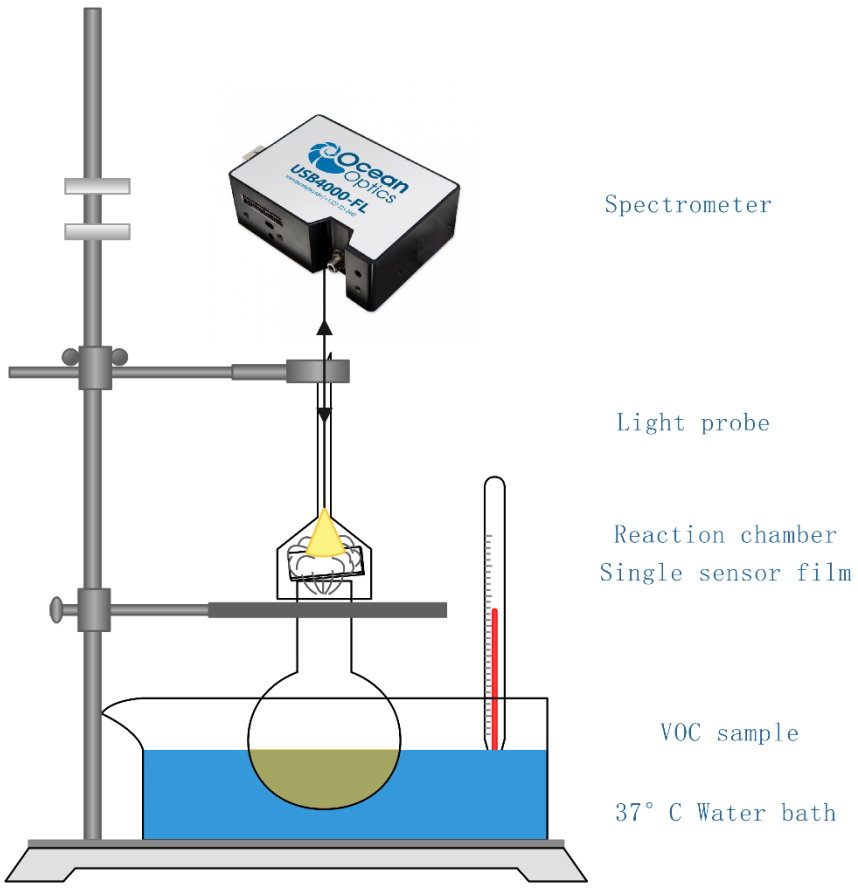
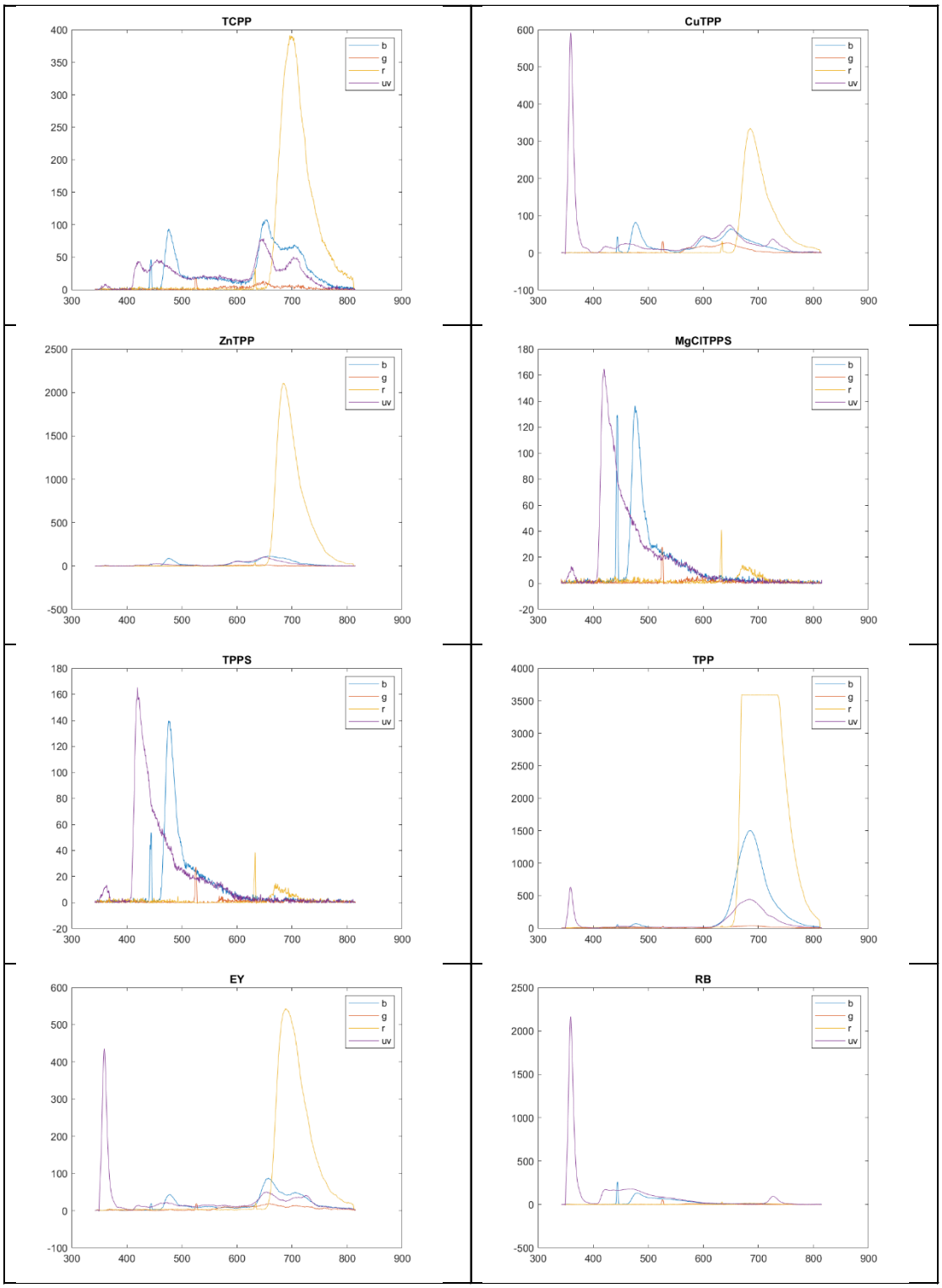


Figure 3-26 Single-element test

Although single-element tests have several cons, they still can perform a quick and simplified test for verifying the sensitive material candidates and checking their responses to various analytes like purified VOCs or urine samples.


TCPP
CuTPP
ZnTPP
MgClTPPS
TPPS
TPP
EY
RB
b
g
r
uv

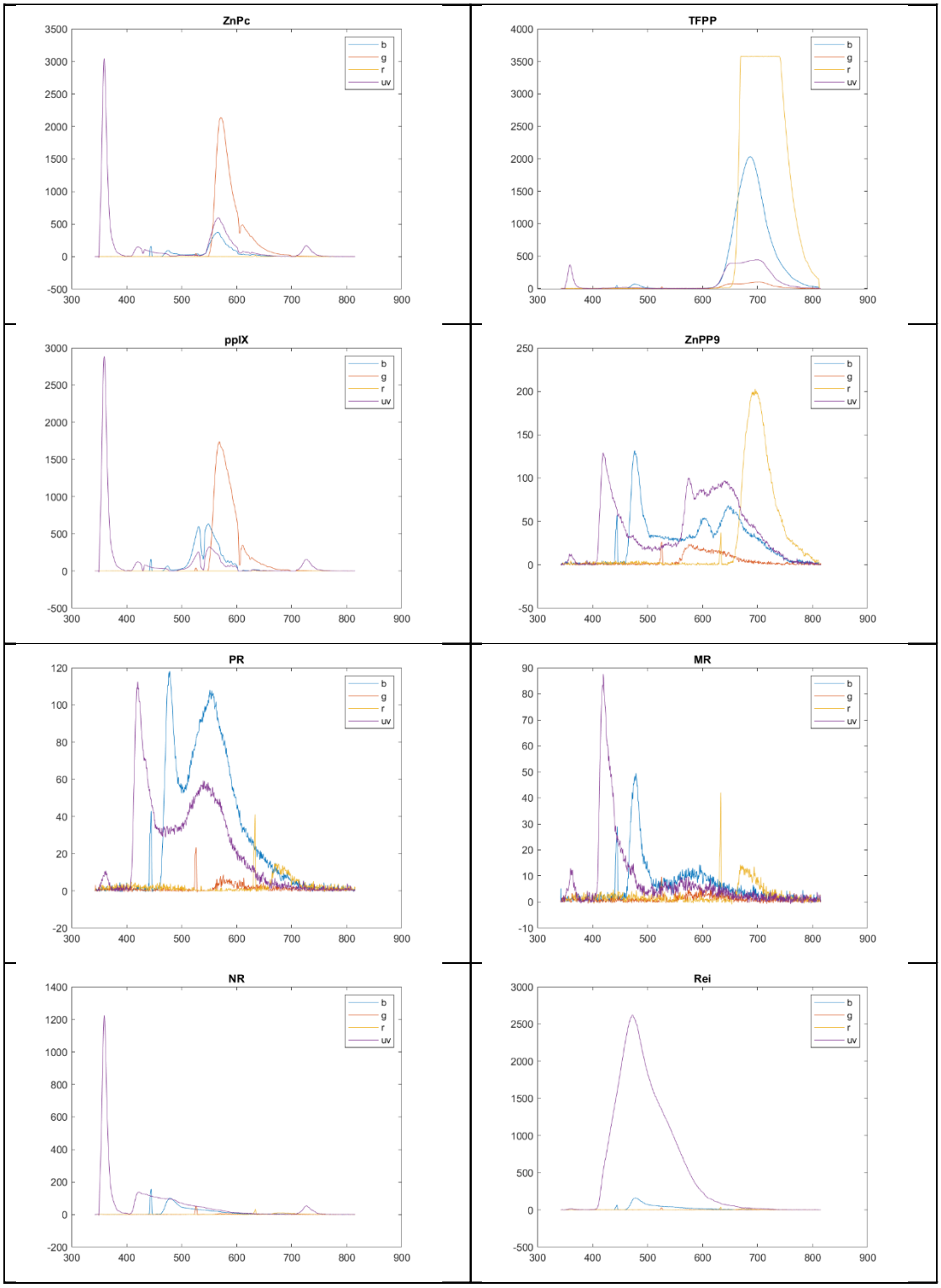


Figure 3-27 Original spectrum of all 16 sensitive material candidates under blue, green, red, and uv laser

Figure 3-27 shows the original spectrum of all 16 sensitive material candidates. As we can see the candidates **(4)**, **(5)**, and **(14)**, namely $(Mg-Cl)$TPPS, TPPS, and methyl red, have very weak fluorescence. For candidates **(4)** and **(5)**, because the sulfonyl hydroxide group is a strong electron withdrawing group, it is reasonable that there is weakening of the coaxial effect of the macrocyclic as well as the fluorescence. Theoretically, methyl red **(14)** should have wide emissions in the range of 320 nm to 480 nm, with maximum at 375nm on excitation at 310nm [139]. However, it did not show any fluorescence when fixed on the PVDF film, probably because of the lack of ionization without solution.

Candidate **(13)** and **(16)**, namely phenol red and Reichardt's dye, failed to present fluorescence changes after exposure to urine vapour. Contrastingly, previous studies have shown that those dyes will change colours when exposed to various VOCs [132]. The reason for the discrepancy may be the influence of high water and low VOC concentrations.

Candidates **(2)** and **(3)** have the same frame structure and almost the same response to urine vapour. Therefore, only **(2)**, namely CuTPP, was kept due to its better signal/noise ratio on excitation at UV and blue laser.

Candidate **(12)**, namely Zn-PpIX, was abandoned due to its uneven deposition onto the PVDF film (see Figure 3-28). It led to large differences between each measurement and poor reproducibility. The reason for such uneven deposition is unclear, it is possibly due to the poor solubility of stabilizer that mixed with Zn-PpIX crystals.

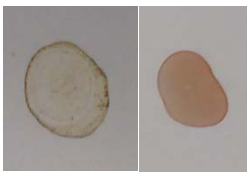


Figure 3-28 Deposition of Zn-PpIX (left) and ZnTPP (right) THF solution on the PVDF substrate, the Zn-PpIX are crystalized on the film, form unevenly died area.

3.3.4. Parameters

One of the most important aims of single element testing is to determine the suitable parameters for optic lasers.

First is the distance between the tip of the light probe and the film surface. Because of the photobleaching effect, dyes exposed to light may lose the fluorescence ability permanently due to a break of a covalent bond or non-specific reactions between molecules by the high energy of light. There are several ways to reduce the effect of photobleaching, such as lowering the incident light intensity, reducing the time of exposure or reducing the frequency of the light thus lowering the photon energy. In current circumstances, the output power of the light source in LAKK-M is not adjustable, and its wavelength is fixed too. So, there are only two ways to reduce the photobleaching effects: by reducing the light energy received per unit area on the surface of the film or reducing the time of exposure.

The time of overall exposure depends on the single exposure time, the number of accumulations, and the process speed of the spectrometer:

$$\text{Overall exposure} = \text{Single exposure time} \times \text{Number of accumulations} + \text{Process time}$$

The processing time required for each spectrum is fixed, while the single exposure time and number of accumulations are adjustable in the LAKK-M PC terminal. In general, longer single exposures can get more photons into the spectrometer, generating a higher intensity signal as well as noise. Whereas, more accumulations can reduce the noise by averaging multiple spectra from each single exposure, thus yielding a higher signal-to-noise ratio. All the sensitive materials used have quite good quantum yields, therefore the exposure time does not have to be set very high. While the number of accumulations is very effective in reducing noise, we still can use a de-noising algorithm later to reduce it as well.

The light emitted from the tip of the probe is distributed in a cone shape and the spot area increases with the distance between probe and surface. While the fluorescence is generated on the top of the film surface and diverges evenly in all directions, only those directly hitting the receiving fibre can be seen by the spectrometer. Therefore, increasing the hanging distance can increase the spot area. This reduces the energy received per unit area on the film surface, but it will also reduce the fluorescence signal received by the spectrometer.

After repeat testing, the distance between the light fibre tips to the film was fixed at 3mm, permitting the light source to illuminate a 2mm diameter circle. Using this setup, fluorescence intensity of the sensitive materials can be maintained for 30s under exposure of UV light and 1min under blue light without being heavily photobleached.

3.3.5. Film reproducibility

Based on the results from single-element tests, 8 out of 16 sensitive material candidates were selected for making the VOCs sensory array: CuTPP **(2)**, quantum yields ϕ_f: 0.17; TPP **(6)**, ϕ_f: 0.12; Eosin Y **(7)**, ϕ_f: 0.67; Rhodamine B **(8)**, ϕ_f: 0.7; Zinc phthalocyanine **(9)**, ϕ_f: 0.3; TFPP **(10)**, ϕ_f: 0.2; Protoporphyrin IX **(11)**, ϕ_f: 0.1; and Nile Red **(15)**, ϕ_f: 0.7. Among them are three free-base porphyrins, one metalloporphyrin, two fluorescence pH dyes, one phthalocyanine metal complex, and one solvatochromic dye. To verify the reproducibility of our methods of making sensory films, 10 repeat measurements were carried out on the selected sensitive materials, and the normalized coefficient of variations (nCV) of all tests were used for evaluating the reproducibility of sensory films.

The only difference between the normalised coefficient of variations and ordinary coefficients of variations is the normalisation of the spectrum before calculation of

coefficient of variations. This was intended to eliminate the interference from excitation light output power.

$$\text{Coefficient of variations: } C_v = \frac{\sigma}{\mu}, \text{where } \sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2} \qquad (3-1)$$

Where σ is standard deviation, μ is mean value, and N is the number of measurements, x_i is the intensity reading of *i*-th wavelength. Figure 3-29 shows the original fluorescence spectrum of Eosin Y under green light. We can see there are quite large differences between each measurement. However, after eliminating the interference from the excitation light output power through normalisation, the spectrum reveals that the differences are not that significant (Figure 3-30). The plot of normalised coefficient of variation (Figure 3-31) also supports this finding: the nCV values are small (<0.01) at the range of emission peak.

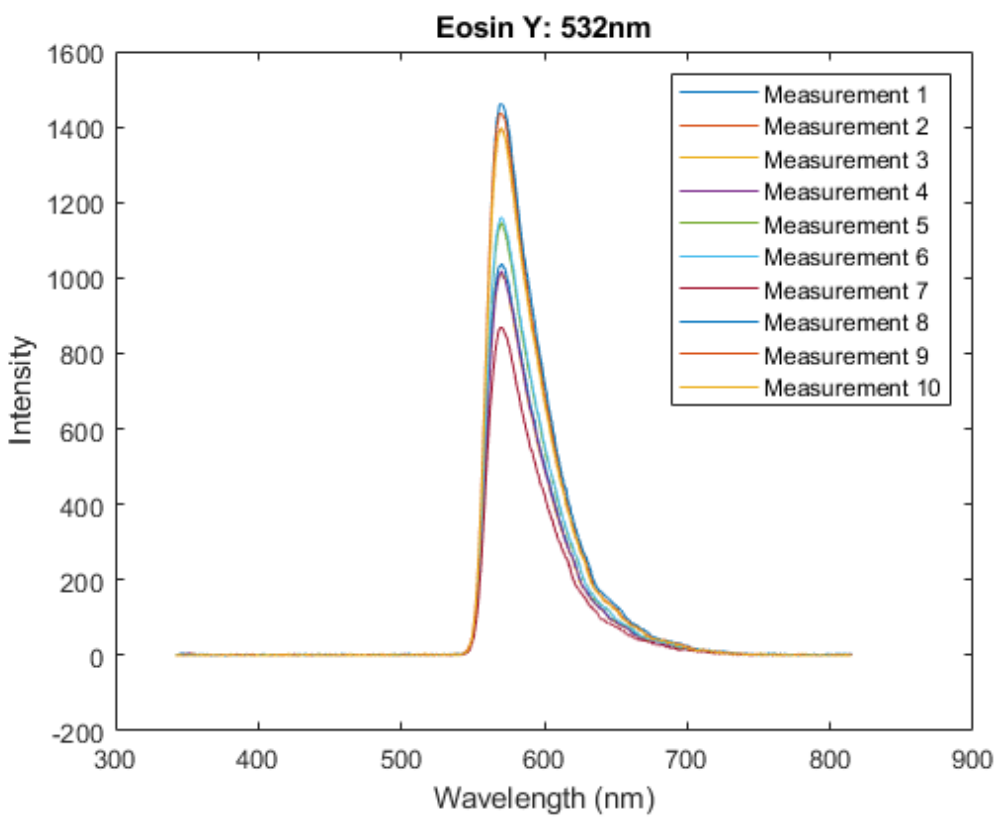


Figure 3-29 Original fluorescence spectrum of Eosin Y under 532nm excitation

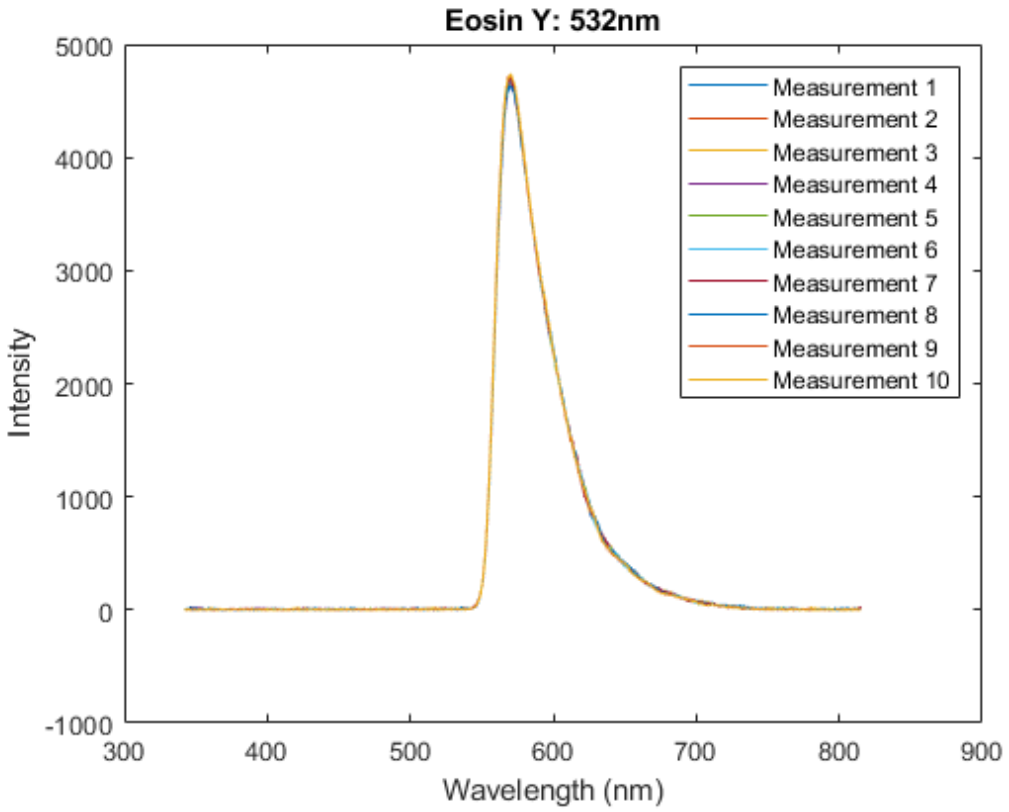


Figure 3-30 Normalized fluorescence spectrum of Eosin Y under 532nm excitation

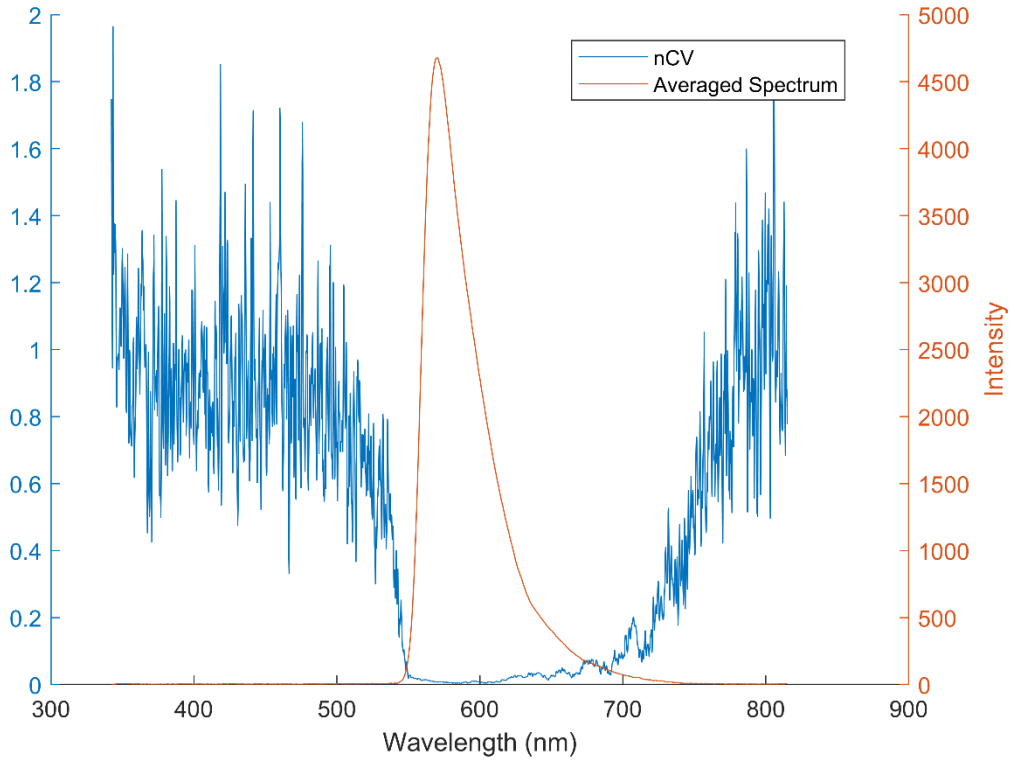


Figure 3-31 Normalized coefficient of variation of 10 repeats of Eosin Y fluorescence spectrum under 532nm excitation

Therefore, using nCV to evaluate the variation between repeat measurements is feasible. As shown in Figure 3-32, all the selected sensitive materials have good reproducibility with very small nCV (<0.05).

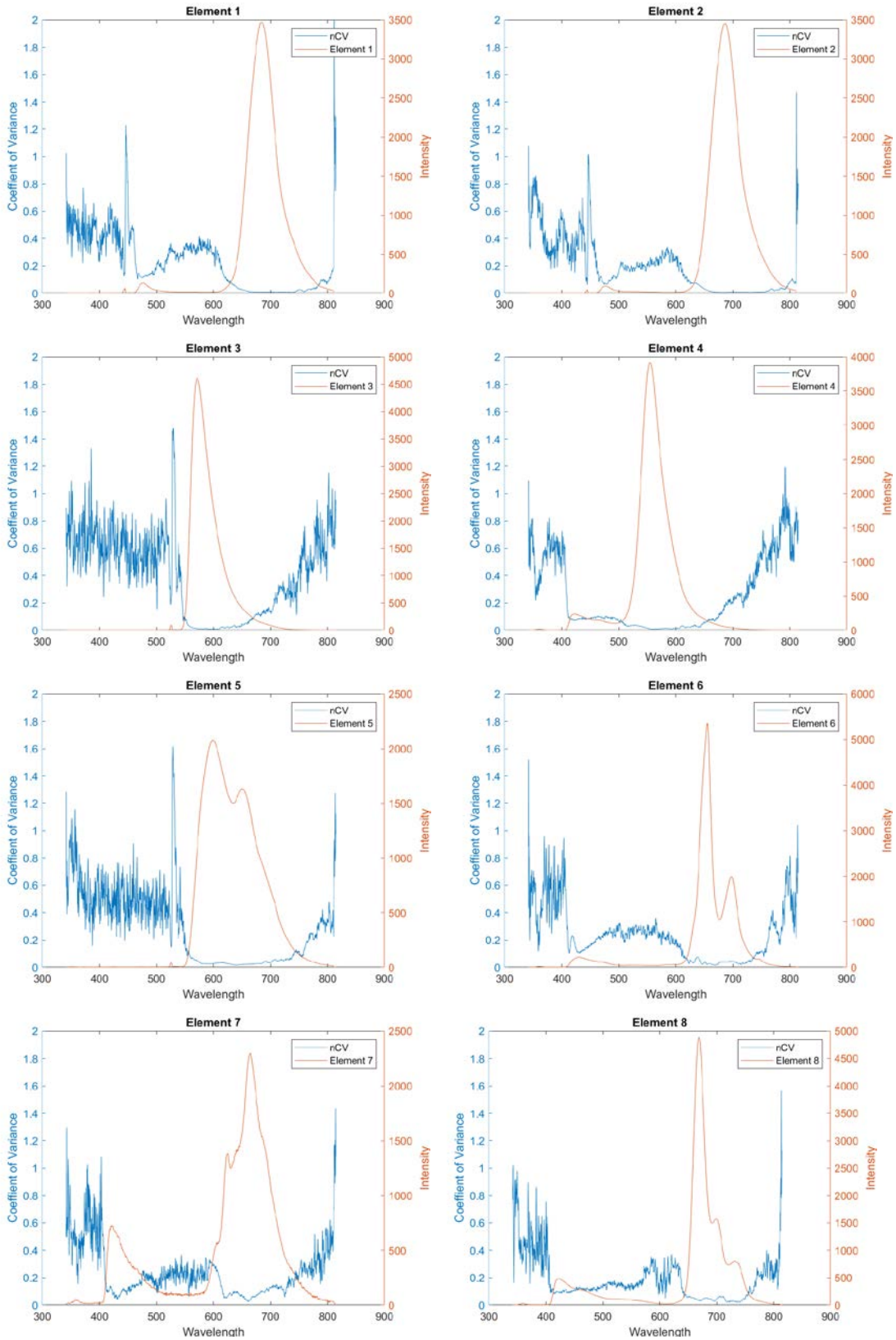
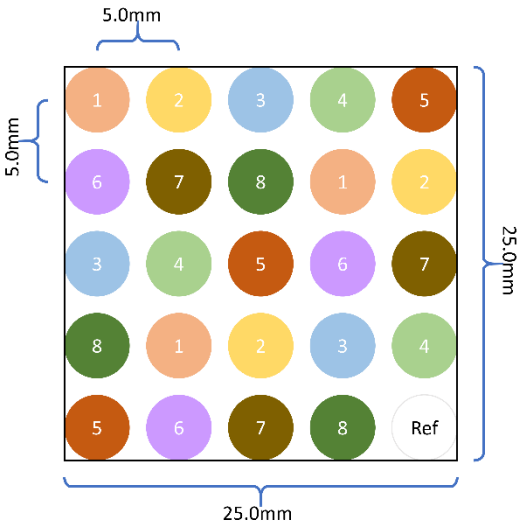

Element 1
Element 2
Element 3
Element 4
Element 5
Element 6
Element 7
Element 8
rCV
Coeffient of Variance
Intensity
Wavelength

Figure 3-32 nCV of 8 selected sensitive materials, measurement number = 10

3.3.6. Sensory array layout

Figure 3-33 shows the layout of the 24-elements in the sensory array film. Unless stated otherwise, all the "sensory arrays" in this thesis followed the same layout and composition as this design: a 25mm×25mm PVDF film with 24 element points with each point spotted with 1.5μL of 1mM sensitive material THF solution. The distance between each point centre is 5mm. The right-bottom corner on the film is the reference point to be used as blank control of the background and location point for film orientation.



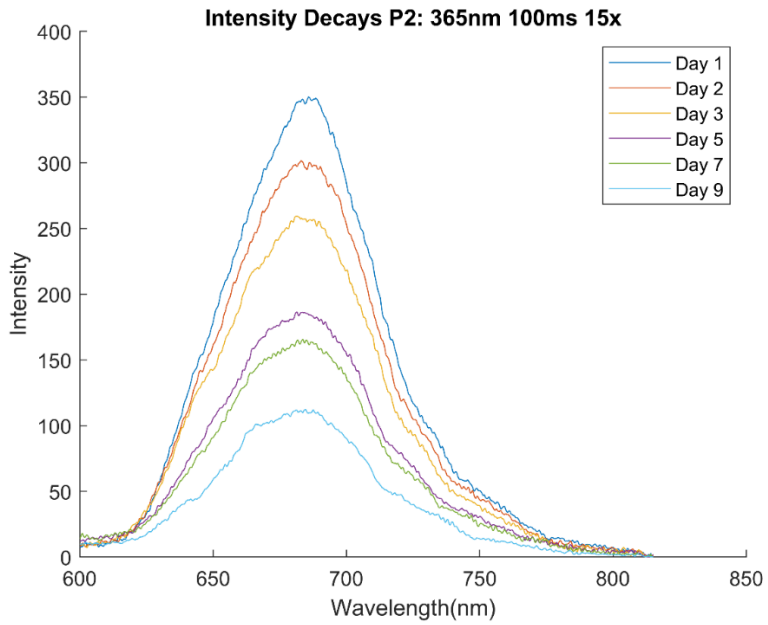
Point number	Sensitive material
1	CuTPP **(2)**
2	TPP **(6)**
3	Eosin Y **(7)**
4	Rhodamine B **(8)**
5	Zinc phthalocyanine **(9)**
6	TFPP **(10)**
7	Protoporphyrin IX **(11)**
8	Nile Red **(15)**

Figure 3-33 24-elements sensory array layout and sensitive material composition

3.3.7. Fluorescence decays

The fluorescence emission intensity decays over longer periods (days to weeks) were observed during the experiments. The following designed experiment was carried out to investigate the shelf life of the sensory films. This test aimed to evaluate the durability of the sensory film and find out the cause of decay.

Two 24 elements PVDF sensory films (with layouts as described previously) were measured 0, +1, +2, +4, +6, +8 days after being freshly prepared. The reference film was stored in an aluminum foil covered and sealed petri-dish at room temperature. The testing film was stored in the same petri-dish with the lid opened in the same room. The room temperature was 21°C and humidity was 60% and remained stable during the tests.



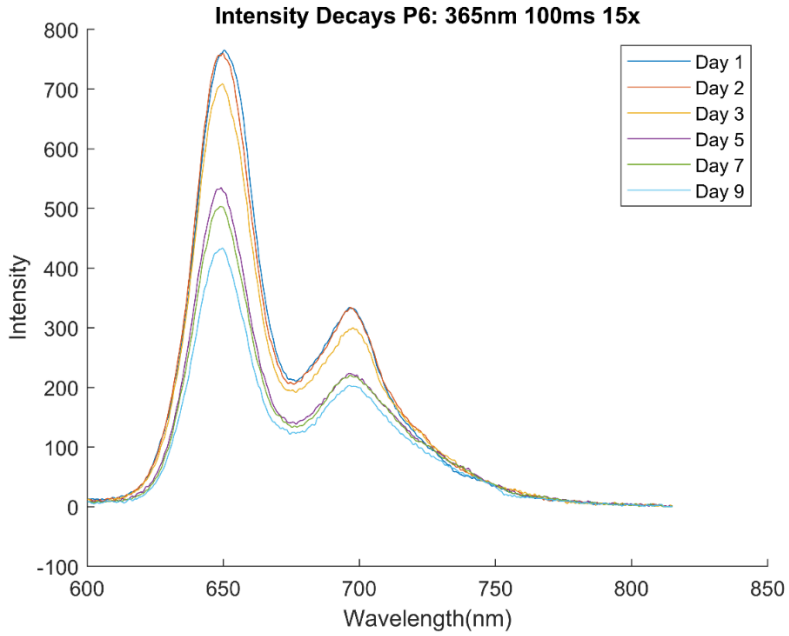


Figure 3-34 emission intensity decays of porphyrin sensitive materials element point P2 (TPP) and P6 (TFPP) on PVDF film.

Figure 3-34 shows the decay of the emission spectrum of element point 2 (TPP) and 6 (TFPP) over a 9-day period in lid-open group; the spectrum has been normalised to eliminate the influence of the optic power of light source. It is clear that porphyrins do lose their fluorescence ability over time when exposed to air, while fluorescence dyes like eosin Y do not (Figure 3-34). In comparison, another sensory film from the same batch of production was stored in the same condition except for the fact it was in a sealed petri-dish for 9 days. In this test, all the element points did not decay during the test period. Therefore, it is reasonable to hypothesise that something in the air, most likely oxygen, quashed the porphyrins' fluorescence ability.

3.4. Conclusion

In this chapter, the physical chemistry and optical principles of fluorescence gas sensitive materials were reviewed. Three major types of fluorescence gas sensitive materials were discussed. According to Gouterman four orbital model, the energy gap between HOMO and LUMO decides the absorption and emission of porphyrin and porphyrin derivatives. When environmental molecules adduct with the porphyrins, their HOMO/LUMO gap will change and the central ions and steric hindrance will affect the selectivity of the coordination of the ligands. For pH indictors, the molecules usually have two forms of structure and switching between them depends on the environmental proton strength, showing pH-dependent spectral properties. Solvatochromic dyes will change colour based on the environmental polarity, usually from solvent polarity.

Based on the aforementioned principles, a total of 16 sensitive materials were tested using a custom-built detection system. This system was composed of a

laser/spectrometer device (LAKK-M); a sample processing module, used for turning the liquid sample into gas phase; a sensory array, made by different manufacturing methods; a reaction chamber, to allow the VOC molecules to interact with the sensor array. The system has several versions to suit different applications and a set of hardware and software parts for control and communications.

This system fulfilled the requirement of a proof-of-principle medical device prototype for urinary bladder cancer diagnosis, it can handle different testing samples with corresponding hardware configurations, with selective sensitive materials the cross-responsive sensor array has ability to respond to different analytes, next two chapters will verify this using purified urinary VOC biomarkers for bladder cancer and urine samples for bladder cancer patients and healthy controls.

4. Distinguish Test of Urinary VOC Biomarkers for Bladder Cancer

The aim of this section of work was to comprehensively test the performance of the fluorescence urinary VOCs detection system we built. Several tests were carried out to achieve this goal:

1. Establishing a data processing protocol, designed for the data generated from this system
2. Building a classification model that may be able to identify the target urinary VOC biomarkers using processed data
3. Determining the performance indexes of the fluorescence urinary VOCs detection system such as sensitivity, selectivity, stability, responding speed, and limits of detection.

4.1. Introduction

Electronic nose, or e-nose, is a device designed to mimic the discrimination of the mammalian olfactory system for smells [140]. Like its imitation target, the artificial olfactory system follows the same principle of identifying the odorant (Figure 4-1).

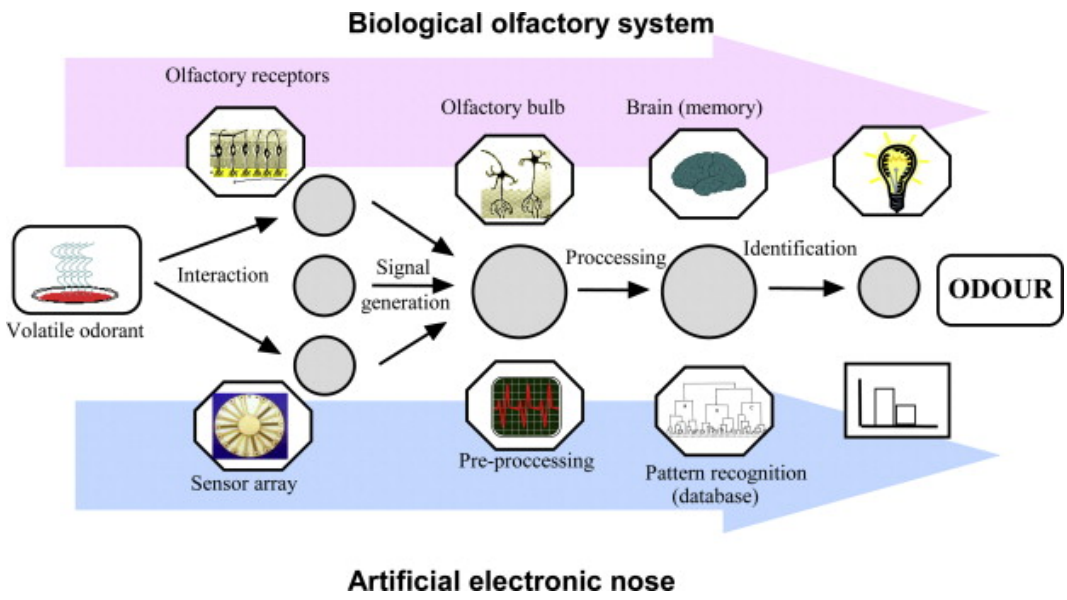


Figure 4-1 Analogy of the biological and the electronic noses (courtesy Ghasemi-Varnamkhasti et al. [141], reproduced with permission from Elsevier).

In a biological system, there are many olfactory receptors on the olfactory sensory neurons. Such receptors can have specific binding to odorant molecules, leading to an action potential in the receptor neuron that converts the chemical signals to bioelectrical signals. In an artificial system, the chemical sensors take the role of olfactory receptors. Depending on the type of sensor, the chemical signals of odorant molecules are converted into electric charge or experience optical or mass changes, eventually turning into electronic signals. There are millions of olfactory receptors in a biological olfactory system; multiple receptors may respond to one odorant molecule, while single receptors may respond to multiple odorant molecules. Similarly, in an artificial olfactory system, multiple sensors combine and build up a sensory array, showing the same “cross-responding” ability of biological system. Through the network of receptors/sensors, a complex information profile of exposed odours was generated that compared these with a pattern recognition database of known odours. This allowed for the identification of the exposed odour.

Therefore, starting from these three aspects, we can summarise current commonly used e-nose techniques.

4.1.1. Sensing technology

In general, the sensing technology can be dived into two types: electronic and non-electronic. The former includes Metal Oxide Semiconductor (MOS), Metal Oxide Semiconductor Field Effect Transistor (MOSFET) and Conducting polymer (CP) sensors. The latter includes Quartz microbalances (QMB or QCM), Surface Acoustic Wave (SAW) and optical sensors.

Metal oxide semiconductor (MOS) based gas sensors are the most invested in type of gas sensor across all e-nose sensor types and have been widely used in environmental monitoring of flammable, toxic or hazardous gases [142]. In general, MOS-based gas sensors have two major components: receptor and transducer. The receptor is composed of a metal oxide surface that has direct contact with analyte gases. Such contact induces the electronic resistance changes of the surface and is read by the transducer. MOS-based gas sensors have a very wide response range (low- to high-ppmV level of wide various chemicals) and a low unit costs, but they require a high working temperature (a few hundred degrees Celsius) and have a long-term usability issue in terms of poisoning by sulfide chemicals or baseline shifting.

Metal-Oxide-Semiconductor Field-Effect Transistor (MOSFET) gas sensors are another type of gas sensing technique, based on the electronic property changes on exposure to chemical vapours. MOSFET gas sensors have a layer of catalyst that covers the insulating gate. Its catalytic products of VOC can infiltrate the insulating gate and cause the threshold voltage to shift, thereby changing the electronic properties of the transistor. A different catalytic layer can be used for detection of various types of VOCs. The MOSFET gas sensor can be manufactured on an integrated circuit with good stability and reliability but requires complex processing techniques.

Conducting polymer (CP) gas sensors are the most widely used e-nose sensing technique, also belonging to electronic based gas sensor types and similar to MOS sensors. The sensitive material coating the top of the components can respond to VOC gases by changing their conductivities. The biggest difference between the CP sensor and the MOS sensor is that the CP sensor relies on conducting particles, such as polypyrole and carbon black, interspersed in an insulating polymer matrix to absorb the analyte molecules rather than metal oxide thin films. The good side of doing this is that the CP coating can work under room temperature, which makes it suitable for

usage in portable device. Meanwhile, the disadvantage of CP sensors is that the coating is very sensitive to water and has a similar baseline shifting in long term uses to MOS sensors.

The Quartz microbalances (QMB or QCM) gas sensor is a non-electronic based gas sensor. Its core component is a polymer coated quartz resonance plate that works at a low MHz resonance frequency. When the polymer coating absorbs analyte gases, the mass of plate will increase and cause the resonance frequency to decrease.; this establishes the linear correlation between gas concentration and component resonance frequency. Its selectivity depends on the material of the polymer coating. Various materials, including porphyrin, were tested for manufacturing QMB gas sensors. The biggest advantage of QMB gas sensor is its structure simplicity and good sensitivity, but it has serious cross-responding issues due to non-specific interactions between polymer coating and interference gases.

The Surface Acoustic Wave (SAW) gas sensor is very similar to the QMB sensor. The SAW sensor also relies on the selective interaction between sensitive polymer coating and analyte gases. However, it is not only focused on the mass changes from material adsorption of gas molecules, but also the conductivity and viscoelasticity influencing the SAW frequency. Compared to QMB, the SAW gas sensor has a higher working frequency and higher sensitivity but lower signal-to-noise ratio. Another major advantage of the SAW gas sensor is that it can be processed using the lithography technique, which has been widely used in large-scale integrated circuit (IC) chip manufacturing. This means it is very suitable for integrated mass production and reduces the overall costs.

Despite the fluorescence gas sensor array described in the previous chapter, there are other optical gas sensing techniques including absorption spectroscopy, optical fibre-based sensors, colorimetric sensors and surface enhanced Raman spectroscopy. The

optical sensing techniques have some unique advantages over the commercially available methods mentioned previously such as: higher detection limit (pptV - low ppbV level compared to ppmV level) and signal-to-noise ratio, room temperature working condition. The biggest disadvantages of optical sensing methods are the expensive costs of optical devices.

4.1.2. Signal processing and Pattern recognition

The raw data obtained from the gas sensors array is usually a 2D array of a series of discrete data. For an array with n sensors, each sensor generates m readings, there are:

$$X_{Gas1} = \begin{bmatrix} Sensor_1 Reading_1 & \cdots & Sensor_1 Reading_m \\ \vdots & \ddots & \vdots \\ Sensor_n Reading_1 & \cdots & Sensor_n Reading_m \end{bmatrix}$$

Then the dataset for each analyte gas X_{Gas} is a n × m 2D array. Depending on the type of sensor, the size of m could be from hundreds to thousands of readings. For most of the electronic and non-electronic type e-nose, the responding curves are composited from a series of transient response signals in time sequence. While for optical type, the readings usually are a series of intensity in frequency sequence. Both of them need normalization to some extent to remove the influence of factors that correlate to time or frequencies.

Other commonly used signal processing in e-noses include resampling, noise reduction, detrend, removal of outliers. Sometimes those process steps are used solely or together to remove interference factors such as artefacts, background noise, trends, or spikes in preparation for further analysis.

The third step shown in Figure 4-1 is identification of odours by memory or pattern recognition. In fact, a single biological olfactory receptor in the human olfactory epithelium can only respond to very few kinds of odours with limited sensitivity and specificity. Contrastingly, the olfactory neuro-network can enhance the ability of a

single olfactory receptor by 1000 times by integrating hundreds of thousands of receptors' information [143]. So does the pattern recognition algorithm in artificial olfactory systems; a single gas sensor only has limited sensitivity and selectivity, but when multiple sensors build up a cross-responsive array and combine with suitable pattern recognition algorithms, their identification ability can significantly enhancement. Commonly used pattern recognition algorithms in e-nose include Linear Discriminant Analysis (LDA), k-nearest-neighbours (k-NN), artificial neural networks (ANN), support vector machine (SVM). Pattern recognition is defined as "the automatic discovery of regularities in data through the use of computer algorithms and with the use of these regularities to take actions such as classifying the data into different categories" [144]. The idea of pattern recognition gave rise to today's hot artificial intelligence and machine learning, and it still very useful in dealing with problems like classification and clustering.

4.2. Materials & Methods

4.2.1. Materials

Four VOC biomarkers were selected from the literature of bladder cancer urinary biomarkers [96, 126, 145, 146]: ethylbenzene (CAS: 100-41-4), hexanal (CAS: 66-25-1), lauric aldehyde (dodecanal; CAS: 112-54-9), and nonanoyl chloride (CAS: 764-85-2). All were purchased from Sigma-Aldrich (Sigma-Aldrich, UK). All chemicals were of analytical grade and were used as received without further purification.

All the mechanical parts and micro controllers were used as described in Chapter 3.

4.2.2. Sample processing module

As described in Chapter 3, the sample processing module for the VOC vapour test consists of three parts: liquid flowmeter, for controlling the VOC liquid flow; gas flowmeter, for controlling the carrier gas flow; and the mixer, for vaporising the VOC

liquid and mixing it with the carrier gas to achieve the desired concentration.

4.2.3. Reaction chamber & sensory film

As described in Chapter 3, the reaction chamber used in VOC tests were Mark. I and Mark. II, with corresponding single-element sensory film and multi-element arrays made from low-fluorescence PVDF membrane.

In short, the Mark. I chamber was designed to validate the feasibility of the whole system, therefore it only has some basic functionalities to allow one single-element of sensory film to be exposed to VOC vapours in a sealed environment, Meanwhile, the detecting probe of LAKK-M can be inserted into the chamber and hover over the top of the film so it can read the fluorescence signal in real-time.

The Mark. II chamber allows the chamber to hold bigger film with a multi-elements array. Apart from that, a mechanism driven by a servo motor allows the automatic switching between sensor points. Although changing the optic filters of LAKK-M still relies on manual operation, this chamber design allows the operator to read all readings of the array without opening up the box, therefore ensuring the accuracy of the flow control.

4.2.4. Experiment workflow

4.2.4.1. Single-element test

To validate the performance of the sensitive material candidates, the reaction chamber Mark. I was made for a single-element test. As described in Chapter 3, the 3D printed reaction chamber has two parts: the upper part has a hole to allow the detection probe of the LAKK-M laser device in and the lower part has a stand to support the sensor film and allow the VOCs vapours to spread into the chamber. The two parts were screwed together to ensure a tightly sealed environment.

Before starting, a single-element PVDF sensory film was mounted in the lower part of the Mark. I chamber, then the upper part was screwed in to combine the two parts. Next, the reaction chamber with the sensory film was mounted on the empty flask, and the fluorescence spectrum under ultraviolet (UV; 365nm), blue (450nm), and green (532nm) excitation light was detected by the LAKK-M device. An accumulation of 15 spectra were recorded using an acquisition time of 150ms per exposure at an excitation power of 4.2mW. After that, 5mL of VOC liquid was added into a flask and warmed to 37℃, and the VOC gas was released and spread into the chamber to react with the sensitive materials on the sensory film. After waiting two minutes for the reaction to take place, the fluorescence spectra for the same excitation lights were recorded.

To further ensure the responses are also applicable to urine, a urine sample from a 26-year-old male volunteer with signed consenting was tested for each fluorescence sensitive material candidate.

To reduce errors, the background signal of the device was measured, and the data showed that the materials of the 3D printing device and substrate can cause background fluorescence. The main reason is that the polymer materials will emit fluorescence, especially under UV and blue light excitation. Finally, we chose black PVA as 3D printing material and low-fluorescence PVDF as a substrate. The background of the substrates and chamber material were measured before each test as a reference.

4.2.5. Data processing

Each raw data file exported from LAKK-M contains the spectrophotometer reading of one element excited with single light, with 2100 wavelength and corresponding intensity readings. Therefore, for a 24-element array, there are 24 separate files for

each element. During the export process, additional information of date and time, test object, sensor identity number, film batch code, and excitation light were added to the file. In MATLAB, two datasets were generated from the files. The first dataset is the reading dataset, containing only intensity readings from each element as a 24×2100 two-dimension double numeric array. The second dataset is the identity dataset, it contains the additional information aforementioned in a 24×5 two-dimension string array.

Four steps of signal processing were then applied to the reading dataset: removing backscattering; normalization; de-noising; and serialization. This is shown as flow diagram in Figure 4-2. The product of data processing is the serialized feature array, which is a sequence of fluorescence peaks of each sensor point.

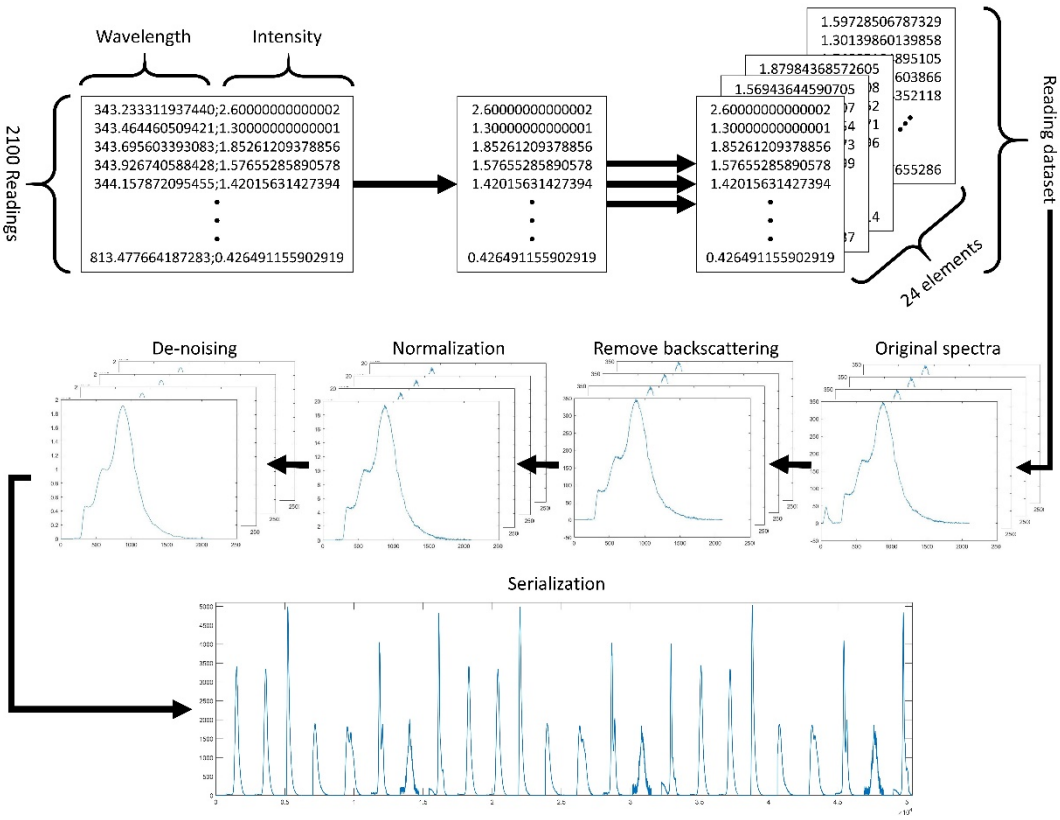


Figure 4-2 Flow diagram of data processing

4.2.6. Differential spectrum

Differential spectrum is a key concept of fluorescence sensor array responses. Differential spectrum is the difference between the fluorescence emission spectrum of one sensory element *before* exposure with any analyte and the fluorescence emission spectrum of the same sensory element *after* exposure with analyte. For each point Δ_{I_λ} on the differential spectrum at the wavelength λ, it has:

$$\Delta_{I_\lambda} = I_\lambda - I'_\lambda \qquad 2$$

Where I_λ is the intensity of the 'Before' spectrum at the point of wavelength λ, and I'_λ is the intensity of the 'After' spectrum at the point of wavelength λ.

This principle also applies on the sequence array; the differential sequence array will be the input of the discrimination algorithm.

4.2.7. Discriminate analysis

Two common dimension reduction methods were used: signal characteristics selection, used for feature selection, and Principal Component Analysis (PCA), used for feature extraction.

Five signal characteristics: Maximum Peak Value (MXV) and Wavelength (MXW), Minimum Peak Value (MNV) and Wavelength (MNW), and Peak Area (PA) were extracted from the spectrum, as shown in Figure 4-3:

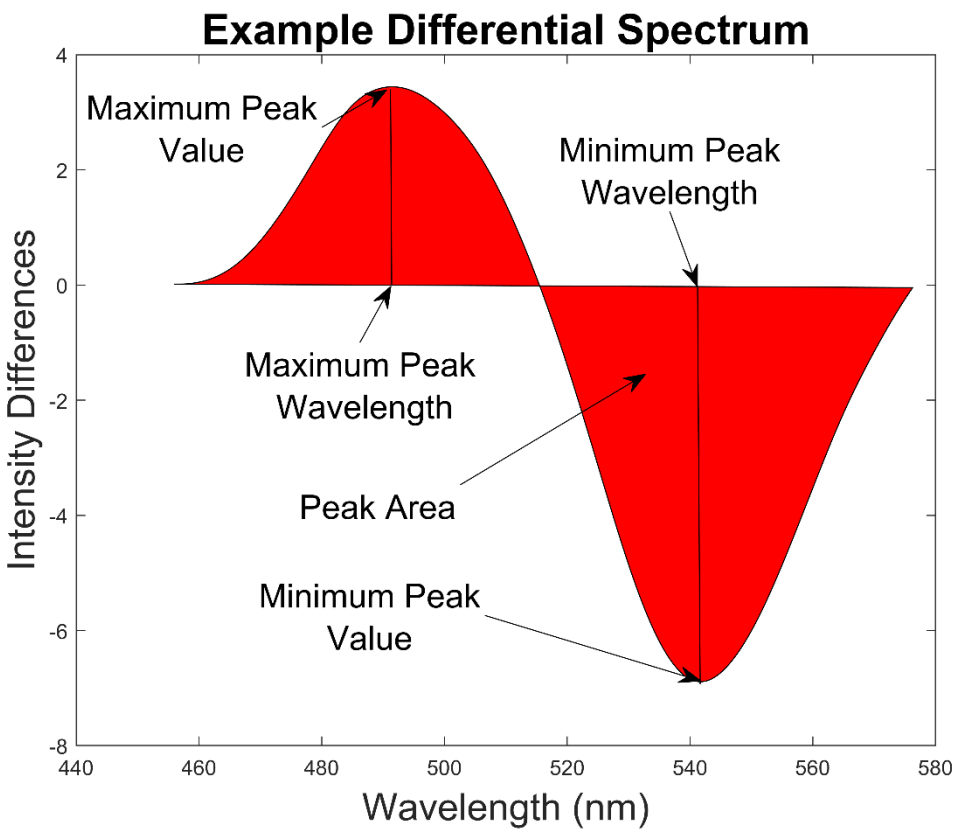


Figure 4-3 Differential Spectrum and feature selection

The signal characteristics from each differential spectrum then formed a **spectrum signal characteristic dataset** $\boldsymbol{X_{SSC}}$. For each $\boldsymbol{X_{SSC}}$ generated from *n* samples tested with *m* elements array, there are:

$$X_{SSC} = \begin{bmatrix} \mathrm{MXV}_{11} & \mathrm{MXW}_{11} & \mathrm{MNV}_{11} & \cdots & \mathrm{MNV}_{1m} & \mathrm{MNW}_{1m} & \mathrm{PA}_{1m} \\ & \vdots & & \ddots & & \vdots & \\ \mathrm{MXV}_{n1} & \mathrm{MXW}_{n1} & \mathrm{MNV}_{n1} & \cdots & \mathrm{MNV}_{nm} & \mathrm{MNW}_{nm} & \mathrm{PA}_{nm} \end{bmatrix} \quad 3$$

The size of X_{SSC} are *n-by-5m*.

Principal Component Analysis (PCA) was used for feature extraction. PCA can generate a new set of variables from linear combinations of original variables without losing too much information, therefore reducing the number of variables used for further analysis. The linear combinations of original variables called principle components (PCs), and its amount is picked based on the explanation of original data.

Usually, PCs that can explain 95% of the raw data are sufficient for use.

Smaller numbers of variable can largely reduce the computation of discriminate analysis and model validation. Linear Discriminate Analysis (LDA) was used for building the classification model for the testing VOCs, while leave-one-out cross validation was used for validating such model.

All the data processing and computations were performed using MATLAB 2018b (MathWorks, USA).

4.3. Results & Discussion

4.3.1. Target VOC biomarkers

As discussed in Chapter 2, there are various origins of the bladder cancer biogenic VOC biomarkers: endogenous and exogenous. For endogenous sources, the most studied pathways are: oxidative stress, cytochrome p450 detoxication, carbohydrates metabolism, and lipid peroxidation [56]. Through researching the literature, four VOC biomarkers were carefully selected for testing the system shown in table 4.1 [96, 126, 145, 146].

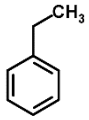
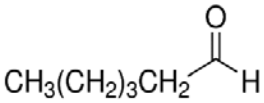
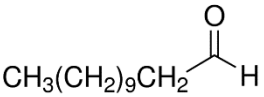
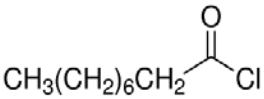
Structural formula	VOCs	CAS	Formula
	Ethylbenzene	100-41-4	C_8H_{10}
	Hexanal	66-25-1	$CH_3(CH_2)_4CHO$
	Lauric aldehyde (dodecanal)	112-54-9	$CH_3(CH_2)_{10}CHO$
	Nonanoyl chloride	764-85-2	$CH_3(CH_2)_7COCl$

Table 4.4-1 Urinary bladder cancer VOC biomarkers tested.

Of the VOCs tested, ethylbenzene is an aromatic hydrocarbon, which smells like gasoline. Hexanal and lauric aldehyde are aliphatic aldehydes with a grassy and soapy flavour. Meanwhile, nonanoyl chloride is acyl halide with a pungent, mustard-like odour.

Both lauric aldehyde and hexanal are aldehydes that are probably generated from lipid peroxidation. Hexanal has been found to be elevated in both breath [84] and urine [126] of cancer patients, while such elevation is only present in urine for lauric aldehyde [145]. From the structures, we can see the only difference between hexanal and lauric aldehyde is the length of the carbon chain. Since both of them are saturated, straight monoaldehydes, they can be considered as the end products of lipid peroxidation and

would be good targets to validate the selective responding of our sensors due to the aforementioned steric hinderance effect.

Ethylbenzene is a metabolism product of aromatic compounds. Usually, ethylbenzene is considered as an air pollutant and should be metabolised by detoxication cascades. In the blood, ethylbenzene can be oxidised by NADP and turns into benzyl alcohol. Benzyl alcohol further turns into benzoic acid by NAD, and finally condensates with glycine and turns into hippuric acid and is discharged in urine. However, the hippuric acid, the end product of ethylbenzene metabolism process, has been found to be of a lower concentration in the urine of those with bladder cancer than that of healthy groups [147]. Therefore, we have reason to believe that the ethylbenzene found in bladder cancer patients' urine is excreted directly into the urine by cancer cells and is not present in the blood in large quantities. This means ethylbenzene is more likely to be from an endogenous source, not an exogenous one such as inhaled air pollutant.

Among the four VOC biomarkers we tested with our system, only nonanoyl chloride may be an exogenous VOC. The detailed biogenic origin and metabolism pathway of nonanoyl chloride is not clear, but from its chemical property, it is unlikely to be from endogenous sources. However, as an acyl halide, the nonanoyl chloride is a typical electrophile and Lewis acid, which can be detected by metalloporphyrin. The reason for using exogenous VOCs is to verify that the system can distinguish the endogenous VOCs from exogenous ones, and to find possible linear relationships between the concentration of VOCs and the sensor array's responding, as well as the limit of detection (LOD).

4.3.2. Sensors responses

Figure 4-4 to Figure 4-7 shows the raw differential spectra of each sensor point in response to each urinary VOCs biomarker of bladder cancer. Those spectra are not

processed. As we can see, the degree of change of intensity varied between ±0.5 to ±100; some spectra have maximum intensity changes at the backscattering region, and some sensor points have weak responses and therefore low signal-to-noise ratio. The raw differential spectra need further signal processing to remove those influencing factors.

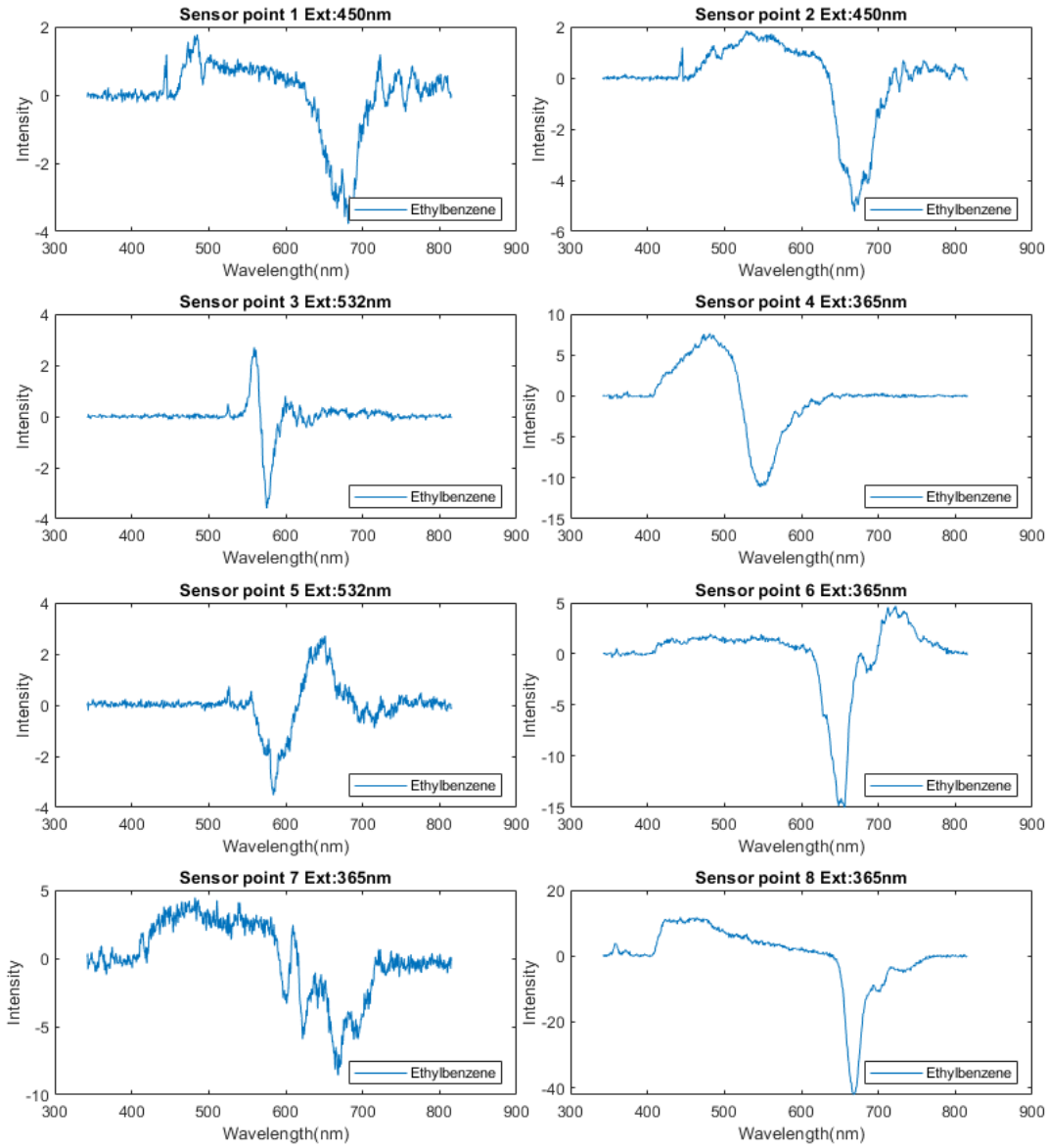


Figure 4-4 Differential Spectrum of each sensor point in response to ethylbenzene

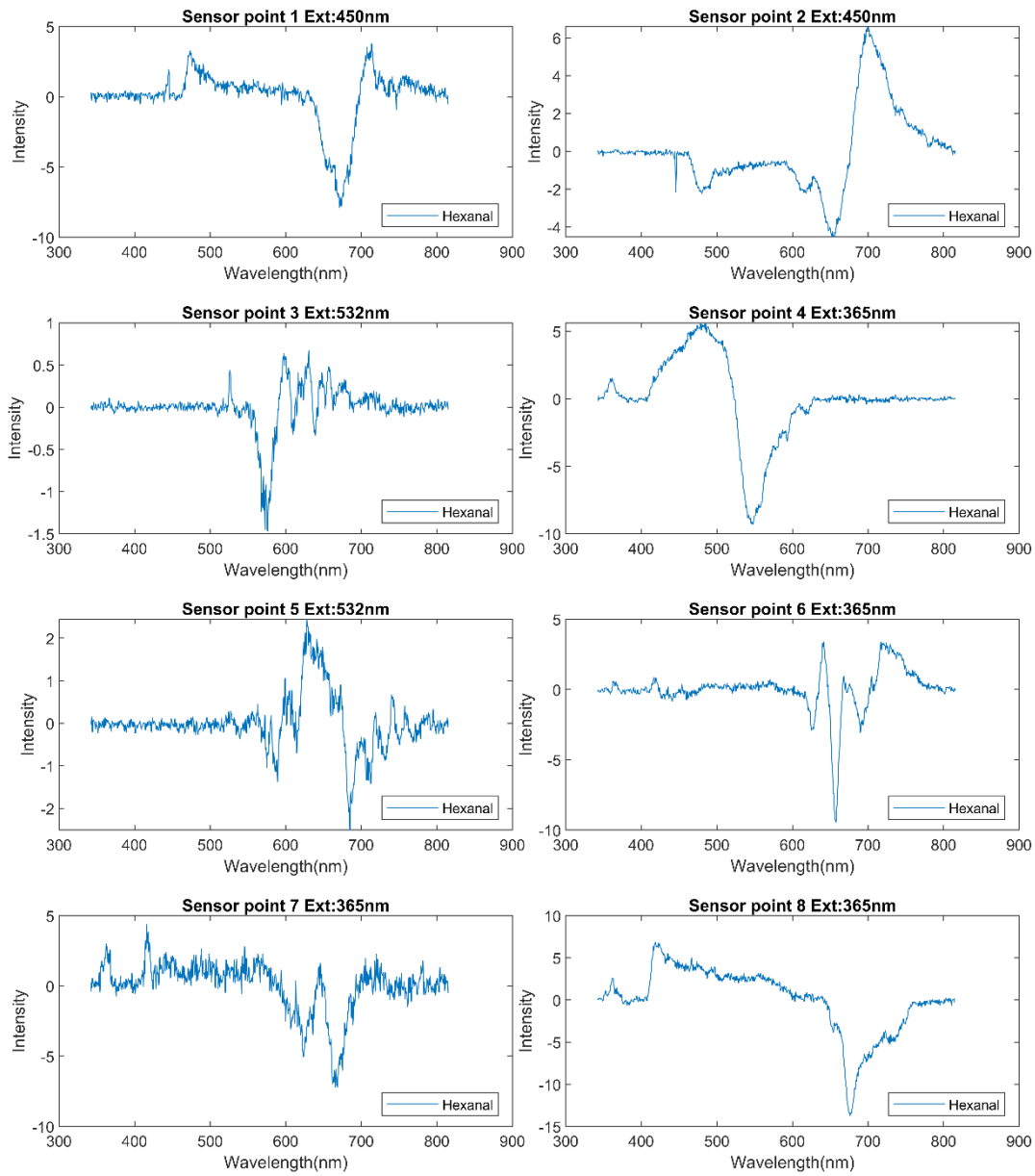


Figure 4-5 Differential Spectrum of each sensor point in response to hexanal

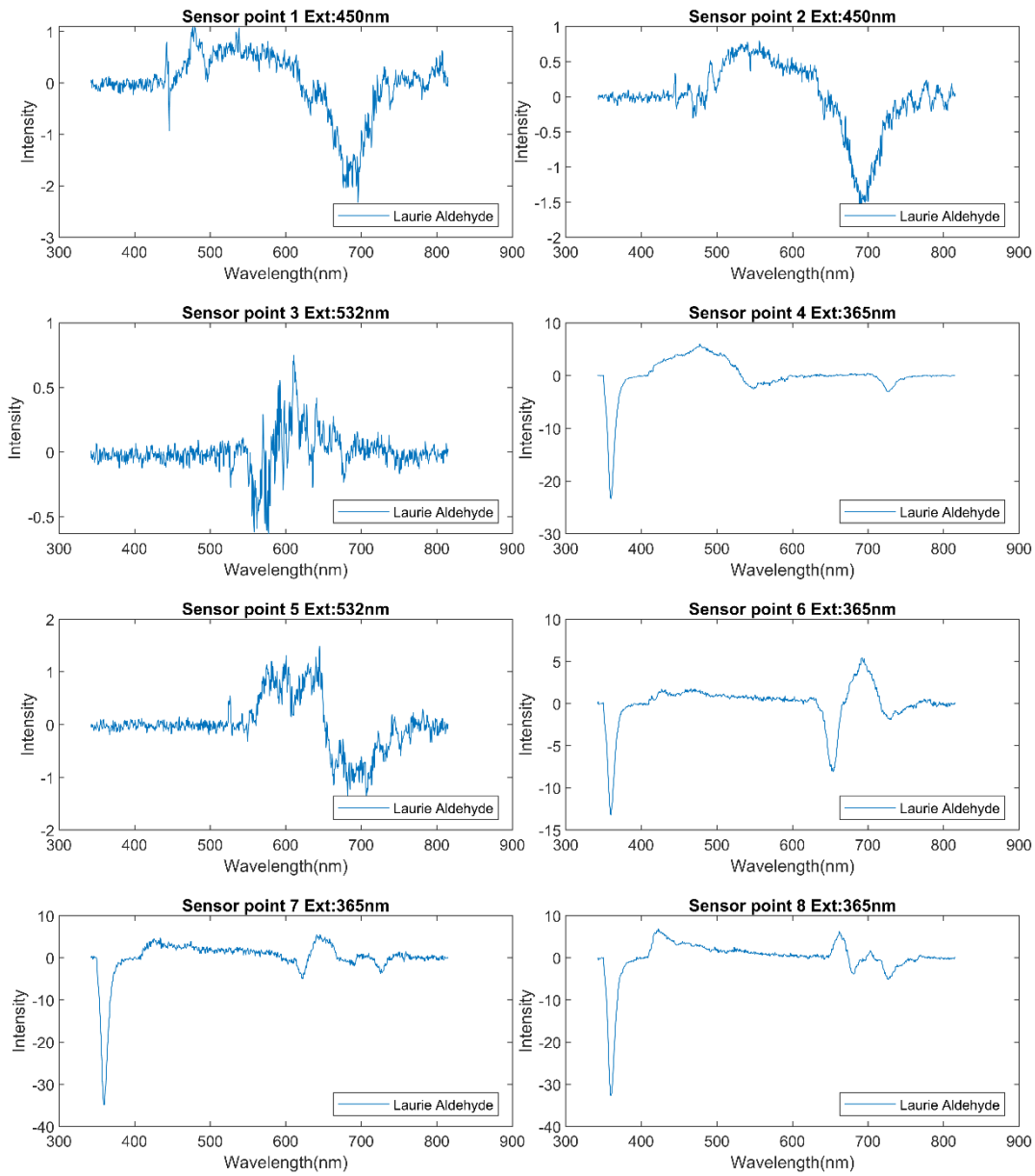


Figure 4-6 Differential Spectrum of each sensor point in response to Lauric aldehyde

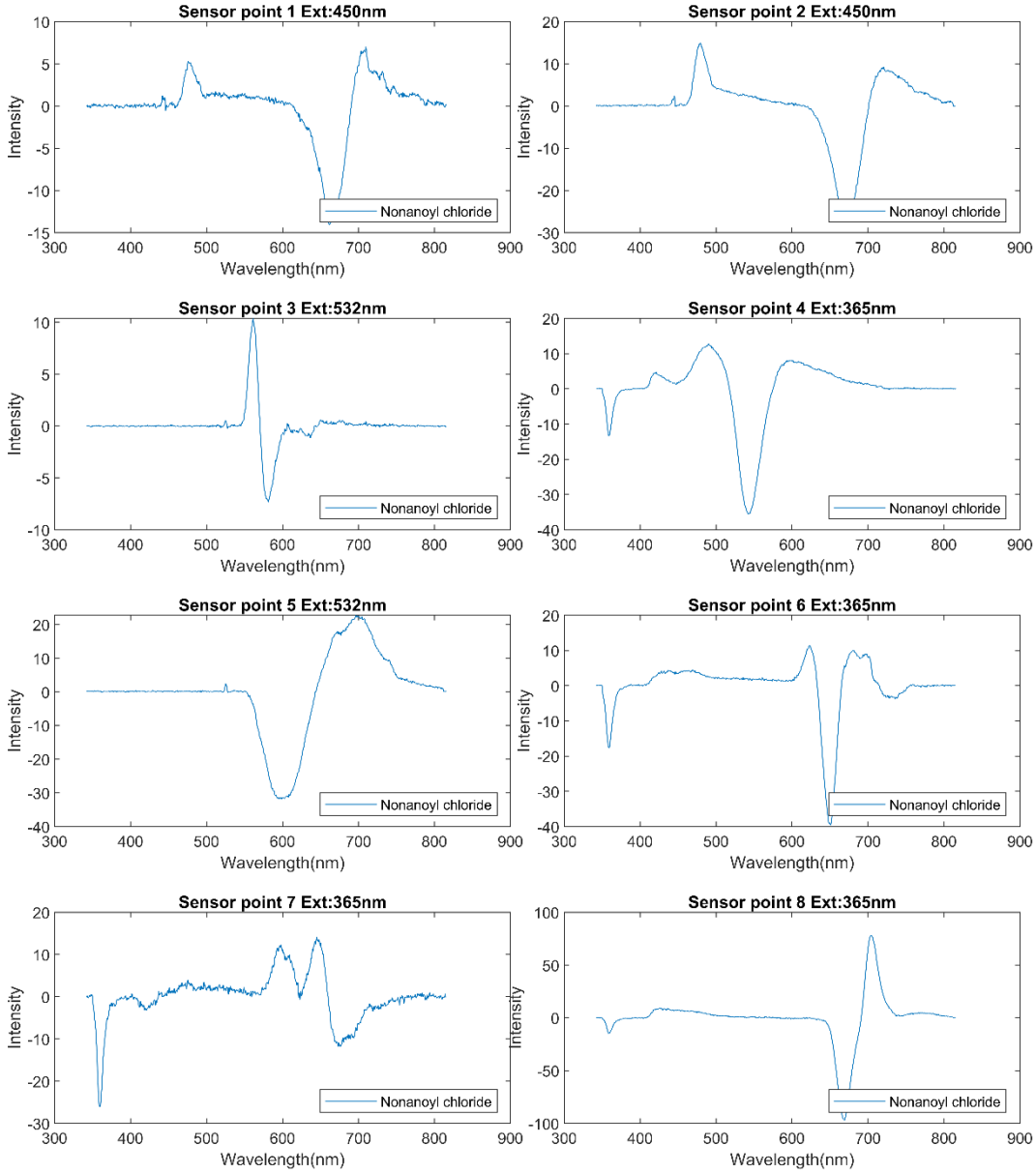


Figure 4-7 Differential Spectrum of each sensor point in response to nonanoyl chloride

4.3.3. Data processing

The raw data file exported from LAKK-M software was a 72 kb .csv file for each single excitation and the structure of the datasets is shown in Figure 4-8:

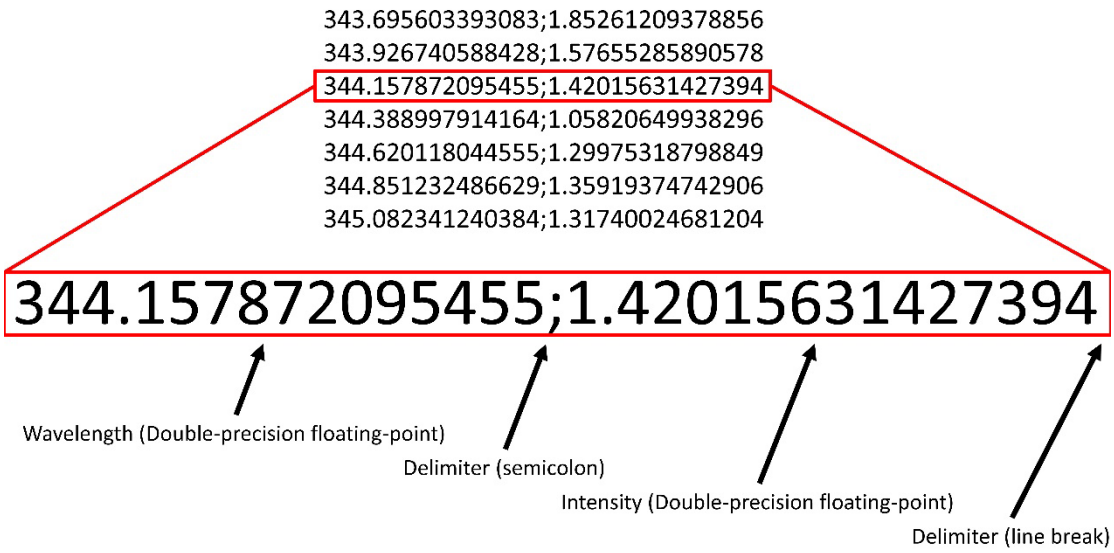


Figure 4-8 Dataset structure of raw csv file exported from LAKK-M program

Each line in the dataset represents a reading from the spectrophotometer, as shown in Figure 4-8, containing one wavelength reading and the corresponding intensity reading, separated with a semicolon delimiter. The numbers are saved as a double precision floating point decimal, which means they have a 15 significant digit precision, much higher than the spectral resolution of the spectrophotometer (0.22nm). The speed-up of processing by reducing the decimal numbers is almost negligible in modern computers, but it may be significant on portable devices since they have slower processors and smaller storage space. By converting the data storage from a double precision floating point decimal to a single precision floating point decimal, the size of one measurement file is reduced by 48.6%.

The backscattering is the signal from excitation light scattering back to the device, in the spectra it looks like a sharp and narrow peak on the same wavelength of the excitation light. Its intensity is affected by several factors: the intensity of incident light; the absorption and reflection of the illuminating material (including sensitive material and substrate); and the angle of the light probe. To ensure the accuracy of

normalization, the backscattering peak was removed at the first step of the data processing.

Normalization is a key process in reducing the influence of different output powers of excitation lights. Because at the first step the backscattering has been removed, the trapezoidal numerical integration of the spectrum can be considered as the emission output of the fluorescence material. Meanwhile for the same material under the same conditions, the emission output of the fluorescence material depends on the output power of the excitation light to a larger extent.

Commonly used normalization standards include total area under the spectrum, internal peaks (unchanged peaks within the studying spectrum) and external peaks (peaks from external standard spectrum). In this study, normalization over the total area under the spectrum was used, as shown in the equation 4-1. The normalized intensity of each point I_{λ}^{Norm} equals the original intensity of the same point on the spectrum curve I_{λ} divided by the trapezoidal numerical integration of the whole spectrum.

$$I_{\lambda}^{Norm} = \frac{I_{\lambda}}{\sum_{i=0}^{n}(\lambda_{i+1} - \lambda_i)I_{\lambda}} \qquad (4-1)$$

Savitzky-Golay filtering was used for denoising [148]. Several commonly used denoising filters were tested: moving average; local regression; fast Fourier transform; and wavelet package denoising. The reason for choosing Savitzky-Golay filtering is because it can remove high-frequency noise efficiently without shifting the fluorescence peaks, while other denoising methods cannot. This is a very important property for keeping the responding information of each of the sensors thus ensuring reproducibility and accuracy.

In MATLAB, there is a built-in function *sgolayfit* in the signal processing toolbox:

```
y = sgolayfilt(x,order,framelen)
```

Where *y* is the output smoothed data, *x* is input data, *order* is the polynomial order of local function to fit the curve and *framelen* is the length of frame of local function to work on. In most cases, the signal-noise-ratio is high and fluorescence signals do not have too much detail; linear function is sufficient (*order* = 1), while the frame length is set to 25.

Serialization is essential for discriminant analysis modelling. It will transform the two-dimensional reading dataset array to a one-dimensional feature array. On the image, it looks like the spectra connect each other end-to-end and become a long sequence of spectrums with multiple peaks. During serialization, a new column is added to the identity dataset that is used for keeping the location information of the spectrum in the original 2D array to correspond to the new 1D feature array.

4.3.4. Discriminate analysis

The idea of discriminate analysis is to establish an association between a categorical variable and a set of interrelated variables [149]. Different applications have different names for the two concepts such as: variables and observations, respond and predictors, dependant variable and independent variables. In discriminate analysis, the first one is called **classes** while the second one is called **features**. Specific to our study, the classes are the sample categories we know, such as cancer or non-cancer, or different names of VOC chemicals whilst the features are the data received from the sensors, or any dataset generated from them.

Consider the characteristics of the differential feature array generated from our data; the **number of features** of each sample (***p***, also called **dimensions**) is much larger than the **number of samples (*n*)**. In this case ($p>>n$), the discriminate will easily get

overfitted, which means the classification model picks too many irrelevant features and reduces the accuracy of the model relative to models with fewer features, known as “curse of dimensionality”. In our case, the differential feature array contains a lot of background signal. Even though we can remove those backgrounds and the regions without fluoresce emission, the scale of the feature array is still much larger than the number of samples we get.

There are two feasible ways to ease the curse of dimensionality: **feature selection** and **feature extraction**. Both methods aim to reduce the number of features without losing the characteristics of the data. Theoretically, increasing sample numbers can also help, but the number of samples required increases exponentially with the dimensions, so it is not the most common method in practice. The strategy of feature selection is to reduce the dimensions by selecting a subset of original features. Meanwhile, the strategy of feature extraction is to find one or more combinations of original features that can explain the samples most (principal component analysis, PCA) or distinguish the samples most (linear discriminant analysis, LDA).

4.3.4.1. Feature selection: signal characteristics of differential spectrums

As shown in the Figure 4-3, a typical differential spectrum of fluorescence peak can be described in these parameters: Maximum Peak Value (MXV) and Wavelength (MXW), Minimum Peak Value (MNV) and Wavelength (MNW), and Peak Area (PA). Thereby, for each differential spectrum, the number of features was reduced from 2100 to 5. Obviously, this method only works when there is only one maximum peak and one minimum peak, and it discards most of the minor details of the differential spectrum. However, as described in Chapter 3, the sensitive materials were carefully selected for those with specific reactions with target bladder cancer biomarkers, which means their responses to the VOC biomarkers are known, thus minimising the influence of information loss.

4.3.4.2. Feature extraction (FS): principal component analysis (PCA)

Principal component analysis (PCA) is available in Statistics and Machine Learning Toolbox in MATLAB:

```
[coeff,score,latent,tsquared,explained,mu] = pca(X)
```

X: input *n*-by-*p* data matrix. In our case, *n* represents the number of samples while *p* represents the number of features. ***coeff***: principal component coefficients,which is a *p*-by-*p* matrix whereby each column represents the coefficient of one principle component (PC). ***score***: represents the scores of each sample in each PC, it can be used as an input for other discriminate analysis. ***latent***: principal component variances represented as a column vector; its size is the number of PCs and every element in it is the contribution of corresponding PC in *score*. ***tsquared***: Hotelling's T-squared statistic (t^2) for each sample in *X*. ***explained***: same size as *latent* and is expressed as the percentage of each *latent* in total, representing the percentage of variance explained by each PC. This can help in choosing the number of PCs for dimensional reduction. ***mu***: estimated means of the features in *X*, in our case it is the mean of all samples.

Particularly, when *p>n*, the maximum number of PCs will be *n-1*, as the dimensions required to explain *n* sample in the multivariance space is one less than the number of samples. Therefore, the *coeff* matrix will be in *p-by-(n-1)* size, and *score* matrix will be *n-by-(n-1)* etc.

As the results of dimensional reduction, **the number of PCs (*q*)** was chosen based on the accumulated *explained*. Usually, we use 95% variance explanation as the threshold; choosing this number of PCs means 95% of the variance is explained by the combination of selected PCs, with only 5% variance unexplained. Through the dimensional reduction by PCA, the feature matrix *X* becomes **reduced *score* matrix**

$\boldsymbol{X_{rdcscore}}$, its size thereby is reduced from *n-by-p* to *n-by-q*.

4.3.4.3. Discriminate Analysis: Linear Discriminant Analysis (LDA)

In MATLAB, *fitcdiscr* is a built-in function in Statistics and Machine Learning Toolbox:

```
Mdl= fitcdiscr(X,Y)
```

It returns trained discriminant analysis classification model *Mdl* based on the input features array *X* and classification *Y*.

In our case, *X* can be **spectrum signal characteristic dataset** $\boldsymbol{X_{SSC}}$ by feature selection or **reduced *score* matrix** $\boldsymbol{X_{rdcscore}}$ by feature extraction, while the *Y* can be either VOCs names or cancer and non-cancer classes.

The Classification Object *Mdl* has many properties; one is *Mdl.Coeffs*, which contains the coefficients for the linear boundary between classes and can be used for visualizing the classifier. For the prediction, the classifier aims to minimise the expected classification cost:

$$\hat{y} = \underset{y=1,\cdots,k}{\arg\min} \sum_{k=1}^{K} \hat{P}(k|x)C(y|x) \qquad 4$$

where

$\hat{y}$ is the predicted classification, K is the number of classes, $\hat{P}(k|x)$ is the posterior probability of class k for observation x and $C(y|x)$ is the cost of classifying an observation as y when its true class is k [150]. In two-classes classification, cost function can evaluate the prediction performance of the classifier.

4.3.4.4. Partial least squares Discriminant Analysis (PLS-DA).

Partial least squares Discriminant Analysis (PLS-DA) is a linear classification algorithm that combines the partial least squares regression with the discrimination analysis for classification purposes. The PLS-DA using **Latent Variables (LVs)** to describe the input features of the algorithm, which are the linear combination of original features. **Loadings** are the coefficients of original features in the linear combination, while **scores** are the coordinates of each samples in the LV projection hyperspace. LVs, loadings, and scores are the most commonly used parameters in visualizing the results of PLS-DA classification.[151]

The PLS-DA was performed using third-class Classification toolbox for MATLAB version 4.2 developed by Milano Chemometrics and QSAR Research Group[152]. This toolbox contains various classification algorithms with a graphic user interface for visualization the classification results.

4.3.5. Classification results

Figure 4-9 shows some of the most significant spectra changes of sensory points during multiple-element array tests after exposure with different urinary VOCs biomarker of bladder cancer. The unfiltered differential spectra can be seen in Figure 4-4 to Figure 4-7.

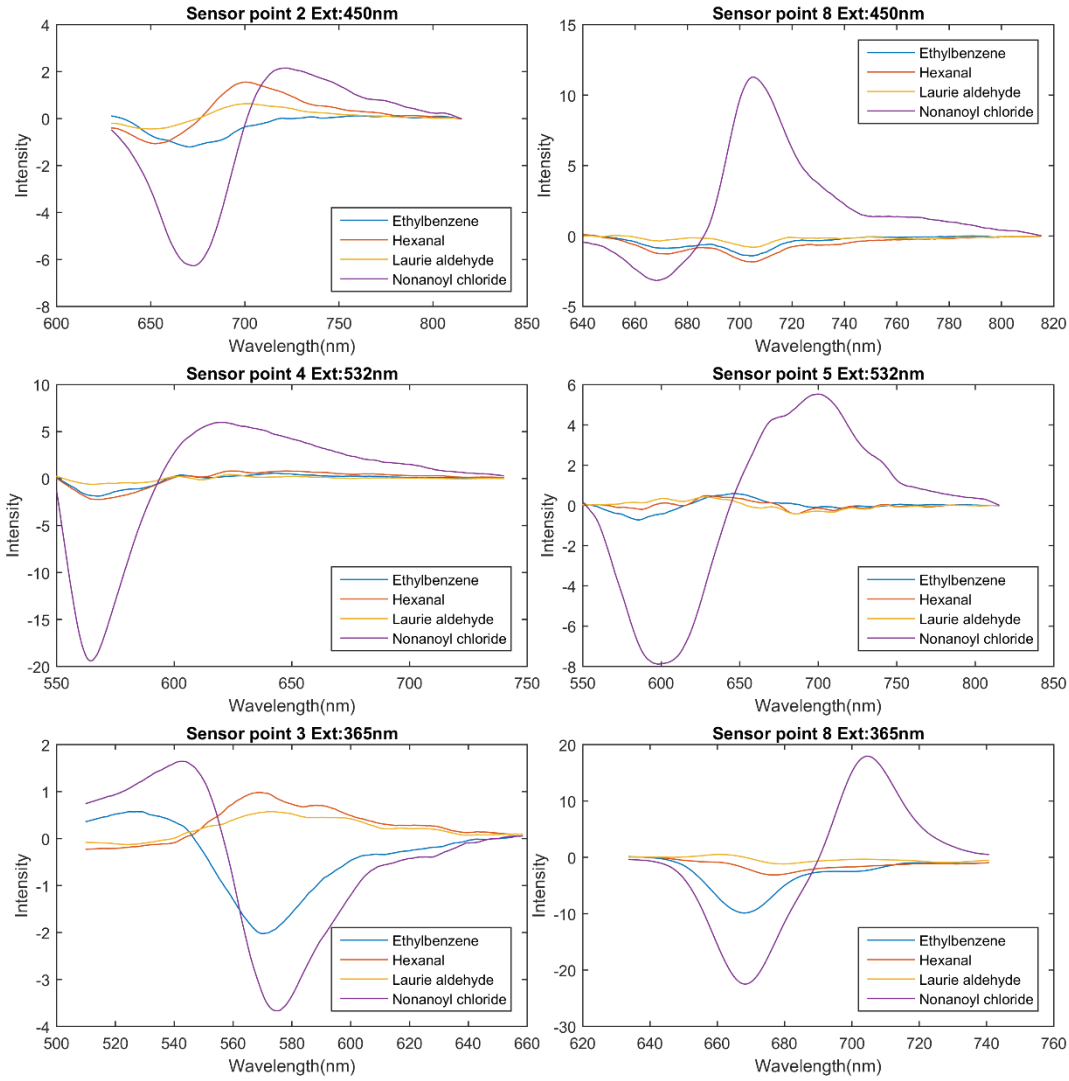


Figure 4-9 Sensor point responses to VOCs under excitation by 450nm, 532nm, and 365nm laser. The differential spectrum has been denoised using Savitzky-Golay filtering [123].

It is very clear that the sensory array has the most significant responses to nonanoyl chloride, then ethylbenzene and hexanal, while the Lauric aldehyde causes the weakest responses. This is due to the chemical properties of those VOCs. Acyl halide (nonanoyl chloride) is the most active compound of the four VOC biomarkers. Usually, aldehyde should be more active than alkylbenzene, but for our sensitive material, the changes of spectral property are more influenced by the π-π conjunction between the VOC molecules and the sensitive materials. Therefore, specific sensor points (see Figure 4-9

point 3 and 8) have higher responses to ethylbenzene than the two aldehydes.

VOC	Test	Sensitivity	Specificity
Ethylbenzene	12	73%	89%
Hexanal	12	38%	94%
Lauric aldehyde	12	100%	90%
Nonanoyl chloride	12	100%	100%

Table 4-2 Leave-one-out cross-validation results of the VOCs test PLS-DA classification model, latent variables = 3, R2 = 0.97, Q2 = 0.83.

As shown in Table 4-2, 12 repeat tests for each urinary bladder cancer biomarkers were conducted. A leave-one-out cross-validation was used for validating the PLS-DA classification model and achieved overall 77.75% sensitivity and 93.25% specificity. Among the four VOC biomarkers, the nonanoyl chloride unexpectedly achieved 100% classification accuracy due to its chemical properties being most different. The hexanal achieved the worst classification accuracy due to the structural and chemical similarity to Lauric aldehyde; the test passed the leave-one-out cross-validation by satisfying the condition of both percentages of explained variances in training (R^2) and cross-validating (Q^2) datasets being larger than 0.4.

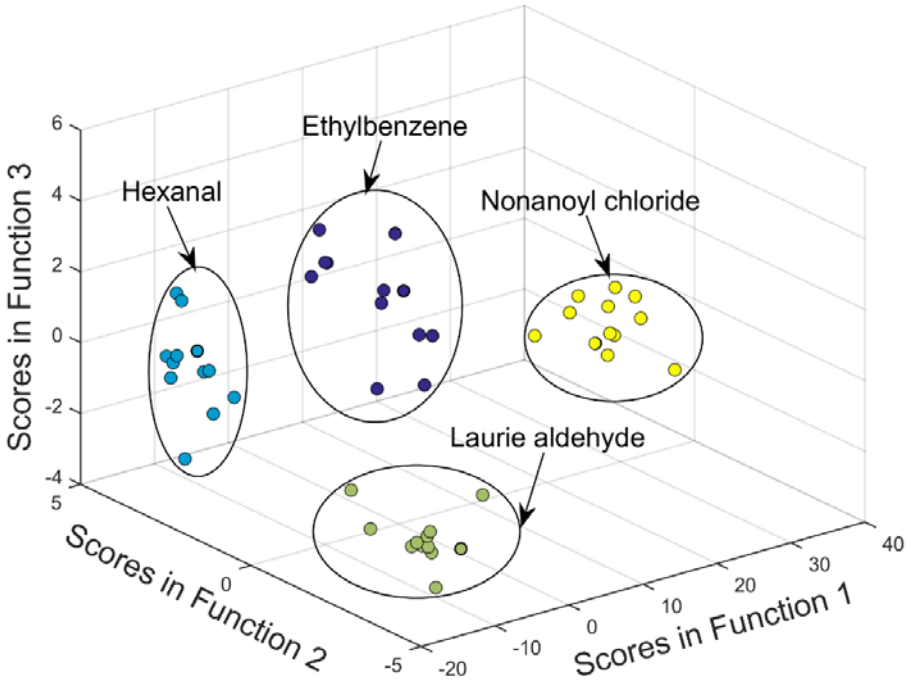


Figure 4-10 FS-LDA plot of urinary VOC bladder cancer biomarkers test (Feature selection)

VOC	Test	Sensitivity	Specificity
Ethylbenzene	12	50%	94%
Hexanal	12	58%	91%
Lauric aldehyde	12	100%	83%
Nonanoyl chloride	12	100%	96%

Table 4-3 Leave-one-out cross-validation results of the VOCs test FS-LDA classification model, R2 = 1.00, Q2 = 0.77

Figure 4-10 shows the feature selection (FS) + LDA score plot of the four urinary VOC

bladder cancer biomarkers tested. From the 3D plot we can clearly see that all four groups are well separated and each element within the groups is well clustered. Three functions explained 100% of variance. The first function explained 96.2% of the variance with canonical correlation of 0.998. The second function explained 2.1% of the variance with canonical correlation of 0.932. The third explained 1.7% of the variance, canonical correlation at 0.915. In combination, these discriminant functions significantly differentiated all VOC groups. The index for describing discrete between groups, Wilk's Lambda $\Lambda < 0.01$, the chi-square test of significance under 123 degrees of freedom: χ^2 (123) = 233.71, $p < 0.01$, by removing the first function indicating that the second and third function can still differentiate the four VOC groups but not significantly, $\Lambda = 0.02$, χ^2 (80) = 94.07, $p = 0.14$. The discriminant function plot showed that: the first function discriminated the nonanoyl chloride group from the other groups, the second function differentiated the hexanal group from the remaining two groups and the third function differentiated the ethylbenzene group from the Laurie aldehyde group.

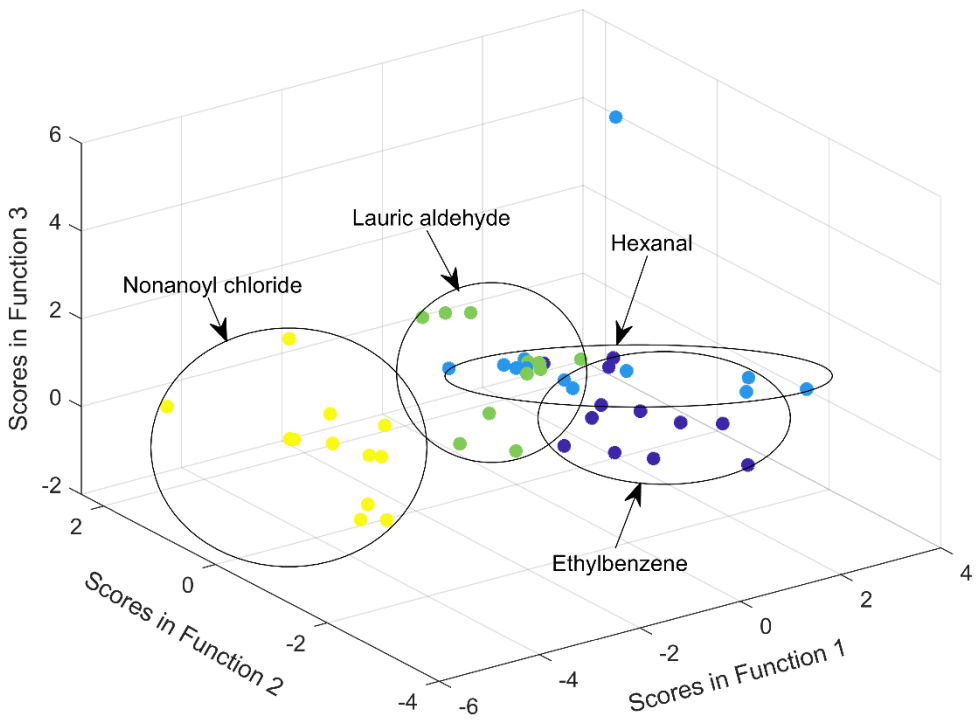


Figure 4-11 PCA-LDA plot of urinary VOC bladder cancer biomarkers test (Feature extraction: 4 components PCA)

VOC	Test	Sensitivity	Specificity
Ethylbenzene	12	75%	92%
Hexanal	12	42%	92%
Lauric aldehyde	12	92%	86%
Nonanoyl chloride	12	100%	100%

Table 4-4 Leave-one-out cross-validation results of the VOCs test PCA-LDA classification model, principle components = 4 (explanation>90%), R2 = 0.79, Q2 = 0.77 (Feature extraction: PCA)

Table 4-4 shows the sensitivities and specificities of PCA-LDA classification model of the four urinary VOC bladder cancer biomarkers tested. Four principle components were kept after PCA (>90% explanation). Compared to FS-LDA (Figure 4-10), the PCA-LDA score plot (Figure 4-11) has worse separation of the hexanal group from ethylbenzene and Lauric aldehyde groups, but it is still easily distinguishable from the remaining three groups. Three functions explained 100% of variance. The first function explained 91.0% of the variance with a canonical correlation of 0.919. The second function explained 7.9% of the variance with a canonical correlation of 0.567. The third explained 1.0% of the variance, canonical correlation at 0.241. However, all three Wilk's Lambda $\Lambda > 0.01$ (0.1, 0.639, 0.942, respectively) indicate that the classification model may have poor performance in separate groups.

This qualitative test shows the ability of different urinary VOC biomarkers in

identifying bladder cancer by the selecting sensitive materials. It also shows the (albeit limited) ability of this method in identifying structurally and chemically similar compounds like hexanal (C6 aldehyde) and Lauric aldehyde (C12 aldehyde). Both feature extraction and feature selection processing before discriminate analysis can achieve good classification performance. The comparison of the performance of three classification models is shown in Table 4-5:

	Sensitivity	Specificity	Notes
PLS-DA	77.75%	93.25%	3 latent variables
PCA-LDA	77.25%	92.50%	4 principle components
FS-LDA	77.00%	91.00%	

Table 4-5 Comparison of the classification performance of PLSDA, PCA-LDA, and FS-LDA. The sensitivities and specificities are generated from leave-one-out cross validation

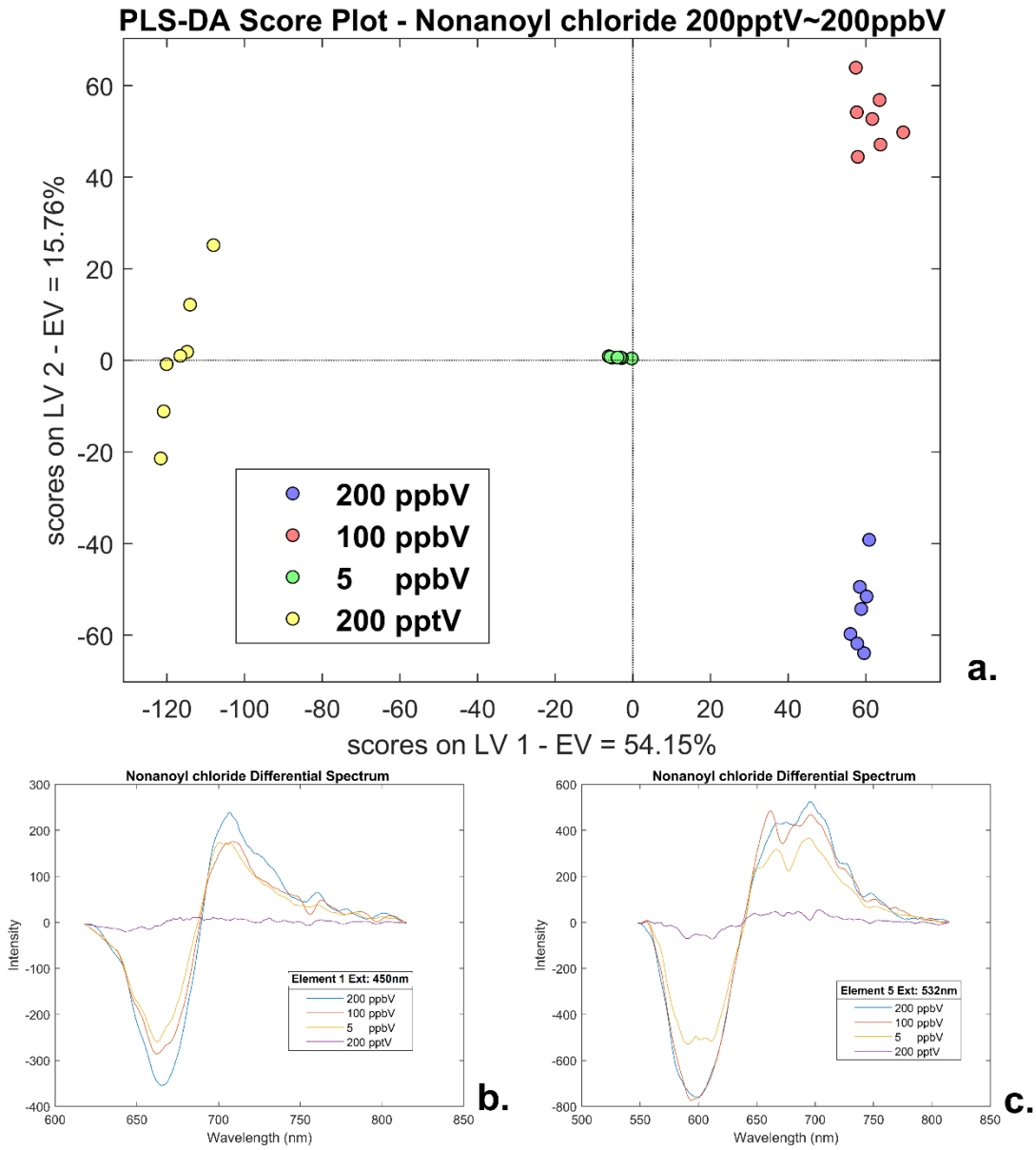


Figure 4-12 a. PLS-DA score plot of various concentration tests of nonanoyl chloride, vapour concentration: 200 ppt ~ 200ppb, reaction time: 120s, carrier gas: nitrogen. b. Differential spectrum of element 1 under 450nm laser excitation when exposed to various concentrations of nonanoyl chloride. c. Differential spectrum of element 5 under 532nm laser excitation when exposed to various concentrations of nonanoyl chloride [123].

Figure 4-12.a shows a clear discrimination grouping of 200 ppb, 100 ppb, 5 ppb, and 200 ppt of nonanoyl chloride by PLS-DA. The degree of dispersion of sample points

within the group shows that, at ppb level, the variation between each measurement is acceptable. Meanwhile, the variation becomes bigger at the sub-ppt level. As shown in Figure 4-12.b and Figure 4.12.c, there are clear changes in spectra intensity that are specific to elements and positively correlated with the concentration of VOC. Within ppbV range, the intensity changes are clear and significant. Contrastingly, once the concentration drops onto pptV (part-per-trillion volume) range, the intensity changes become insignificant and confused with background noise, which makes it harder for the PLSDA algorithm to identify.

4.3.6. Cost efficiency

A typical commercial hand-held e-nose usually costs more than $5000. For example, a ppbRAE Plus system (used for homeland security and indoor air quality) costs about $6200, while a Cyranose 320 system with wide range of chemical detection costs about $8000. For our system, the main cost is from the optical system. An Ocean optic USB4000-FL fluorescence spectrometer costs about $3500 and a low-power LED light source costs around $1000. In combination, the cost of the components of the lab-based system are at the same level as commercially available systems, while the material cost of each sensory array is as low as $0.5 per test. Notably, the aforementioned commercial e-noses are all integrated, which means the sensory setup is fixed and cannot be changed. Meanwhile, our system uses a low-cost sensory array that is customised for urinary VOCs test. This means, if necessary, it can be changed to adapt to other applications, or be upgraded further if new sensitive materials are found, without needing to change the main design of the device.

4.4. Conclusion

In this chapter, the basic principle of sensing technology was reviewed. The artificial olfactory device or e-nose follows the same workflow as the natural animal olfactory system to identify the odours from gas samples. Compare to electronic (MOS, MOSFET, CP etc.) based sensors, non-electronic (QMB, SAW, optical etc.) based sensors usually have a lower detection limit and better sensitivity, but the relevant detecting systems are more complicated and sometimes more expensive.

In this chapter, the fluorescence sensory array built based on selected sensitive materials mentioned in last chapter were tested using VOC test purposed-built device setups. Four literature reviewed urinary VOC biomarkers of bladder cancer: ethylbenzene, hexanal, Lauric aldehyde (dodecanal), and nonanoyl chloride were tested.

Using discrimination analysis with algorithms LDA and PLS-DA, the classification models of four urinary VOC biomarkers were built. In verification of such models, leave-one-out cross validation was used. As a result, all four urinary VOC biomarkers were classified correctly, while the leave-one-out cross validation shows that the PLS-DA classification model achieved overall 77.75% sensitivity and 93.25% specificity. In further tests, the limit of detection of the system was found between low-ppmV to medium-ppbV with acceptable discrimination.

In conclusion, the experiments have shown that the fluorescence VOCs sensitive materials and their sensory array can respond to the presence of urinary VOC biomarkers of bladder cancer to reflect changes in their fluorescence spectra. Further statistical analysis revealed that the classification model can reach a high sensitivity (77.75%) and specificity (93.25%). With suitable statistical tools, the system has the potential to become a cost-effective point-of-care non-invasive diagnostic method for urinary bladder cancer.

5. Clinical Tests of Human Urine Samples

5.1. Introduction

In 2004, Willis et al. published a study using six trained detection dogs to identify human urine from bladder cancer patients and healthy controls [153]. The dogs successfully identified 22 out of 54 cancer patients, better than 14% accuracy rate by chance. Later, several animal olfactory studies were carried out to identify the urine samples from patients with lung [47, 48], breast [47], and prostate [49-51] cancer. Those studies revealed that the odours from human urine have information about certain health conditions, and those odours are detectable.

Analytical chemistry has powerful tools, like gas chromatography (GC), liquid chromatography (LC) and mass spectroscopy (MS), for investigating the detailed composition of testing materials, including urine and the origin of its odour: volatile organic compounds (VOCs). Back in 1999, even earlier than the very first dog test, Spanel et al. used selected ion flow tube mass spectrometry (SIFT-MS) to study the concentration of formaldehyde in the headspace of urine from patients with bladder (14) and prostate cancer (24), as well as 14 healthy volunteers [154]. They found that in both bladder and prostate cancer patients' urine the concentration of formaldehyde was elevated, and the bladder cancer patients' urine had even higher formaldehyde concentrations than prostate cancer patients. The authors speculated that this was because the bladder tumor has closer contact with urine than the prostate tumor, whereby formaldehyde must progress to urine from blood stream.

Jobu et al., using GC-MS, studied urine samples from 9 bladder cancer patients before and after TURBT and 7 healthy controls [145]. They found 12 peak area increased metabolites in patients' urine, and 5 of them were unique to bladder cancer patients' urine when the tumor was not removed. They suggested that the 5 unique VOCs,

namely ethylbenzene, nonanoyl chloride, dodecanal, (Z)-2-nonenal, and 5-dimethyl-3(2H)-isoxazolone could be used as urinary VOCs biomarkers.

Silva CL et al. studied urine samples from 33 cancer patients (14 leukemia, 12 colorectal and 7 lymphoma cancers) and 21 healthy volunteers using GC-qMS analysis [112]. The authors found that a total of 82 volatile metabolites were identified in both groups, while the compound classes and concentrations were dramatically different between cancer patients and healthy volunteers. Benzene derivatives, terpenoids and phenols were the most common classes from the cancer group, whereas ketones and sulphur compounds were the main classes of the control group. 2-methyl-3-phenyl-2-propenal, p-cymene, anisole, 4-methyl-phenol and 1,2-dihydro-1,1,6-trimethyl-naphthalene were observed to have significantly increased concentrations in cancer groups. This study confirmed the hypothesis that biogenic VOCs as cancer biomarkers do exist among various tpes of cancer, and the differences of biogenic VOCs occur in both classes and concentrations.

Issaq HJ et al. used high performance liquid chromatography (HPLC)-MS to identify the urine samples of 48 healthy volunteers and 41 patients with bladder cancer [155]. The authors achieved 100% sensitivity and specificity using the OPLS-DA classification model and 98% sensitivity and 96% specificity using the unsupervised PCA model. Compared to other studies, this research focused more on analysing the VOCs profile of urine samples as a whole instead of targeting any specific VOC biomarkers and achieved remarkably good results, which showed the potential of full urinary VOCs analysis in diagnostic application.

Cauchi M et al. studied urine samples from 72 patients with bladder cancer and 205 controls using GC-MS with three different statistical analyses and machine learning algorithms PLS-DA, random forests, and support vector machine (SVM) [126]. They achieved 89% overall accuracy (90% sensitivity and 88% specificity) in classification

of cancer and non-cancer groups using PLS-DA. This study revealed that pattern recognition algorithms, especially PLS-DA, are feasible in the analysis of high dimensional data like chromatogram or spectra when coupled with suitable cross-validation methods. It was also shown that the “dimensional curse” is not inevitable and building a reliable and robust classification model based on urinary VOCs profile is possible.

Alberice et al. discovered 27 metabolites from the urine samples of 48 bladder cancer patients using LC-MS and capillary electrophoresis (CE)-MS [95]. This study also found that several metabolites (namely betaine, cysteine, histidine and tyrosine) have elevated levels in high-risk compared to low-risk patients, while tryptophan was found to be particularly significant in low-risk patients. They also found that several elevated metabolites in bladder cancer patients’ urine were related to bladder cancer recurrence and regulated by some biophysical pathways linked to well-known cancer mechanisms like immunity and oxygen stress. Although the biomarkers found in this study are not VOCs in the usual sense, the idea of correlating the metabolism products (including VOCs and non-VOCs) with prognosis of cancer is still instructive to our study.

Besides, there are more studies that have found non-VOC urinary metabolite biomarkers using chromatography and MS coupling techniques. Pasikanti et al., using GC-ToFMS, studied urine samples from 24 bladder cancer patients and 51 non-cancer controls, achieving 100% sensitivity and specificity by the OPLS-DA classification model [93]. Then, they compared the spectra with a standard database and found 15 significant differences in urinary metabolites between cancer and non-cancer groups. Jin et al., using high-performance liquid chromatography-quadrupole time-of-flight mass spectrometry (HPLC-QTOFMS), studied urine profiles of 138 bladder cancer patients and 121 controls [156]. They successfully distinguished not only cancer/non-cancer, but also muscle-invasive/non-muscle-invasive cancer groups. Further study

found 12 differential metabolites and some of them (namely phosphoenolpyruvate, pyruvate, and acetyl-CoA) are related to glycolysis and beta-oxidation pathways. Zhou et al. studied urine samples from 50 bladder cancer patients and 35 healthy controls using GC-MS and built a classification model using PCA and PLS-DA [157]. They also collected urine samples from 59 bladder cancer patients and 37 healthy controls and used an external validation set to validate the classification model. They achieved 85.7% sensitivity and specificity in a cross-validation in training set, and 70.5% sensitivity and 70.3% specificity in the external validation set.

A recent study by Rodrigues et al. provided an ex-vivo point of view in the study of VOCs profile of bladder cancer [146]. Two transitional cell carcinoma cell lines (J82, 5637), one squamous cell carcinoma cell line (Scaber), and one non-tumorigenic cell line (SV-HUC-1) were cultured and the VOCs from the extracellular medium were analyzed using GC-MS. Three VOC metabolites, namely 2-pentadecanone, dodecanal and γ-dodecalactone were found to be significant in the identification of bladder cancer and normal cell lines.

In recent years, besides from the approaches of finding more biomarkers and improving the accuracy of GC/LC and MS, studies also introduced e-nose in identifying urinary VOC biomarkers. Weber CM et al., using GC in conjunction with 12 MOS and 10 MOSFET gas sensor arrays, analysed urine samples from 30 bladder cancer patients and 59 controls [54]. The data analysis algorithms used were PCA and PLS-DA. The classification models achieved accuracies varying from 62.2% to 70.0% among identification between cancer groups and a control group of those with different urological conditions. Khalid T et al. analyzed 98 urine samples from 24 patients with bladder cancer and 74 controls with non-malignancy urological symptoms using an in-house fabricated GC-MOS sensor device [124]. The LDA model successfully classified 24/24 of cancer groups and 70/74 of control groups and a leave-one-out

cross-validation of the PLSDA model achieved 95.8% and 94.6% sensitivity and specificity in prediction of the same samples, respectively. Heers et al. tested urine samples from 30 bladder cancer patients and 30 healthy controls with handheld CP e-nose Cyranose™ 320 [125]. The LDA model successfully classified 28/30 cancer groups and 26/30 healthy controls. They also tested the influence of different frozen temperatures (-20°C and -80°C) and concluded that no significant effect was observed. Horstmann et al. reported a pilot study using MOS e-nose in the detection of urinary VOCs from 15 participants with a clinical suspicion of primary or recurrent bladder cancer and from 21 of those without cancer but with benign urological conditions [158]. Histological tests found 8 out 15 cancer group members have urothelial cancer, while the e-nose only detected 5 of them (75% sensitivity), while the system correctly identified 24 out of 28 in the control group (86% specificity).

Last but not least, the recent publication from our research group (Zhu et al. [159]) used a novel fluorescence gas sensor array system and PLS-DA algorithm to test urine samples from 30 bladder cancer patients and 30 age- and gender-matched healthy controls. We successfully identified 54 out of 60 urine samples and achieved 90.00% accuracy with 86.67% sensitivity and 93.33% specificity overall. Furthermore, in classification of high-grade and low-grade bladder cancer, the novel system achieved 76.67% accuracy with 78.57% sensitivity and 75.00% specificity overall. In the following part of this chapter, the study will be detailed, and the updated results will be fully analyzed and discussed.

5.2. Materials & Method

5.2.1. Sample processing module

As described in Chapter 3, the sample processing module for urine test was based on

a bubbler-safety bottle setup and driven by a RS Pro D250 micro diaphragm gas pump (RS Components Ltd, UK), the default flow speed is 380mL/min under 5V working voltage.

5.2.2. Reaction chamber design

For the urine test, a smaller and more compact design of reaction chamber is made based on the VOC test's chambers (Mark. III). The aim of this version is to adapt the nature of low concentration and small amounts of urine vapour. In addition to the smaller box shaped gas chamber, two semi-cylindrical chambers are attached to the inlet and outlet of the gas chamber. The semi-cylindrical chambers are mainly used for pre-mixing the urine vapour at the very beginning of each flow to balance the vapour concentration around the film surface and work as a buffer, preventing the film from contamination from urine spill. In this design, the sensory film is restricted to 25mm×25mm square size and allowed to have a maximum of 25 elements as small as 1mm diameter each.

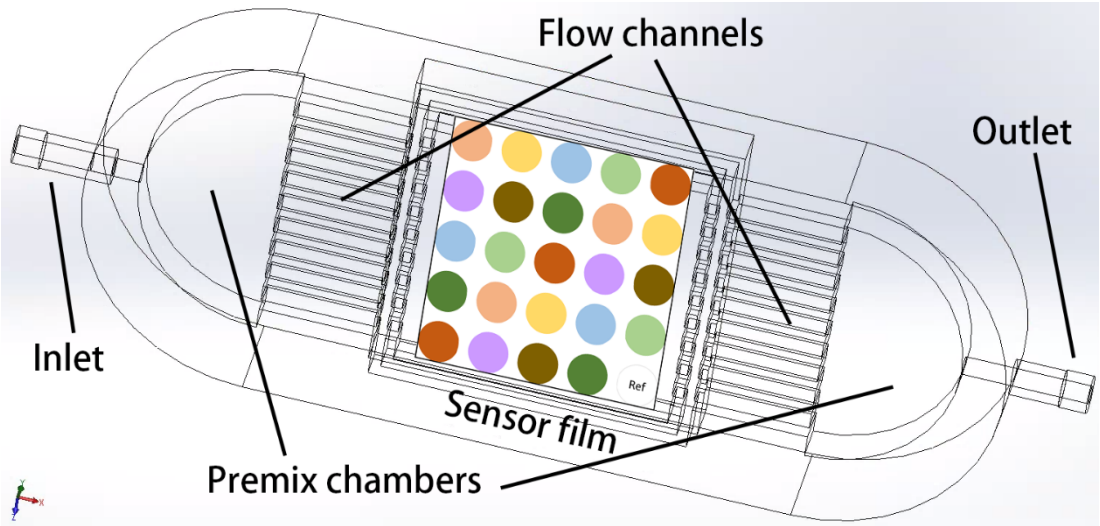


Figure 5-1 Reaction chamber Mark. III

5.2.3. Patient Recruitment

Ethical approval for this study was obtained through the East of Scotland Research Ethics Service (REC: 17/ES/0003) titled "Urinary volatile organic compounds (VOCs) in the diagnosis of urothelial bladder cancer" with sponsorship from TASC (University of Dundee/NHS Tayside). The application was made on 23 December 2016 for 80 participants recruited for 12 months and approved on 22 February 2017. An application for amendment in order to extend the 12-month period was made on 26 March 2018 and approved by the same authority on 27 March 2018. In total, 42 bladder cancer patients and 40 non-cancer controls were recruited in this study. The inclusion criteria were: 16-90 years old, cystoscopically or histologically confirmed new bladder tumour (cancer group), OR 16-90 years old, no urological infections or cancerous disease. All participants received a participant information sheet (PIS) and an introduction letter for this study and asked to provide signed informed consent form on the day of recruitment. The documents of recruitment are attached on Appendix.2.

5.2.4. Sample collecting protocol

There were three different sources for participant recruitment:

1. Patients receiving cystoscopy examination at the endoscopy department in Ninewells hospital;
2. Patients about to receive the transurethral resection of bladder tumour (TURBT) at the surgery department in Ninewells hospital;
3. Patients visiting the urology outpatient clinic in Ninewells hospital.

For the participants from each source, there are slight differences in the protocols for collecting urine sample:

For case 1, patients need to empty their bladder prior to the cystoscopy examination. A small part of the voided urine was used for a dipstick test to check for any urinary infections and the remains were then temporary stored in the urine container covered with lid. When a patient was found to have a tumour in the bladder during the

cystoscopy examination, the clinic nurse introduced him/her to the study and provided the recruitment documents. Once he/she agreed to participant and signed the consent form, his/her urine sample was be transferred to a sealed sterilized urine beaker and frozen in -20 °C immediately.

For case 2, the TURBT waiting list was checked by a senior urologist and surgeon in advance and suitable potential participants were selected based on the patient's clinical report. Once the potential participant arrived to the pre-surgery ward, the clinic nurse introduced the study to him/her and provided the recruitment documents, along with a leaflet describing the procedure of passing the urine to the urine beaker without contamination. Once the patient agreed to participant and signed the consent form, he/she was provided with a sealed sterilized urine beaker to collect the urine sample. Successfully collected urine samples were then frozen under -20 °C immediately.

For case 3, most of the controls came from this source. The clinic nurse was in charge of filtering the potential participants from patients visiting the urology outpatient department in Ninewells hospital. The inclusion criteria were: age and gender matching the existing cancer group and no urological infections or known cancerous disease. Similar to case 2, the potential participants were given the recruitment documents and guidance notes on collecting urine samples. Upon signing the consent form, the urine samples were collected and stored immediately.

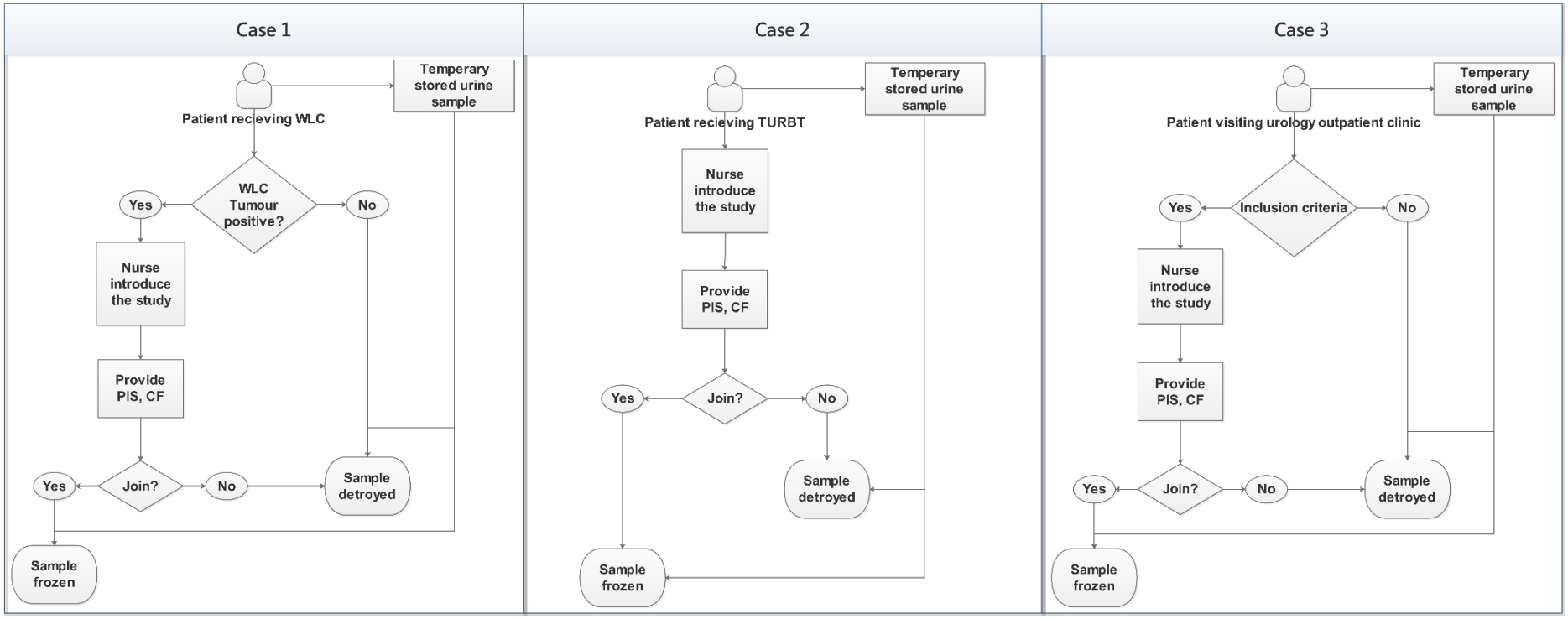


Figure 5.5-2 Recruitment flowchart for each case. WLC: white light cystoscopy; PIS: patient information sheet; CF: consent form; TURBT: transurethral resection of bladder tumour.

5.2.5. System workflow

On the day of testing, frozen urine samples were moved out of the -20°C freezer to defrost in a 37°C water bath. Simultaneously, the “before” spectra of the sensor array film were measured using the LAKK-M device. Then, the defrosted urine sample was centrifuged at 5000rpm for 5min to remove the sediment. 1.5mL of the urine supernatant was then transferred to a 15mL tube and 1.5mL of 1M NaOH solution was added and incubated for 30min in the 37°C water bath. After incubation, the tube with urine solution was mounted to the sample processing module, the pump then turned on for 2min to circulate the urine vapour within the reaction chamber. After “pre-mix”, the sensory array film was put into the reaction chamber and the pump was turned on for another 2min to allow the sensitive materials on the film to interact with the urinary VOCs. After reaction, the sensory array film was measured again for “after” spectra, and the tubing and reaction chamber were purged with clean air to remove any residuals.

5.3. Results & Discussion

For all cases, urine collection was part of the standard operating procedure, no additional procedures were added to the routine. All the recruitment documents were paper archived: the signed consent forms were kept by research team, while the study introduction letters and PIS were archived in participants’ clinic notes for future reference. All participants’ age, gender and medication history were collected and study-related diagnoses (including histopathological results) of cancer patients were authorised to be obtained by the research group. All information was stored in a password protected NHS computer and only authorised members of the research team

were able to access it.

5.3.1. Patient demographic

Table 5-1 shows the original patient demographic of this study. In total, 83 participants were recruited, with 42 of them having suspected bladder cancer (tumour found through imaging examination before TURBT and biopsy examination) and the remaining 41 being healthy volunteer controls. 10 out of 42 cancer group members were recruited from a pre-surgery ward in Ninewells hospital, while the remaining 32 were recruited from the endoscopy department in same hospital. All the control group members were recruited from Ninewells hospital's urology outpatient clinic.

Characteristic	Cancer (n=42)	Control (n=41)
Age (Years):		
Means	72.88	67.73
Range	33-92	25-85
Gender:		
Male	33 (78.6%)	35 (85.4%)
Female	9 (21.4%)	6 (14.6%)

Table 5-1 Participant demographics of original recruitment

The average age of the cancer group was 72.88 years old, ranging from 33 to 92 years old, with a male to female ratio 33:9, mean ages 73.06 and 72.22, respectively. The average age of the control group was 67.45 years old and ranged from 25 to 85 with a male to female ratio of 34:6, aged 71.74 and 43.17 years old, respectively.

All urine samples from cancer group members were collected before they had TURBT and biopsy. The pathological results of the biopsy were analysed by the pathology department in Ninewells hospital and discussed by doctors during routine multidisciplinary team (MDT) meetings in which the clinical diagnosis was given. The

diagnosis opinions recorded in MDT meeting documents were considered as a decisional diagnosis and were used as the “true condition” for discrimination analysis and further discussion.

Among all 42 cancer group members, 1 patient (Research code: BC5) was diagnosed as “No evidence of malignancy” by pathological examination and was removed from cancer group. Three patients (Research codes: BC10, BC37, BC40) failed to provide sufficient or qualified samples for pathological examinations hence no pathological opinions were given; those 3 patients were removed from the cancer group as well.

Updated participant demographic and cancer groups’ pathological diagnosis opinions are shown in Table 5-2:

Characteristic	Cancer (n=38)	Control (n=41)
Age (Years):		
Means	72.74	67.73
Range	33-92	25-85
Gender:		
Male	31 (81.6%)	35 (85.4%)
Female	7 (18.4%)	6 (14.6%)
Grading:		
G1	1 (2.6%)	
G2-Low	19 (50.0%)	
G2-High	9 (23.7%)	
G3	9 (23.7%)	
Staging:		
pTa	27 (71.0%)	
pT1	6 (15.8%)	
pT2	5 (13.2%)	

Table 5-2 Updated participant demographic and cancer group members pathological diagnosis results

5.3.2. Urine sample preparation

The vapour-liquid equilibrium of VOCs in water solution follows Raoult’s law:

$$\mathrm{p} = p_A^* x_A + p_B^* x_B + \cdots$$

It states that the partial vapor pressure of each component p_i of an ideal mixture of liquids is equal to the vapour pressure of the pure component p_i^* multiplied by its mole fraction x_i in the mixture.

In reality, because of the interactions between components, the vapour-liquid equilibrium of the real solution does not always follow Raoult's law, as shown in Figure 5-3. The total vapour pressure of a solution may have positive or negative deviation from Raoult's law, meaning the vapour-liquid equilibrium may be different to the calculation. When the interaction force between different molecular components in the solution is lower than the force between dissimilar molecular components, the pressure and polarity will cause more molecules to escape the solution, therefore the vapour pressure would be greater than the Raoult's law calculation, showing positive deviation. Opposingly, the negative deviation of Raoult's law happens when the adhesive and cohesive forces between the mixture are stronger than the purified solution; the molecules are more likely to stay in the solution, causing lower vapour pressure and lesser gaseous state molecules to release.

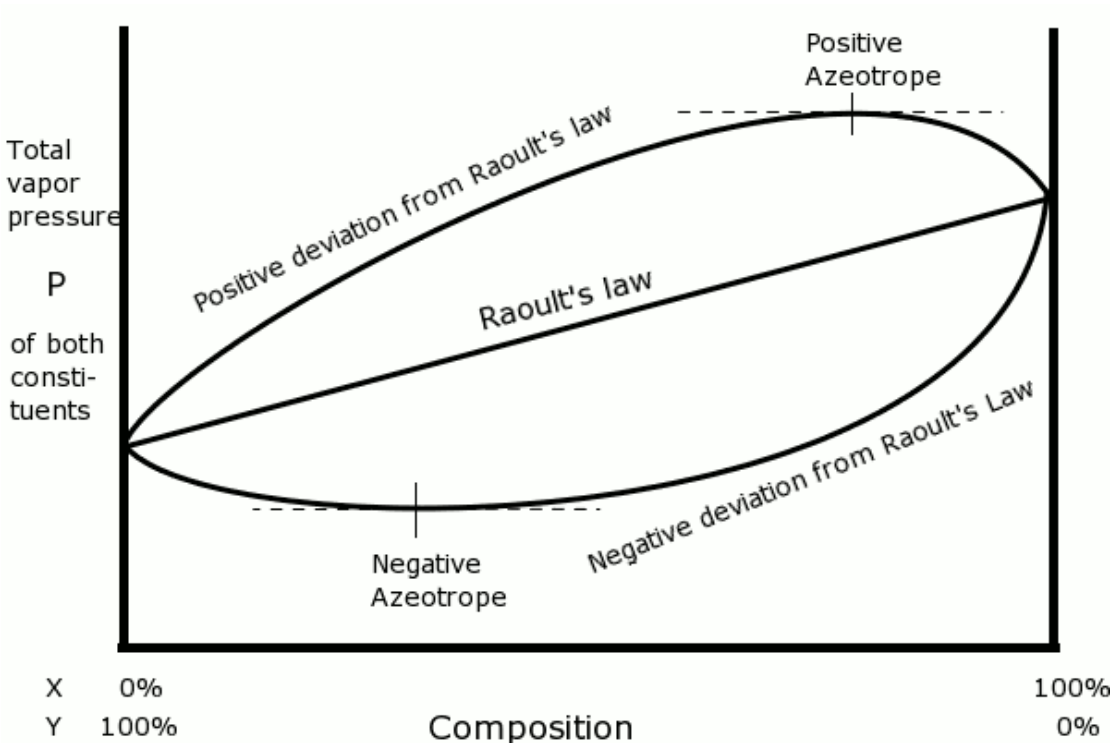


Figure 5-3 Vapour pressure curves of different Azeotrope solution compared to ideal Raoult's law[162].

One of the most important influencing factors of molecular interaction forces in solution is the ionic strength. The ionic strength is the concentration of ions in the solution, it can affect many properties of the solution such as the dissociation constant or the solubility of different compounds. Usually, increasing the ionic strength of the solution can enhance the VOCs released because the charge around the ions in the solution can attract the component molecules in the solution and reduce the interaction forces between existing solution components, thus enhancing the positive deviation of Raoult’s law. The most commonly used ionic additions for urine analysis are salt (e.g. NaCl), acid (e.g. HCl) and alkali (e.g. NaOH) additions.

For the acid and alkali additions, another important influencing factor is pH value. Change of pH value can largely influence the VOC’s profile of urine because the H^+ and OH^- ions will push the ionisation equilibrium of a weak acid or base toward the unionised forms, thus enhancing the corresponding compounds released from the

solution. In addition, the acid or base condition may promote further reactions with existing compounds in the solution and produce secondary volatile or non-volatile compounds [160]. For example, esters are expected to hydrolyse greater under basic conditions, thus leading to more alcohols and less esters releases from the solution. In general, acid and alkali additions cause more species and higher concentrations of VOCs release into the urine vapour than salt additions [161].

In this study, all the urine samples were treated with the same amount of basic solution (1:1 1M NaOH solution) to ensure the same condition, no further VOC enhancement was carried out. Although acid addition may result in higher enhancement of VOC's urine vapour, the volatile nature of acid itself may bring interference to pH sensitive materials in our sensory array. Besides, the hydrocarbon enhancement by acid addition is harder for our sensory array to detect, while the alcohol enhancement by alkali addition is an ideal detection target for the sensory array we made.

5.3.3. Effect of storage and freezing-thawing

Smith et al. studied the VOCs profile when it was fresh and after 2hrs, 4hrs, 6hrs and 8hrs of storage at room temperature, and subsequently frozen-thawed urine samples using GC-MS [161]. They found that deep freezing can preserve the VOC's profile of the urine sample without appreciable loss of volatility. The time between urine collection and freezing did not affect the VOC's profile, but they observed some peak area changes after 8hrs of storage at room temperature prior to freezing and estimated that this was because of microbial production or utilisation of urine components. This study suggests that the urine sample can be stored for a short time before freezing, and that frozen-thawed urine samples are still useable for VOC's analysis.

Due to the nature of our methods, it was not possible to carry out a quantitative study

of the VOC's profile differences between fresh and frozen-thawed urine samples like GC-MS does, but the comparison of the response signals of the device is still meaningful. As shown in Figure 5-4, most of the sensor elements have no differences in response signals between fresh and frozen-thawed urine samples, though element 3, 4, and 8 showing some peak shifts in the differential spectra. Those signals are not gone through denoising filter in order to maintain the raw responses from fluorescence sensors. The differences of element 3, 4, and 8 are mainly because the interference from background noise, the level of shifts is very weak therefore cannot be distinguished. In combination of conclusion from previous studies and our tests, we can think that the impact to VOCs profiles of human urine from single frozen-thawed cycle is negligible. However repeated freezing-thawing are not suggested, it because the microbes may grow and produce VOCs during thawing process and cause interference to the urinary VOC profiles. This is the same reason for avoids long delays prior to the urine sample freezing (longer than 8 hours of storage in room temperature), microbes may grow and produce metabolites, including VOCs. In this project, all urine samples were frozen within 2 hours of collection, and were tested within one month of storage.

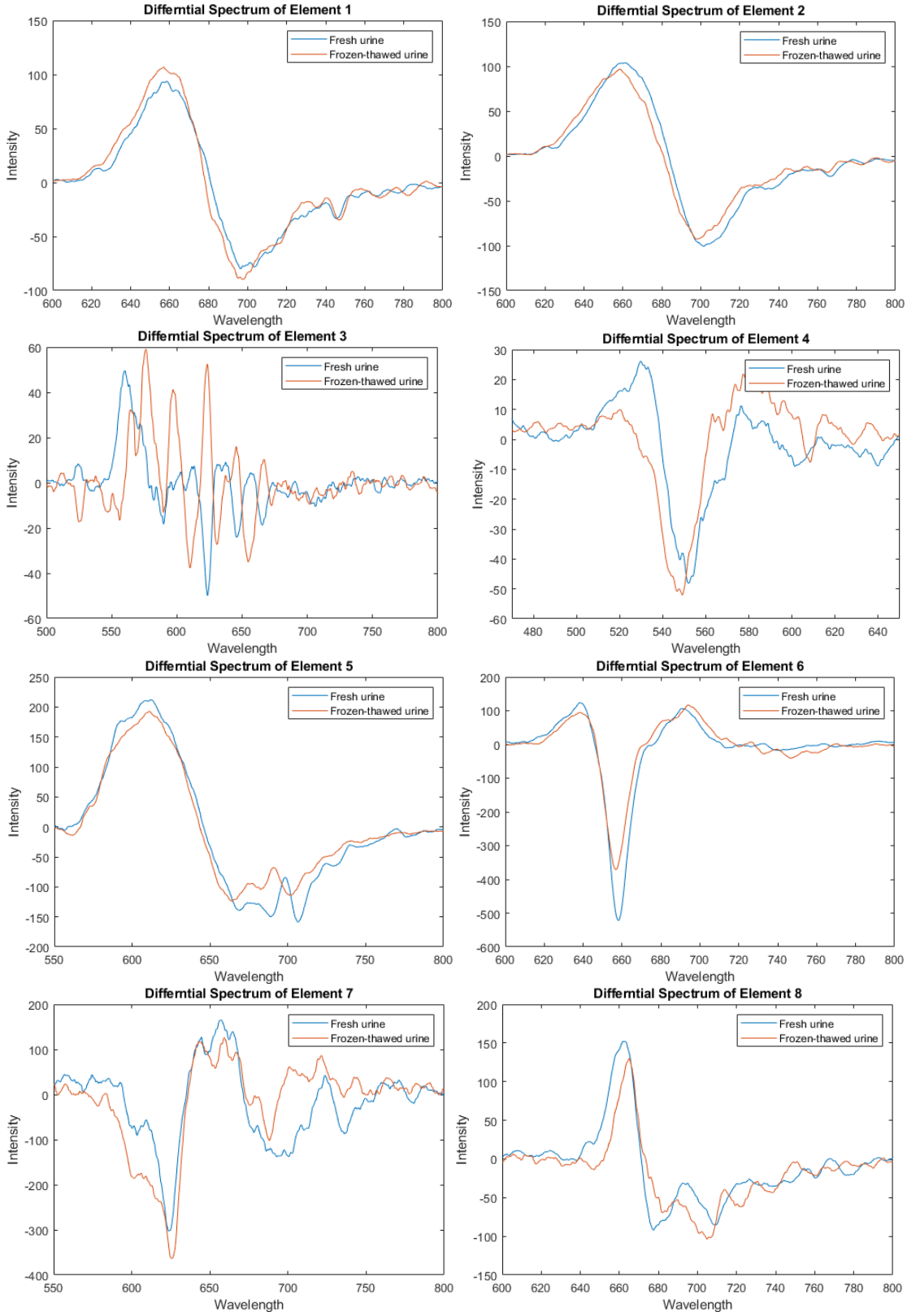

Differntial Spectrum of Element 1
Differntial Spectrum of Element 2
Differntial Spectrum of Element 3
Differntial Spectrum of Element 4
Differntial Spectrum of Element 5
Differntial Spectrum of Element 6
Differntial Spectrum of Element 7
Differntial Spectrum of Element 8
Fresh urine
Frozen-thawed urine
Intensity
Wavelength

Figure 5-4 Comparison of signal profile of fresh and frozen-thawed urine samples

5.3.4. Discrimination analysis

The discrimination analysis methods are the same as VOCs tests: feature selection/extraction and LDA, serial differential spectrum PLS-DA for a classification model and leave-one-out cross validation for validating such model.

In short, the feature selection/extraction are two methods to reduce the sample dimensions and improve the model calculation performance. The feature selection method in this study was signal characteristics extraction. It described the differential spectrum through several characteristics: Maximum Peak Value (MXV) and Wavelength (MXW), Minimum Peak Value (MNV) and Wavelength (MNW), and Peak Area (PA). The spectrum signals characteristic dataset X_{SSC} was a combination of each elements' differential spectrum signal characteristics on the sensory array and was then used as an input dataset for LDA. The feature extraction method used in this study was PCA. The PCA described the features of the differential spectrum by linear transformation of the raw dataset and generated a new smaller dataset that could describe the original dataset to the greatest extent. The new dataset, reduced *score* matrix $X_{rdcscore}$, was then used as an input dataset for LDA modelling.

LDA was used for building the classification model. It used a linear combination of features to generate a function that aimed to maximise the differences between each group and to minimise the differences within each group. The function, called cost function, can be used for predicting the classification of unknown samples.

PLS-DA was used independently without prior feature selection or extraction, it is the combination of partial least squares regression with discrimination analysis. The PLS algorithm works as feature extraction and noise reduction and its properties are ideal for a dataset that has more variables than sample numbers, like ours.

5.3.4.1. Classification of cancer and non-cancer groups (diagnosis)

The classification model of cancer and non-cancer is the approach is to distinguish the bladder cancer group from healthy control group, it is the primary target of the study. To achieve that, all three previous discussed discriminate analysis algorithms: FS-LDA, PCA-LDA, PLS-DA were used for building the classification model.

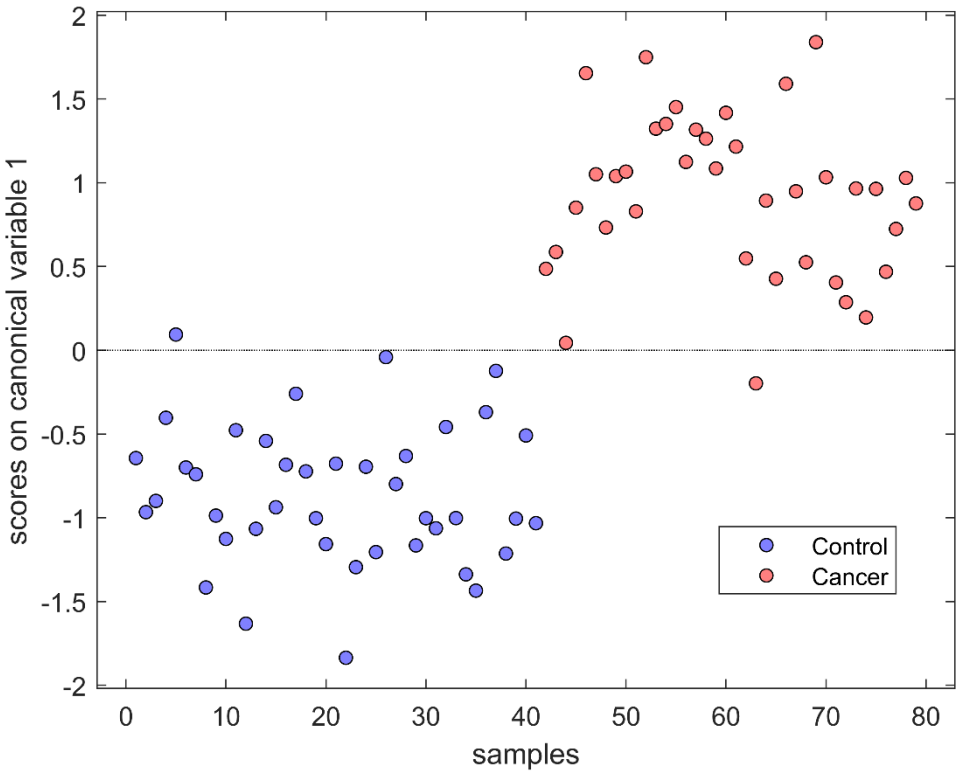


Figure 5-5 Canonical score plot of FS-LDA classification model

Total population 79	Bladder Cancer BC 38	Health Control HC 41	Prevalence 0.481012658	**Accuracy (ACC)** 82.28%
Predicted BC 34	True positive 29	False positive 5	**Positive predictive value (PPV)** 0.852941176	False discovery rate (FDR) 14.71%
Predicted HC 45	False negative 9	True negative 36	False omission rate (FOR) 20.00%	**Negative predictive value (NPV)** 0.8
	True positive rate (TPR) 76.32%	False positive rate (FPR) 12.20%	Positive likelihood ratio (LR+) 6.257894737	Diagnostic odds ratio (DOR) 23.2
	False negative rate (FNR) 23.68%	**True negative rate (TNR)** 87.80%	Negative likelihood ratio (LR−) 0.269736842	F1 score 0.805555556

Table 5-3 Leave-one-out cross-validation results of FS-LDA model

Figure 5-5 and Table 5-3 show the classification results of the FS-LDA model, cross-validated by leave-one-out cross validation. The FS-LDA model successfully identified 65 out of 79 urine samples with an overall accuracy of 82.28%. The cross-validation sensitivity and specificity were 76.32% and 87.80%, respectively.

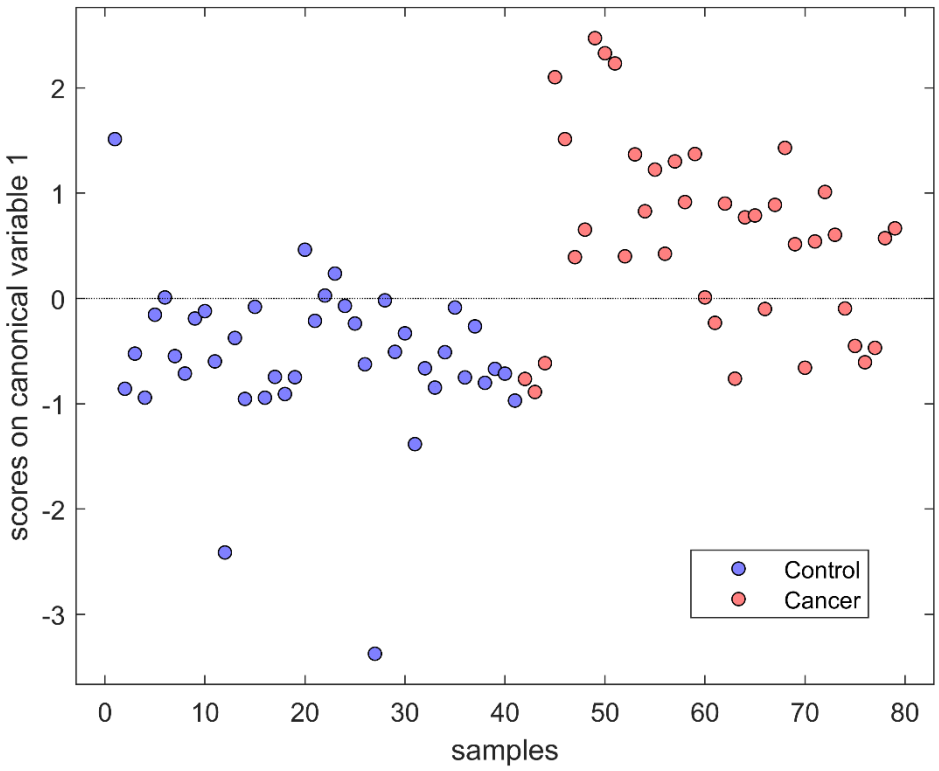


Figure 5-6 Score plot PCA-LDA classification model, principle components = 5

Total population 79	Bladder Cancer BC 38	Health Control HC 41	Prevalence 0.481012658	**Accuracy (ACC)** 78.48%
Predicted BC 31	True positive 26	False positive 5	**Positive predictive value (PPV)** 0.838709677	False discovery rate (FDR) 16.13%
Predicted HC 48	False negative 12	True negative 36	False omission rate (FOR) 25.00%	**Negative predictive value (NPV)** 0.75
	True positive rate (TPR) 68.42%	False positive rate (FPR) 12.20%	Positive likelihood ratio (LR+) 5.610526316	Diagnostic odds ratio (DOR) 15.6
	False negative rate (FNR) 31.58%	**True negative rate (TNR)** 87.80%	Negative likelihood ratio (LR−) 0.359649123	F1 score 0.753623188

Table 5-4 Leave-one-out cross-validation results of PCA-LDA classification model (principle components = 5)

Figure 5-6 and Table 5-4 show the classification results of PCA-LDA model, cross-validated by leave-one-out cross validation. The PCA-LDA model successfully identified 62 out of 79 urine samples with an overall accuracy of 78.48%. The cross-validation sensitivity and specificity were 68.42% and 87.80%, respectively.

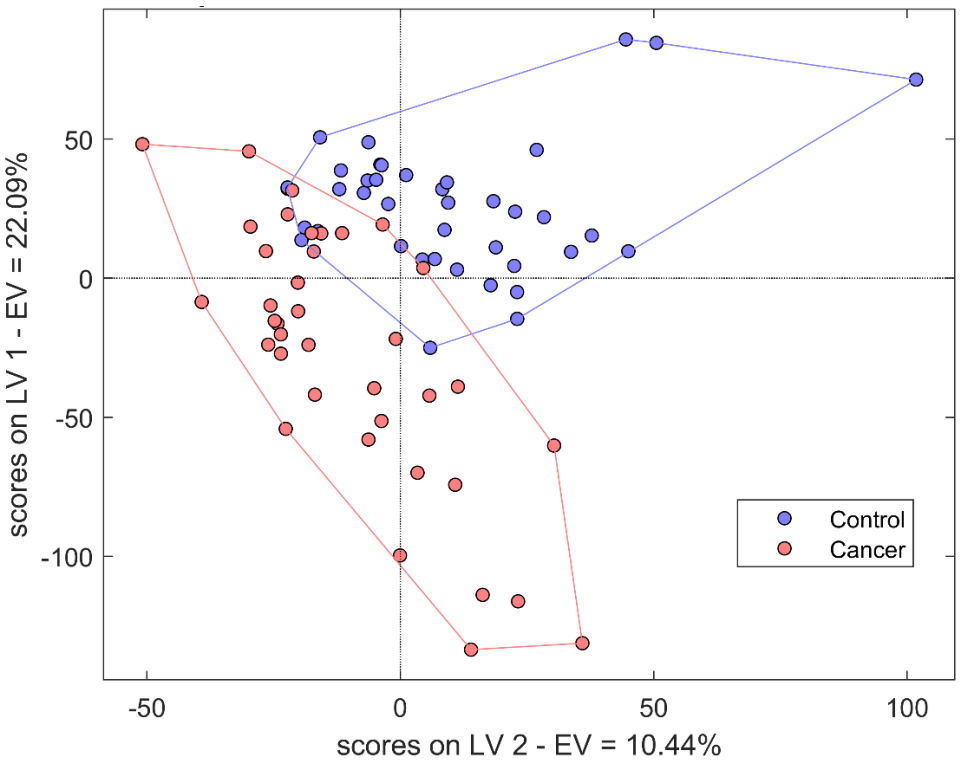


Figure 5-7 Score plot of PLSDA classification model, latent variables = 4

Total population 79	Bladder Cancer BC 38	Health Control HC 41	Prevalence 0.481012658	**Accuracy (ACC)** 86.08%
Predicted BC 37	True positive 32	False positive 5	**Positive predictive value (PPV)** 0.864864865	False discovery rate (FDR) 13.51%
Predicted HC 42	False negative 6	True negative 36	False omission rate (FOR) 14.29%	**Negative predictive value (NPV)** 0.857142857
	True positive rate (TPR) 84.21%	False positive rate (FPR) 12.20%	Positive likelihood ratio (LR+) 6.905263158	Diagnostic odds ratio (DOR) 38.4
	False negative rate (FNR) 15.79%	**True negative rate (TNR)** 87.80%	Negative likelihood ratio (LR−) 0.179824561	F1 score 0.853333333

Table 5-5 Leave-one-out cross-validation results of PLSDA classification model (4 latent variables)

Table 5-5 shows the classification and leave-one-out cross validation results of the PLSDA model and the number of latent variables (4). As shown above, the PLSDA model successfully identified 68 out of 79 urine samples with an overall accuracy of 86.08%. The cross-validation sensitivity and specificity were 84.21% and 87.80%, respectively. As shown in Figure 5-7, the score plot can clearly separate the two groups with very little overlay.

To further validate the model, a 20% random Mote Carlo cross-validation was performed. This method randomly divides the dataset into training (80% of the samples) and testing (20% of samples) datasets, and the algorithm attempts to build the model using only the training dataset and to verify the model using the testing dataset. This is repeated 1,000 times. This is a strong cross-validation method and can largely challenge the prediction ability of the model. Compared to other cross-validation methods, the major disadvantage of random Mote Carlo is that it requires a large amount of calculation of resources and is more time consuming.

In a total of 1,000 iterations of 16,000 predictions, the PLSDA model hit 13,081 times correctly with 81.76% overall accuracy, 77.42% sensitivity and 85.82% specificity. The detailed confusion matrix is shown in Table 5-6.

Total population 16000	Bladder Cancer BC 7742	Health Control HC 8258	Prevalence 0.483875	**Accuracy (ACC)** 81.76%
Predicted BC 7165	True positive 5994	False positive 1171	**Positive predictive value (PPV)** 0.836566643	False discovery rate (FDR) 16.34%
Predicted HC 8835	False negative 1748	True negative 7087	False omission rate (FOR) 19.78%	**Negative predictive value (NPV)** 0.802150538
	True positive rate (TPR) 77.42%	False positive rate (FPR) 14.18%	Positive likelihood ratio (LR+) 5.459860607	Diagnostic odds ratio (DOR) 20.75299818
	False negative rate (FNR) 22.58%	**True negative rate (TNR)** 85.82%	Negative likelihood ratio (LR−) 0.263087799	F1 score 0.804185953

Table 5-6 Confusion matrix of 20% random Mote Carlo cross-validation of PLSDA model, latent variables = 4

In summary, three discriminate analysis were performed using same dataset, by comparison, the PLSDA model achieved highest sensitivity and specificity, FS-LDA second, and the PCA-LDA third. One of the main concerns of discriminate analysis is the possibility of overfitting, which means the discriminate algorithm picked up too many unnecessary features of input data and overwhelmed the real difference. To prevent that from happening, cross-validation is needed.

The leave-one-out cross-validation is a trade-off of bias and variance: each time it uses all except one sample to train the model, which makes it almost has no bias, but when the sample number is small, it tends to be easier to having overfitting, therefore the 20% random Mote Carlo cross-validation was used to further investigate whether the overfitting happened or not. Luckily, both the sensitivity and specificity have not fallen too much, which means that the possibility of overfitting is minimized.

One of the interesting phenomena is that all three discriminate analysis achieved the same specificity, with 5 false positive hits, the study code of the five easy-to-confused samples are: C2, C7, C23, C34, C39. All of them are men, average age 72.2 (62-84), which is nearly 10 years older than the average ages of all control group, this suggests those five men might have some sort of metabolic disorders or hidden influence factors yet to know and requires further studies. Although the exact reasons of why the urine samples from these five participants have similar sensor responds to cancer patients is still unclear, this results also proved that all three discriminate algorithms can pick up the same special features from the spectral data that different from the labeled groups, hence mutually verified the ability of choosing most significant features of the three algorithms.

5.3.4.2. Classification of high-grade and low-grade cancer groups (grading)

As described in Chapter 1, the biogenic VOC's profile may express differently among different phenotypes of cancer cells, which means the high-grade and low-grade tumours may have different VOCs profiles and that this might be detected by our system.

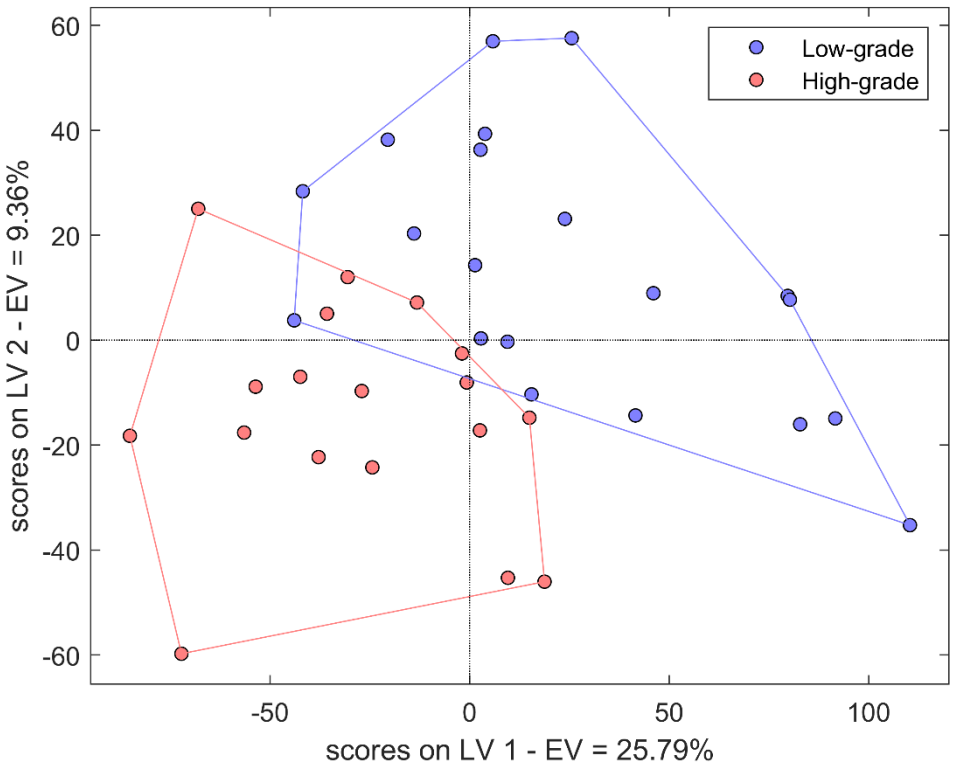


Figure 5-8 Score plot of PLSDA grading model, latent variables = 10

Total population 38	High Grade HG 18	Low Grade LG 20	Prevalence 0.473684211	**Accuracy (ACC)** 71.05%
Predicted HG 17	True positive 12	False positive 5	**Positive predictive value (PPV)** 0.705882353	False discovery rate (FDR) 29.41%
Predicted LG 21	False negative 6	True negative 15	False omission rate (FOR) 28.57%	**Negative predictive value (NPV)** 0.714285714
	True positive rate (TPR) 66.67%	False positive rate (FPR) 25.00%	Positive likelihood ratio (LR+) 2.666666667	Diagnostic odds ratio (DOR) 6
	False negative rate (FNR) 33.33%	**True negative rate (TNR)** 75.00%	Negative likelihood ratio (LR−) 0.444444444	F1 score 0.685714286

Table 5-7 Leave-one-out cross-validation results of PLSDA grading model (10 latent variables)

As shown in Table 5-7, the PLSDA grading model successfully identified 27 out of 38 urine samples with an overall accuracy of 71.05% and cross-validation sensitivity and specificity of 66.67% and 75.00%, respectively. However, with such a small sample size, the model may have overfitted, therefore further cross-validation is needed.

Total population 5581	High Grade HG 2545	Low Grade LG 3036	Prevalence 0.456011467	**Accuracy (ACC)** 71.74%
Predicted HG 2502	True positive 1735	False positive 767	**Positive predictive value (PPV)** 0.693445244	False discovery rate (FDR) 30.66%
Predicted LG 3079	False negative 810	True negative 2269	False omission rate (FOR) 26.31%	**Negative predictive value (NPV)** 0.736927574
	True positive rate (TPR) 68.17%	False positive rate (FPR) 25.26%	Positive likelihood ratio (LR+) 2.698473116	Diagnostic odds ratio (DOR) 6.336560594
	False negative rate (FNR) 31.83%	**True negative rate (TNR)** 74.74%	Negative likelihood ratio (LR−) 0.425857699	F1 score 0.687537151

Table 5-8 Confusion matrix of 20% random Mote Carlo cross-validation of PLSDA grading model, latent variables = 10, 3996 out of 8000 samples failed to assigning groups.

Similar to the classification model, a 20% random Mote Carlo cross-validation was used for validating the PLSDA grading model. In a total of 1,000 iterations of 8,000

grading predictions, the PLSDA model hits 4,004 times correctly with 3996 failed attempts to assign groups. Among the successfully assigned samples, the grading prediction hit 68.17% sensitivity and 74.74% specificity. The detailed confusion matrix is shown in Table 5-8. This result suggests that the sample size is too small for training the grading model, therefore nearly 50% of the testing samples cannot be assigned to groups. This also means the grading model still has room for improvement in future larger scale population studies.

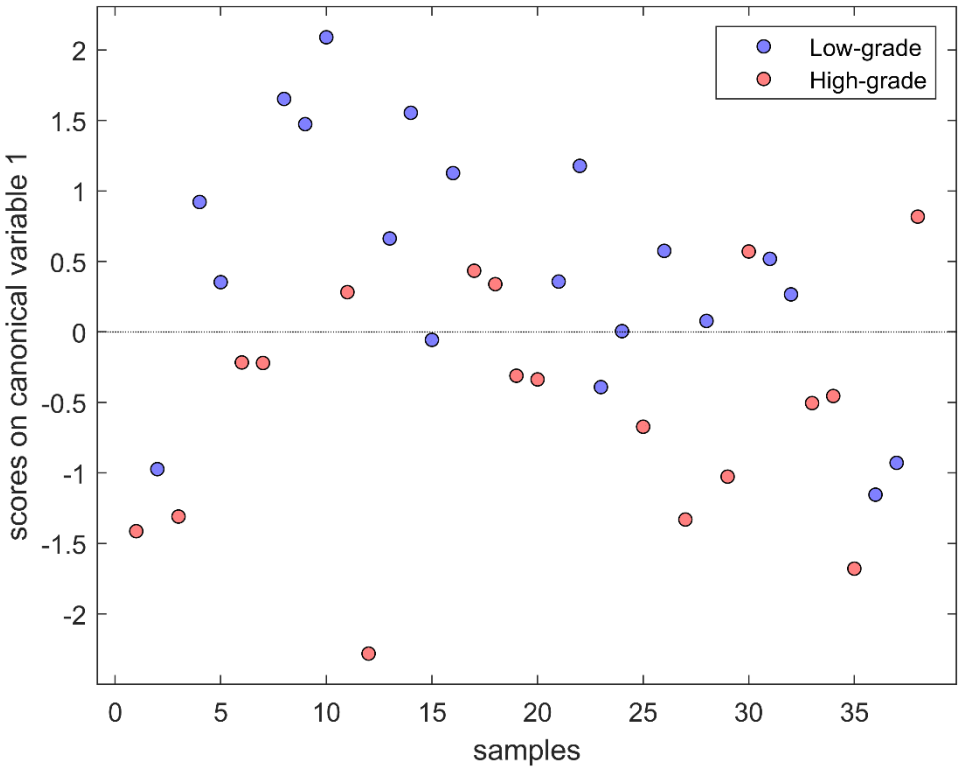


Figure 5-9 Canonical score plot of PCA-LDA grading model, principle components = 3

Total population 38	High Grade HG 18	Low Grade LG 20	Prevalence 0.473684211	**Accuracy (ACC)** 71.05%
Predicted HG 17	True positive 12	False positive 5	**Positive predictive value (PPV)** 0.705882353	False discovery rate (FDR) 29.41%
Predicted LG 21	False negative 6	True negative 15	False omission rate (FOR) 28.57%	**Negative predictive value (NPV)** 0.714285714
	True positive rate (TPR) 66.67%	False positive rate (FPR) 25.00%	Positive likelihood ratio (LR+) 2.666666667	Diagnostic odds ratio (DOR) 6
	False negative rate (FNR) 33.33%	**True negative rate (TNR)** 75.00%	Negative likelihood ratio (LR−) 0.444444444	F1 score 0.685714286

Table 5-9 Leave-one-out cross-validation results of PCA-LDA grading model (3 principle components)

As shown in Table 5-9, the 3 components of the PCA-LDA grading model achieved exactly the same cross-validation results as the PLS-DA grading model. However, the sample size limited the explanatory power of the model, therefore further larger population studies are needed exploring the potential of grading based on urinary VOCs detection.

5.3.4.3. Classification of invasiveness of cancer groups (staging)

Among all 38 cancer group members (see Table 5-2), there were 27 members in Ta stage when urine samples were collected, 6 in T1 and 5 in T2. Here, we divide the samples in two groups: a non-invasive group (including all members with pTa diagnosis) and an invasive group (including members with pT1 and pT2 diagnosis). Since the non-invasive group are the majority (more than 70%), the classification model was unbalanced and therefore the results are only for reference. The potential of cancer staging based on urinary VOCs detection needs to be confirmed with larger cohort studies in the future.

Total population 38	Invasive 11	Non-invasive 27	Prevalence 0.289473684	**Accuracy (ACC)** 57.89%
Predicted Inv 11	True positive 3	False positive 8	**Positive predictive value (PPV)** 0.272727273	False discovery rate (FDR) 72.73%
Predicted Ninv 27	False negative 8	True negative 19	False omission rate (FOR) 29.63%	**Negative predictive value (NPV)** 0.703703704
	True positive rate (TPR) 27.27%	False positive rate (FPR) 29.63%	Positive likelihood ratio (LR+) 0.920454545	Diagnostic odds ratio (DOR) 0.890625
	False negative rate (FNR) 72.73%	**True negative rate (TNR)** 70.37%	Negative likelihood ratio (LR−) 1.033492823	F1 score 0.272727273

Table 5-10 Leave-one-out cross-validation results of PLSDA staging model (4 latent variables)

From a biophysical point of view, the biogenic VOCs profile are correlated with the metabolism alternation of the cancer cells. Cancer cells with different aggressiveness may have different VOC profiles. With the cancer progression, the more aggressive tumour tissue may start to penetrate deep into the muscle layer which can cause non-invasive cancer to become invasive. In this case, the VOC's profile could become similar to the former non-invasive but high-grade tumour, or, it could also adapt to the new microenvironment and develop new metabolism alternations and generate different VOC profiles. The limited evidence of our VOC's staging model supports the former hypothesis; the classification of different grading is better than staging, which means that the cancer cells metabolism phenotype, determined according to their aggressiveness, have a larger impact on VOC profiles and distinction than their microenvironment (decided by tumour tissue progression). Again, this part of the work is limited by the small sample size and such hypotheses are based on limited evidence; future larger cohort studies may help to further clarify this problem.

5.3.5. Comparison with other similar studies

Study	Study size *n* (*cancer+control*)	Techniques	Sensitivity	Specificity	Note
[153]	144 (36+108)	Detection dogs	0.41	N/A	
[155]	89 (48+41)	HPLC-MS	1.00	1.00	No CV
[126]	277 (72+205)	GC-MS	0.90	0.88	
[93]	75 (24+51)	GC-ToFMS	1.00	1.00	Non-VOC
[157]	85 (50+35) training 96 (59+37) testing	GC-MS	0.86 0.70	0.86 0.70	CV External validate
[54]	89 (30+59)	GC+e-nose	0.62	0.70	
[124]	98 (24+74)	GC-MOS	0.96	0.95	
[125]	60 (30+30)	CP	0.75	0.86	No CV
Mean	**126.6**		**0.80**	**0.87**	
OURS	**79 (38+41)**	**fluorescence**	**0.84**	**0.88**	

Table 5-11 Comparison of other studies of urine-based diagnosis of bladder cancer. N/A: Not available; CV: cross-validation

Table 5-11 summarized the clinical studies of using urinary metabolome/VOCs biomarkers for diagnosing bladder cancer. Note some of the studies only performed classification without the cross-validation, which may have over-fitting occurs of the model, therefore the classification accuracy may show higher than the real performance. The average study population is 126.6 participants, with mean sensitivity of 0.80, specificity of 0.87. If only compare to the e-nose studies, the average study population is 82.3 participants, with mean sensitivity of 0.78 and specificity of 0.84. No matter compares to overall averages or other e-nose studies our study has better sensitivity and specificity.

5.4. Conclusion

In this chapter, current studies of diagnosing bladder cancer using urinary VOC biomarkers were reviewed. Comparing to the most used detection method GC/LC-MS,

the e-nose is one of the potential techniques in development of point-of-care urinary VOC diagnosis devices. At the point of writing this thesis, only a few studies have had trials using real human urine from bladder cancer patients; the results look promising, but more attention with larger sample sizes and continuous investment is needed.

In this chapter, the system used in Chapter 3 was fully upgraded to adapt the requirement of urine tests. To achieve this, the sample processing module and reaction chamber were redesigned, a small air-bubbler-like device was used for replacing the flowmeter system. Ethic approval for this study was obtained through the East of Scotland Research Ethics Service (REC: 17/ES/0003). In total 42 bladder cancer patients and 40 non-cancer controls were recruited. After confirmation of pathological diagnostic results, the sample population was adjusted to 79, with 38 bladder cancer patients and 41 non-cancer controls.

Using the PLSDA algorithm, a classification model was built. Leave-one-out and 20% random Mote Carlo cross-validation were used for validating the performance of the model. As a result, the PLSDA model successfully identified over 80% of urine samples (86.08% with leave-one-out, 81.76% with Mote Carlo cross-validation) with 77.42%-84.21% sensitivity and 85.82%-87.80% specificity.

A grading model was built based on the PLSDA algorithm and leave-one-out and 20% random Mote Carlo cross-validation were used for validating the performance of the model. As a result, the PLSDA model successfully identified 27 out of 38 urine samples with leave-one-out cross-validation 4,004 out of 8000 times, with 3996 failed group assignment using 20% random Mote Carlo cross-validation. The cross-validation sensitivity and specificity of leave-one-out cross-validation was 66.67% and 75.00%, respectively. However, due to the large numbers of group assignment failures, we could only carefully consider the fluorescence urinary VOCs detection system with the PLSDA model. The fluorescence urinary VOCs detection system has potential in

assisting grading of bladder cancer however more experiments with a larger cohort may help to confirm this.

In conclusion, the fluorescence urinary VOCs detection system successfully classified urine samples of bladder cancer patients from those of healthy controls with good sensitivity and specificity. This system also shows potential in the classification of high-grade and low-grade bladder cancer patients, but further experiments are needed to confirm this. These results indicated that using low-cost fluorescence gas sensor arrays for non-invasive urinary bladder cancer diagnosis is feasible. Future development of novel low-cost portable devices for bladder cancer diagnosis and surveillance could benefit from this technique.

6. Conclusion and Future Works

The aim of this thesis was to explore new techniques for bladder cancer diagnosis and management, and to address the difficulties of existing diagnosis methods. Conventional flexible cystoscopy lacks sensitivity of low-grade cancer and carcinoma in situ, while photodynamic enhanced cystoscopy improved the sensitivity but reduced specificity, especially with inflamed tissue and tissue that treated with chemotherapy and surgery. Conventional non-invasive methods like urinary cytology have very low sensitivity and largely depends on the experience and judgement from pathologists. Novel urinary biomarker tests like NMP22 and UroVysion have fair sensitivity and specificity and significant lower cost comparing to interventional cystoscopy and biopsy, however, those techniques were considered as supplementary rather than primary diagnosis methods for bladder cancer.

One of the reasons limiting the uses of urinary biomarker tests is the ease of use, UroVysion and FISH involves additional process step during conventional cystoscopy and biopsy procedure, while NMP22 or BTA assay does not but they also have lower sensitivities, when consider the benefit of urinary tests for bladder cancer diagnosis, the extra impact and cost on existing processes are not negligible.

In order to address the concerns of diagnostic accuracy, ease of use and cost, I think using volatile organic compounds (VOCs) as diagnostic biomarkers for bladder cancer might be a good idea, biogenic VOCs contain information about nutrients concerning human health, from ancient to modern time people have found thousands of VOCs in human urine and have linked them with various diseases. Specific to bladder cancer, the urinary VOCs are demonstrating different levels or species between cancer cells and healthy cells, it may be induced by various metabolism and nutritional changes during tumour growth such as oxidative stress, cytochrome p450 detoxication,

carbohydrates metabolism (glycolysis/gluconeogenesis pathways), and lipid peroxidation. In chapter 2, I reviewed the latest metabolism pathways and possible explanation of cancer onset and specific VOCs presenting, and how those VOCs be linked with cancer grading and staging, this literature review provided a comprehensive biological and medical background of bladder cancer and urinary VOCs, and gave a theoretical support for application of urinary VOCs in diagnosis and management of bladder cancer.

In chapter 3, I introduced the physical chemistry and optical principles of fluorescence gas sensitive materials. Three major types of fluorescence gas sensitive materials: Lewis acid/base dyes, such as porphyrins and metalloporphyrins; dyes responding to Bronsted acidic or basic pH indicators; dyes with large permanent dipoles, such as zwitterionic solvatochromic dyes, were selected for manufacturing the fluorescence urinary VOCs sensors. A total of 16 sensitive materials were tested using a custom-built detection system. This system was composed of a laser/spectrometer device (LAKK-M); a sample processing module, used for turning the liquid sample into gas phase; a sensory array, made by different manufacturing methods; a reaction chamber, to allow the VOC molecules to interact with the sensor array. The system has several versions to suit different applications and a set of hardware and software parts for control and communications. This system fulfilled the requirement of a proof-of-principle medical device prototype for urinary bladder cancer diagnosis, it can handle different testing samples with corresponding hardware configurations, with selective sensitive materials the cross-responsive sensor array has ability to respond to different analytes.

In chapter 4, I reviewed the basic principle of sensing technology and the common techniques used in artificial olfactory device e-nose. As an imitation of the natural animal olfactory system, the artificial olfactory device or e-nose follow the same

workflow to identify the odours from gas samples. This workflow including olfactory receptors generating biological/electric signals from odorous molecules, transmission of the signals to the olfactory central nervous system or pattern recognition algorithm, and comparing the pattern with odour memory/databases, and finally identifying the odour by recall memories/classification functions. This chapter also reviewed the most commonly used artificial olfactory sensing technologies including electronic (MOS, MOSFET, CP etc.) and non-electronic (QMB, SAW, optical etc.) based sensors. Electronic based sensors have advantages, like low costs and easy integration with IC, but also have disadvantages, like baseline shifting and vulnerability to interferences. Non-electronic based sensors usually have a lower detection limit and better sensitivity, but the relevant detecting systems are more complicated and sometimes more expensive. In this chapter, the fluorescence sensor array and the detection system were tested with 4 literature reviewed urinary VOC biomarkers of bladder cancer: ethylbenzene, hexanal, Lauric aldehyde (dodecanal), and nonanoyl chloride. Using discrimination analysis with algorithms LDA and PLS-DA, the classification models of four urinary VOC biomarkers were built. In verification of such models, leave-one-out cross validation was used. As a result, all four urinary VOC biomarkers were classified correctly, while the leave-one-out cross validation shows that the PLS-DA classification model achieved overall 77.75% sensitivity and 93.25% specificity. In further tests, the limit of detection of the system was found between low-ppmV to medium-ppbV with acceptable discrimination.

In chapter 5, I further reviewed and compared the current studies of diagnosing bladder cancer using urinary VOC biomarkers, from detection dog olfactory tests to the latest inexpensive portable e-nose tests. The most frequently used detection method of urinary VOCs is GC/LC-MS, it can provide qualitative and quantitative information about the compounds profile of human urine; it is powerful and widely used in

academic research in this area. However, due to its costs in both money and labour, an inexpensive and portable detection method for urinary VOCs is needed. E-nose is one of the potential techniques in development of point-of-care urinary VOC diagnosis devices. At the point of writing this thesis, only a few studies have had trials using real human urine from bladder cancer patients; the results look promising, but more attention with larger sample sizes and continuous investment is needed. In this chapter, the detection system previously used was fully upgraded to adapt the requirement of urine tests. The sample processing module and reaction chamber were redesigned, a small air-bubbler-like device was used for replacing the flowmeter system. 38 bladder cancer patients and 41 non-cancer controls were recruited for clinical trials of the fluorescence sensor array and detection system. A classification model for diagnostic purposes was built based on the PLSDA algorithm. Leave-one-out and 20% random Mote Carlo cross-validation were used for validating the performance of the model. As a result, the PLSDA model successfully identified over 80% of urine samples (86.08% with leave-one-out, 81.76% with Mote Carlo cross-validation) with 77.42%-84.21% sensitivity and 85.82%-87.80% specificity. Using the same data, a grading model was built based on the PLSDA algorithm and leave-one-out and 20% random Mote Carlo cross-validation were used for validating the performance of the model. As a result, the PLSDA model successfully identified 27 out of 38 urine samples with leave-one-out cross-validation 4,004 out of 8000 times, with 3996 failed group assignment using 20% random Mote Carlo cross-validation. The cross-validation sensitivity and specificity of leave-one-out cross-validation was 66.67% and 75.00%, respectively. However, due to the large numbers of group assignment failures, we could only carefully consider the fluorescence urinary VOCs detection system with the PLSDA model. The fluorescence urinary VOCs detection system has potential in assisting grading of bladder cancer however more experiments with a larger cohort may help to confirm this.

In conclusion, this study successfully developed a fluorescence cross-response sensor array system for diagnosing bladder cancer by detecting the urinary volatile organic compounds. This system was composed of a laser/spectrometer device (LAKK-M); a sample processing module, used for turning the liquid sample into gas phase; a sensory array, made by different manufacturing methods; a reaction chamber, to allow the VOC molecules to interact with the sensor array. On this system, a distinguishing test of four urinary VOC biomarkers: ethylbenzene, hexanal, lauric aldehyde, and nonanoyl chloride, was undertaken and achieved a sensitivity of 77.75% and a specificity of 93.25%. In a proof-of-principle clinical trial involving 79 participants (38 bladder cancer patients and 41 healthy controls), this system using a PLSDA model successfully identified over 80% of urine samples (86.08% with leave-one-out, 81.76% with Mote Carlo cross-validation) with 77.42%-84.21% sensitivity and 85.82%-87.80% specificity. Further study found this system has potential ability in distinguishing urine samples from high-grade and low-grade bladder cancer patients but further experimental data in support is needed.

The ultimate goal of this project is to develop a point-of-care device for bladder cancer diagnosis and surveillance. However, the patients who participated in this project were all newly-diagnosed bladder cancer patients and did not include anyone with recurring cancer or those under treatment, such patients should be included in future works. For grading purposes, we found that the current sample size was not big enough for building a significant classification model. Future work should use larger sample sizes.

In order to obtain a large sample size, more patient recruitment and larger study cohorts are needed. For possible future multi-centre clinic trails, the current device needs further improvement in order to meet the required standard. For point-of-care devices, besides from the diagnostic accuracy, the convenience of usage is another important factor, including portability, user-friendliness, efficiency and stability of the device.

At the moment the device is still a pre-prototype, which means it works normally and functionally but is still in the laboratory setup phase and has much room for improvement. To reduce the overall size of the device, the urine sample preparation module could use an ultrasonic automizer instead of two bubbler bottles. The fluorescence spectrometer could also be replaced by miniature spectrometers like HAMAMATSU™'s award winning finger-tip sized spectrometer series.

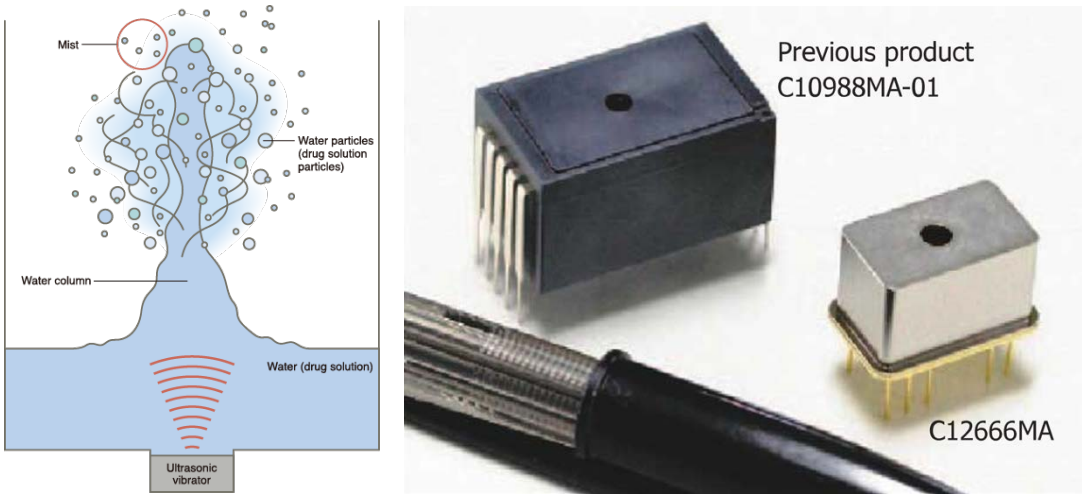


Figure 6-1 Potential techniques to minimize the size of the devices: a. ultrasonic automizer. b. finger-tip sized spectrometers made by HAMAMATSU™

Another possible improvement relates to the operation protocol of the system. At the moment, we need to run three different programs at the same time to control the whole system: one for the spectrometer, one for the flowmeter or sample processing module, and one for the mechanical part to control and monitor the reaction chamber. The operator needs to rapidly switch between each program for the process to proceed and must manually change the spectral filters of the LAKK-M device. This could be further integrated into one central controller which would greatly reduce the workload of the operator and speed up the overall examination process.

Figure 6-2 Example smart phone user interface of future mobile urinary VOCs diagnosis device. Left to right: patient information enrolment page; fluorescence sensor data collection page; discriminant analysis score demonstration page.

This study shows a potential technique for developing a novel low-cost, non-invasive diagnostic device for bladder cancer. However, as reviewed in the Chapters 2 and 3, in addition to bladder or urological cancer, many other cancer types including breast [103-105], thyroid [106], colorectal [107-112], esophagogastric [113], gastric [114, 115], lung [116-121], and pancreatic [122] cancers have been found to have unique biogenic VOC signatures. Such findings have been reported for decades, but no clinical diagnostic tool based on these findings has been put to use. This is strange and there are potential reasons to explain this. Firstly, the study sizes are small due to the limitations of equipment (e.g. not every hospital has analytical devices like GC-MS or HPLC-MS) and there is a lack of multi-centre clinical trials. Secondly, many biogenic VOC pathways are still unclear. When we look back on all the successful cancer biomarkers on the market, they all have clear and specific pathways that correlate to

one or more characteristics of the cancer being targeting. However, biogenic VOC biomarkers sometimes lack such certainty, and this shortcoming remains to be solved by molecular biology and metabolomic studies.

Our study has made a modest contribution to finding a solution to the first difficulties encountered in the clinical application of biogenic VOCs. With a low-cost, non-invasive point-of-care VOC diagnosis device, patient and doctors have access to this novel diagnostic technique for the first time. This could inform larger scale multi-centre trails of biogenic VOCs diagnosis and other possible clinical applications in the future.

7. References

1. Office for National Statistics: **Cancer Statistics Registrations, England (Series MB1), No. 43, 2012**. In: *Series MB1*. Edited by Office for National Statistics; 2014.
2. Scotland ISD: **Cancer Incidence in Scotland (2012)**. In: *Cancer Incidence in Scotland (2012).* Edited by Scotland ISD; 2014.
3. Welsh Cancer Intelligence and Surveillance Unit: **Cancer in Wales**. In: *Cancer in Wales.* Edited by Welsh Cancer Intelligence and Surveillance Unit; 2014.
4. Northern Ireland Cancer Registry: **Incidence & Survival 1993 - 2012**. In *2012 Cancer incidence, survival, mortality and prevalence data.* Edited by Northern Ireland Cancer Registry; 2014.
5. National Institute for Health and Care Excellence: **Bladder cancer: diagnosis and management**. In: *Costing report: Bladder cancer.* vol. NG2: National Institute for Health and Care Excellence; 2015.
6. Drake R, Vogl AW, Mitchell AWM: **Gray's Anatomy for Students**. In., 2 edn. London: Elsevier Health Sciences; 2009.
7. Miyamoto H, Yang Z, Chen Y-T, Ishiguro H, Uemura H, Kubota Y, Nagashima Y, Chang Y-J, Hu Y-C, Tsai M-Y *et al*: **Promotion of Bladder Cancer Development and Progression by Androgen Receptor Signals**. *Journal of the National Cancer Institute* 2007, **99**(7):558-568.
8. Ramirez D, Gupta A, Canter D, Harrow B, Dobbs RW, Kucherov V, Mueller E, Streeper N, Uhlman MA, Svatek RS *et al*: **Microscopic haematuria at time of diagnosis is associated with lower disease stage in patients with newly diagnosed bladder cancer**. *BJU international* 2016, **117**(5):783-786.
9. Mariani AJ, Mariani MC, Macchioni C, Stams UK, Hariharan A, Moriera A: **The significance of adult hematuria: 1,000 hematuria evaluations including a risk-benefit and cost-effectiveness analysis**. *The Journal of urology* 1989, **141**(2):350-355.
10. Grossfeld GD, Litwin MS, Wolf JS, Hricak H, Shuler CL, Agerter DC, Carroll PR: **Evaluation of asymptomatic microscopic hematuria in adults: the American Urological Association best practice policy--part I: definition, detection, prevalence, and etiology**. *Urology* 2001, **57**(4):599-603.
11. Carel RS, Silverberg DS, Kaminsky R, Aviram A: **Routine urinalysis (dipstick) findings in mass screening of healthy adults**. *Clinical chemistry* 1987, **33**(11):2106-2108.
12. Pesch B, Nasterlack M, Eberle F, Bonberg N, Taeger D, Leng G, Feil G, Johnen G, Ickstadt K, Kluckert M *et al*: **The role of haematuria in bladder cancer**

screening among men with former occupational exposure to aromatic amines. *BJU international* 2011, **108**(4):546-552.

13. Willis D, Kamat AM: **Nonurothelial bladder cancer and rare variant histologies**. *Hematology/oncology clinics of North America* 2015, **29**(2):237-252, viii.
14. Humphrey PA, Moch H, Cubilla AL, Ulbright TM, Reuter VE: **The 2016 WHO Classification of Tumours of the Urinary System and Male Genital Organs-Part B: Prostate and Bladder Tumours**. *Eur Urol* 2016, **70**(1):106-119.
15. Kamat AM, Hahn NM, Efstathiou JA, Lerner SP, Malmström P-U, Choi W, Guo CC, Lotan Y, Kassouf W: **Bladder cancer**. *The Lancet* 2016, **388**(10061):2796-2810.
16. Sanli O, Dobruch J, Knowles MA, Burger M, Alemozaffar M, Nielsen ME, Lotan Y: **Bladder cancer**. *Nature Reviews Disease Primers* 2017, **3**:17022.
17. Eble JN, Sauter G, Epstein JI, Sesterhenn IA: **Pathology and genetics of tumours of the urinary system and male genital organs**. *World Health Organization Classi Classification of Tumours, 3rd Edition, Volume 7* 2004.
18. Moch H, Cubilla AL, Humphrey PA, Reuter VE, Ulbright TM: **The 2016 WHO Classification of Tumours of the Urinary System and Male Genital Organs-Part A: Renal, Penile, and Testicular Tumours**. *Eur Urol* 2016, **70**(1):93-105.
19. Gress DM, Edge SB, Greene FL, Washington MK, Asare EA, Brierley JD, Byrd DR, Compton CC, Jessup JM, Winchester DP: **Principles of cancer staging**. *AJCC Cancer Staging Manual 8th Ed* 2016:3-30.
20. Bernard H. Bochner, Donna E. Hansel, Jason A. Efstathiou, Badrinath Konety, Cheryl T. Lee, James M. McKiernan, Elizabeth R. Plimack, Victor E. Reuter, Srikala Sridhar, Raghunandan Vikram *et al*: **Urinary Bladder**. *AJCC Cancer Staging Manual 8th Ed* 2016:757-765.
21. Rochester M, Kelly J: **Bladder Cancer**. In: *Medical Therapy in Urology.* 1st edn. Edited by Shergill IS, Arya M, Grange PR, Mundy AR: Springer London; 2010: 1-16.
22. Yeung C, Dinh T, Lee J: **The health economics of bladder cancer: an updated review of the published literature**. *PharmacoEconomics* 2014, **32**(11):1093-1104.
23. Sangar VK, Ragavan N, Matanhelia SS, Watson MW, Blades RA: **The economic consequences of prostate and bladder cancer in the UK**. *BJU international* 2005, **95**(1):59-63.
24. Palmer S, Sokolovski SG, Rafailov E, Nabi G: **Technologic developments in the field of photonics for the detection of urinary bladder cancer**. *Clin Genitourin Cancer* 2013, **11**(4):390-396.

25. Cross W, Khafagy R: **Urine Cytology**. In: *Imaging and Technology in Urology: Principles and Clinical Applications.* 1st edn. Edited by Payne S, Eardley I, O'Flynn K. London: Springer London; 2012: 171-174.
26. Mowatt G, Zhu S, Kilonzo M, Boachie C, Fraser C, Griffiths TR, N'Dow J, Nabi G, Cook J, Vale L: **Systematic review of the clinical effectiveness and cost-effectiveness of photodynamic diagnosis and urine biomarkers (FISH, ImmunoCyt, NMP22) and cytology for the detection and follow-up of bladder cancer**. *Health technology assessment (Winchester, England)* 2010, **14**(4):1-331, iii-iv.
27. Nabi G, Greene DR, O'Donnell M: **How Important is Urinary Cytology in the Diagnosis of Urological Malignancies?** *European Urology* 2003, **43**(6):632-636.
28. Chan E, Balassanian R, Tabatabai ZL, Lou H, Vohra P: **Improved diagnostic precision of urine cytology by implementation of The Paris System and the use of cell blocks**. *Cancer Cytopathology* 2018, **126**(9):809-816.
29. Meilleroux J, Daniel G, Aziza J, d'Aure DM, Quintyn-Ranty M-L, Basset CML, Evrard SM, Courtade-Saidi MM: **One year of experience using the Paris System for Reporting Urinary Cytology**. *Cancer Cytopathology* 2018, **126**(6):430-436.
30. McIntire PJ, Snow JT, Robinson BD, Rao RA, Goyal A, Heymann JJ, Siddiqui MT: **Improved Correlation of Urinary Cytology Specimens Using The Paris System in Biopsy-Proven Upper Tract Urothelial Carcinomas**. *Cancer Cytopathology* 2018, **126**(7):498-504.
31. Gupta M, VandenBussche C, Bivalacqua T: **Urinary cytology and the Paris system for reporting urinary cytology: Implications for urological management**. *Cytopathology* 2018, **29**(4):368-370.
32. Landman J, Chang Y, Kavaler E, Droller MJ, Liu BCS: **Sensitivity and specificity of NMP-22, telomerase, and BTA in the detection of human bladder cancer**. *Urology* 1998, **52**(3):398-402.
33. Chou R, Gore JL, Buckley D, Fu R, Gustafson K, Griffin JC, Grusing S, Selph S: **Urinary Biomarkers for Diagnosis of Bladder Cancer: A Systematic Review and Meta-analysis**. *Annals of internal medicine* 2015, **163**(12):922-931.
34. Miyake M, Goodison S, Giacoia EG, Rizwani W, Ross S, Rosser CJ: **Influencing factors on the NMP-22 urine assay: an experimental model**. *BMC Urology* 2012, **12**(1):23.
35. Kinders R, Jones T, Root R, Bruce C, Murchison H, Corey M, Williams L, Enfield D, Hass GM: **Complement factor H or a related protein is a marker for transitional cell cancer of the bladder**. *Clinical Cancer Research* 1998, **4**(10):2511-2520.

36. Goodison S, Rosser CJ, Urquidi V: **Bladder cancer detection and monitoring: assessment of urine-and blood-based marker tests**. *Molecular diagnosis & therapy* 2013, **17**(2):71-84.
37. Miyake M, Goodison S, Rizwani W, Ross S, Grossman HB, Rosser CJ: **Urinary BTA: indicator of bladder cancer or of hematuria**. *World journal of urology* 2012, **30**(6):869-873.
38. Lavery HJ, Zaharieva B, McFaddin A, Heerema N, Pohar KS: **A prospective comparison of UroVysion FISH and urine cytology in bladder cancer detection**. *BMC Cancer* 2017, **17**(1):247.
39. Chang SS, Bochner BH, Chou R, Dreicer R, Kamat AM, Lerner SP, Lotan Y, Meeks JJ, Michalski JM, Morgan TM: **Treatment of non-metastatic muscle-invasive bladder cancer: AUA/ASCO/ASTRO/SUO guideline**. *The Journal of urology* 2017, **198**(3):552-559.
40. Fradet Y, Lockhart C: **Performance characteristics of a new monoclonal antibody test for bladder cancer: ImmunoCyt™**. *Can J Urol* 1997, **4**(3):400-405.
41. Comploj E, Mian C, Ambrosini-Spaltro A, Dechet C, Palermo S, Trenti E, Lodde M, Horninger W, Pycha A: **uCyt+/ImmunoCyt and cytology in the detection of urothelial carcinoma**. *Cancer Cytopathology* 2013, **121**(7):392-397.
42. Greene KL, Berry A, Konety BR: **Diagnostic Utility of the ImmunoCyt/uCyt+ Test in Bladder Cancer**. *Reviews in urology* 2006, **8**(4):190-197.
43. Tan WS, Tan WP, Tan MY, Khetrapal P, Dong L, deWinter P, Feber A, Kelly JD: **Novel urinary biomarkers for the detection of bladder cancer: A systematic review**. *Cancer treatment reviews* 2018, **69**:39-52.
44. Williams H, Pembroke A: **Sniffer dogs in the melanoma clinic?** *The Lancet* 1989, **333**(8640):734.
45. **Paddy the dog 'sniffs out' owner's breast cancer** [https://www.bbc.co.uk/news/uk-england-hampshire-34345620]
46. Willis CM, Church SM, Guest CM, Cook WA, McCarthy N, Bransbury AJ, Church MRT, Church JCT: **Olfactory detection of human bladder cancer by dogs: proof of principle study**, vol. 329; 2004.
47. McCulloch M, Jezierski T, Broffman M, Hubbard A, Turner K, Janecki T: **Diagnostic Accuracy of Canine Scent Detection in Early- and Late-Stage Lung and Breast Cancers**. *Integrative Cancer Therapies* 2006, **5**(1):30-39.
48. Amundsen T, Sundstrom S, Buvik T, Gederaas OA, Haaverstad R: **Can dogs smell lung cancer? First study using exhaled breath and urine screening in unselected patients with suspected lung cancer**. *Acta oncologica (Stockholm, Sweden)* 2014, **53**(3):307-315.

49. Gordon RT, Schatz CB, Myers LJ, Kosty M, Gonczy C, Kroener J, Tran M, Kurtzhals P, Heath S, Koziol JA *et al*: **The use of canines in the detection of human cancers**. *Journal of alternative and complementary medicine (New York, NY)* 2008, **14**(1):61-67.
50. Cornu J-N, Cancel-Tassin G, Ondet V, Girardet C, Cussenot O: **Olfactory detection of prostate cancer by dogs sniffing urine: a step forward in early diagnosis**. *European urology* 2011, **59**(2):197-201.
51. Taverna G, Tidu L, Grizzi F, Torri V, Mandressi A, Sardella P, La Torre G, Cocciolone G, Seveso M, Giusti G *et al*: **Olfactory System of Highly Trained Dogs Detects Prostate Cancer in Urine Samples**. *The Journal of urology* 2015, **193**(4):1382-1387.
52. Brooks SW, Moore DR, Marzouk EB, Glenn FR, Hallock RM: **Canine olfaction and electronic nose detection of volatile organic compounds in the detection of cancer: a review**. *Cancer investigation* 2015, **33**(9):411-419.
53. Wagenstaller M, Buettner A: **Characterization of odorants in human urine using a combined chemo-analytical and human-sensory approach: a potential diagnostic strategy**. *Metabolomics* 2013, **9**(1):9-20.
54. Weber CM, Cauchi M, Patel M, Bessant C, Turner C, Britton LE, Willis CM: **Evaluation of a gas sensor array and pattern recognition for the identification of bladder cancer from urine headspace**. *Analyst* 2011, **136**(2):359-364.
55. Lei J-c, Hou C-j, Huo D-q, Luo X-g, Bao M-z, Li X, Yang M, Fa H-b: **A novel device based on a fluorescent cross-responsive sensor array for detecting lung cancer related volatile organic compounds**. *Review of Scientific Instruments* 2015, **86**(2):025106.
56. Zimmermann D, Hartmann M, Moyer MP, Nolte J, Baumbach JI: **Determination of volatile products of human colon cell line metabolism by GC/MS analysis**. *Metabolomics* 2007, **3**(1):13-17.
57. Warburg O: **On the origin of cancer cells**. *Science* 1956, **123**(3191):309-314.
58. Liberti MV, Locasale JW: **The Warburg Effect: How Does it Benefit Cancer Cells?** *Trends in Biochemical Sciences* 2016, **41**(3):211-218.
59. Vander Heiden MG, Cantley LC, Thompson CB: **Understanding the Warburg Effect: The Metabolic Requirements of Cell Proliferation**. *Science* 2009, **324**(5930):1029-1033.
60. Tali F, Jens H, Ingrid K, Guy L, John CC, Jose ST, Maya I, Hossam H, Nir P: **Cancer metabolism: the volatile signature of glycolysis— in vitro model in lung cancer cells**. *Journal of Breath Research* 2017, **11**(1):016008.
61. Yamagishi K, Onuma K, Chiba Y, Yagi S, Aoki S, Sato T, Sugawara Y, Hosoya N, Saeki Y, Takahashi M *et al*: **Generation of gaseous sulfur-containing compounds in tumour tissue and suppression of gas diffusion**

as an antitumour treatment. *Gut* 2012, **61**(4):554-561.

62. Moncada S, Higgs EA, Colombo SL: **Fulfilling the metabolic requirements for cell proliferation**. *The Biochemical journal* 2012, **446**(1):1-7.
63. Hilvo M, Denkert C, Lehtinen L, Muller B, Brockmoller S, Seppanen-Laakso T, Budczies J, Bucher E, Yetukuri L, Castillo S *et al*: **Novel theranostic opportunities offered by characterization of altered membrane lipid metabolism in breast cancer progression**. *Cancer research* 2011, **71**(9):3236-3245.
64. Barr Dana B, Wang Richard Y, Needham Larry L: **Biologic Monitoring of Exposure to Environmental Chemicals throughout the Life Stages: Requirements and Issues for Consideration for the National Children's Study**. *Environmental Health Perspectives* 2005, **113**(8):1083-1091.
65. Apel K, Hirt H: **Reactive oxygen species: metabolism, oxidative stress, and signal transduction**. *Annu Rev Plant Biol* 2004, **55**:373-399.
66. Halliwell B: **Reactive oxygen species in living systems: Source, biochemistry, and role in human disease**. *The American Journal of Medicine* 1991, **91**(3):S14-S22.
67. Lennon SV, Martin SJ, Cotter TG: **Dose-dependent induction of apoptosis in human tumour cell lines by widely diverging stimuli**. *Cell Proliferation* 1991, **24**(2):203-214.
68. Janero DR: **Malondialdehyde and thiobarbituric acid-reactivity as diagnostic indices of lipid peroxidation and peroxidative tissue injury**. *Free Radical Biology and Medicine* 1990, **9**(6):515-540.
69. Shibamoto T: **Analytical methods for trace levels of reactive carbonyl compounds formed in lipid peroxidation systems**. *Journal of Pharmaceutical and Biomedical Analysis* 2006, **41**(1):12-25.
70. Halliwell B, Gutteridge J: **Oxygen toxicity, oxygen radicals, transition metals and disease**. *Biochemical journal* 1984, **219**(1):1.
71. Repetto M, Semprine J, Boveris A: **Lipid peroxidation: chemical mechanism, biological implications and analytical determination**. In: *Lipid peroxidation.* 1st edn.: InTech; 2012.
72. Michiels Y, Puyvelde P, Sels B: **Barriers and Chemistry in a Bottle: Mechanisms in Today's Oxygen Barriers for Tomorrow's Materials**. *Applied Sciences* 2017, **7**(7):665.
73. Frankel EN: **Lipid oxidation**. *Progress in Lipid Research* 1980, **19**(1):1-22.
74. Schaich K: **Lipid oxidation: theoretical aspects**. *Bailey's industrial oil and fat products* 2005.
75. Liou G-Y, Storz P: **Reactive oxygen species in cancer**. *Free Radical Research* 2010, **44**(5):479-496.
76. Kumari S, Badana AK, G MM, G S, Malla R: **Reactive Oxygen Species: A**

Key Constituent in Cancer Survival. *Biomarker insights* 2018, **13**. doi: 10.1177/1177271918755391..

77. Han D, Williams E, Cadenas E: **Mitochondrial respiratory chain-dependent generation of superoxide anion and its release into the intermembrane space**. *The Biochemical journal* 2001, **353**(Pt 2):411-416.
78. Talior I, Yarkoni M, Bashan N, Eldar-Finkelman H: **Increased glucose uptake promotes oxidative stress and PKC-delta activation in adipocytes of obese, insulin-resistant mice**. *American journal of physiology Endocrinology and metabolism* 2003, **285**(2):E295-302.
79. Zhou J, Deo BK, Hosoya K, Terasaki T, Obrosova IG, Brosius FC, 3rd, Kumagai AK: **Increased JNK phosphorylation and oxidative stress in response to increased glucose flux through increased GLUT1 expression in rat retinal endothelial cells**. *Investigative ophthalmology & visual science* 2005, **46**(9):3403-3410.
80. Wallace DC: **A Mitochondrial Paradigm of Metabolic and Degenerative Diseases, Aging, and Cancer: A Dawn for Evolutionary Medicine**. *Annual Review of Genetics* 2005, **39**(1):359-407.
81. Hrycay EG, Bandiera SM: **Chapter Two - Involvement of Cytochrome P450 in Reactive Oxygen Species Formation and Cancer**. In: *Advances in Pharmacology. Volume 74*, 1st edn. Edited by Hardwick JP: Academic Press; 2015: 35-84.
82. Hrycay EG, Bandiera SM: **Monooxygenase, Peroxidase and Peroxygenase Properties and Reaction Mechanisms of Cytochrome P450 Enzymes**. In: *Monooxygenase, Peroxidase and Peroxygenase Properties and Mechanisms of Cytochrome P450*. 1st edn. Edited by Hrycay EG, Bandiera SM. Cham: Springer International Publishing; 2015: 1-61.
83. Androutsopoulos VP, Spyrou I, Ploumidis A, Papalampros AE, Kyriakakis M, Delakas D, Spandidos DA, Tsatsakis AM: **Expression Profile of CYP1A1 and CYP1B1 Enzymes in Colon and Bladder Tumors**. *PLOS ONE* 2013, **8**(12):e82487.
84. Hakim M, Broza YY, Barash O, Peled N, Phillips M, Amann A, Haick H: **Volatile Organic Compounds of Lung Cancer and Possible Biochemical Pathways**. *Chemical Reviews* 2012, **112**(11):5949-5966.
85. Waris G, Ahsan H: **Reactive oxygen species: role in the development of cancer and various chronic conditions**. *Journal of carcinogenesis* 2006, **5**:14-14.
86. Calabrese EJ, Baldwin LA: **Toxicology rethinks its central belief**. *Nature* 2003, **421**(6924):691.
87. Kumar B, Koul S, Khandrika L, Meacham RB, Koul HK: **Oxidative Stress Is Inherent in Prostate Cancer Cells and Is Required for Aggressive**

Phenotype. *Cancer research* 2008, **68**(6):1777.

88. Lewis A, Du J, Liu J, Ritchie JM, Oberley LW, Cullen JJ: **Metastatic Progression of Pancreatic Cancer: Changes in Antioxidant Enzymes and Cell Growth**. *Clinical & Experimental Metastasis* 2005, **22**(7):523-532.
89. Radisky DC, Levy DD, Littlepage LE, Liu H, Nelson CM, Fata JE, Leake D, Godden EL, Albertson DG, Angela Nieto M *et al*: **Rac1b and reactive oxygen species mediate MMP-3-induced EMT and genomic instability**. *Nature* 2005, **436**:123.
90. Hu Y, Rosen DG, Zhou Y, Feng L, Yang G, Liu J, Huang P: **Mitochondrial manganese-superoxide dismutase expression in ovarian cancer: role in cell proliferation and response to oxidative stress**. *The Journal of biological chemistry* 2005, **280**(47):39485-39492.
91. Trachootham D, Zhou Y, Zhang H, Demizu Y, Chen Z, Pelicano H, Chiao PJ, Achanta G, Arlinghaus RB, Liu J *et al*: **Selective killing of oncogenically transformed cells through a ROS-mediated mechanism by β-phenylethyl isothiocyanate**. *Cancer Cell* 2006, **10**(3):241-252.
92. Pasikanti KK, Esuvaranathan K, Ho PC, Mahendran R, Kamaraj R, Wu QH, Chiong E, Chan EC: **Noninvasive urinary metabonomic diagnosis of human bladder cancer**. *J Proteome Res* 2010, **9**(6):2988-2995.
93. Pasikanti KK, Esuvaranathan K, Hong Y, Ho PC, Mahendran R, Raman Nee Mani L, Chiong E, Chan EC: **Urinary metabotyping of bladder cancer using two-dimensional gas chromatography time-of-flight mass spectrometry**. *J Proteome Res* 2013, **12**(9):3865-3873.
94. Pelicano H, Carney D, Huang P: **ROS stress in cancer cells and therapeutic implications**. *Drug resistance updates : reviews and commentaries in antimicrobial and anticancer chemotherapy* 2004, **7**(2):97-110.
95. Alberice JV, Amaral AF, Armitage EG, Lorente JA, Algaba F, Carrilho E, Marquez M, Garcia A, Malats N, Barbas C: **Searching for urine biomarkers of bladder cancer recurrence using a liquid chromatography-mass spectrometry and capillary electrophoresis-mass spectrometry metabolomics approach**. *Journal of chromatography A* 2013, **1318**:163-170.
96. Rodrigues D, Jerónimo C, Henrique R, Belo L, de Lourdes Bastos M, de Pinho PG, Carvalho M: **Biomarkers in bladder cancer: A metabolomic approach using in vitro and ex vivo model systems**. *International Journal of Cancer* 2016, **139**(2):256-268.
97. Lee D-K, Na E, Park S, Park JH, Lim J, Kwon SW: **In Vitro Tracking of Intracellular Metabolism-Derived Cancer Volatiles via Isotope Labeling**. *ACS Central Science* 2018, **4**(8):1037-1044.
98. Hanai Y, Shimono K, Oka H, Baba Y, Yamazaki K, Beauchamp GK: **Analysis of volatile organic compounds released from human lung cancer cells and**

from the urine of tumor-bearing mice. *Cancer cell international* 2012, **12**(1):7-7.

99. Wilson PF, Freeman CG, McEwan MJ, Allardyce RA, Shaw GM: **SIFT-MS Measurement of VOC Distribution Coefficients in Human Blood Constituents and Urine**. *Applied Occupational and Environmental Hygiene* 2003, **18**(10):759-763.
100. Pascual G, Avgustinova A, Mejetta S, Martin M, Castellanos A, Attolini CS, Berenguer A, Prats N, Toll A, Hueto JA *et al*: **Targeting metastasis-initiating cells through the fatty acid receptor CD36**. *Nature* 2017, **541**(7635):41-45.
101. **Directive 2010/75/EU of the European Parliament and of the Council of 24 November 2010 on industrial emissions (integrated pollution prevention and control) Text with EEA relevance**. In. Edited by UNION TEPATCOTE, vol. 2010/75/EU. Official Journal of the European Union; 2010.
102. Mendell MJ: **Indoor residential chemical emissions as risk factors for respiratory and allergic effects in children: a review**. *Indoor Air* 2007, **17**(4):259-277.
103. Barash O, Zhang W, Halpern JM, Hua QL, Pan YY, Kayal H, Khoury K, Liu H, Davies MP, Haick H: **Differentiation between genetic mutations of breast cancer by breath volatolomics**. *Oncotarget* 2015, **6**(42):44864-44876.
104. Li J, Peng Y, Liu Y, Li W, Jin Y, Tang Z, Duan Y: **Investigation of potential breath biomarkers for the early diagnosis of breast cancer using gas chromatography–mass spectrometry**. *Clinica Chimica Acta* 2014, **436**:59-67.
105. Wang C, Sun B, Guo L, Wang X, Ke C, Liu S, Zhao W, Luo S, Guo Z, Zhang Y *et al*: **Volatile Organic Metabolites Identify Patients with Breast Cancer, Cyclomastopathy, and Mammary Gland Fibroma**. *Scientific Reports* 2014, **4**:5383.
106. Guo L, Wang C, Chi C, Wang X, Liu S, Zhao W, Ke C, Xu G, Li E: **Exhaled breath volatile biomarker analysis for thyroid cancer**. *Translational Research* 2015, **166**(2):188-195.
107. Amal H, Leja M, Funka K, Lasina I, Skapars R, Sivins A, Ancans G, Kikuste I, Vanags A, Tolmanis I *et al*: **Breath testing as potential colorectal cancer screening tool**. *International Journal of Cancer* 2015, **138**(1):229-236.
108. Wang C, Ke C, Wang X, Chi C, Guo L, Luo S, Guo Z, Xu G, Zhang F, Li E: **Noninvasive detection of colorectal cancer by analysis of exhaled breath**. *Analytical and bioanalytical chemistry* 2014, **406**(19):4757-4763.
109. de Meij TG, Larbi IB, van der Schee MP, Lentferink YE, Paff T, Terhaar sive Droste JS, Mulder CJ, van Bodegraven AA, de Boer NK: **Electronic nose can discriminate colorectal carcinoma and advanced adenomas by fecal volatile biomarker analysis: proof of principle study**. *International journal*

of cancer 2014, **134**(5):1132-1138.

110. Batty CA, Cauchi M, Lourenço C, Hunter JO, Turner C: **Use of the analysis of the volatile faecal metabolome in screening for colorectal cancer**. *PloS one* 2015, **10**(6):e0130301.
111. Arasaradnam RP, McFarlane MJ, Ryan-Fisher C, Westenbrink E, Hodges P, Thomas MG, Chambers S, O'Connell N, Bailey C, Harmston C: **Detection of colorectal cancer (CRC) by urinary volatile organic compound analysis**. *PLoS One* 2014, **9**(9):e108750.
112. Silva C, Passos M, Camara J: **Investigation of urinary volatile organic metabolites as potential cancer biomarkers by solid-phase microextraction in combination with gas chromatography-mass spectrometry**. *British journal of cancer* 2011, **105**(12):1894.
113. Kumar S, Huang J, Abbassi-Ghadi N, Mackenzie HA, Veselkov KA, Hoare JM, Lovat LB, Spanel P, Smith D, Hanna GB: **Mass Spectrometric Analysis of Exhaled Breath for the Identification of Volatile Organic Compound Biomarkers in Esophageal and Gastric Adenocarcinoma**. *Annals of surgery* 2015, **262**(6):981-990.
114. Amal H, Leja M, Funka K, Skapars R, Sivins A, Ancans G, Liepniece-Karele I, Kikuste I, Lasina I, Haick H: **Detection of precancerous gastric lesions and gastric cancer through exhaled breath**. *Gut* 2016, **65**(3):400-407.
115. Xu Zq, Broza YY, Ionsecu R, Tisch U, Ding L, Liu H, Song Q, Pan Yy, Xiong Fx, Gu Ks *et al*: **A nanomaterial-based breath test for distinguishing gastric cancer from benign gastric conditions**. *British Journal Of Cancer* 2013, **108**:941.
116. Bousamra M, Schumer E, Li M, Knipp RJ, Nantz MH, van Berkel V, Fu X-A: **Quantitative analysis of exhaled carbonyl compounds distinguishes benign from malignant pulmonary disease**. *The Journal of Thoracic and Cardiovascular Surgery* 2014, **148**(3):1074-1081.
117. Corradi M, Poli D, Banda I, Bonini S, Mozzoni P, Pinelli S, Alinovi R, Andreoli R, Ampollini L, Casalini A *et al*: **Exhaled breath analysis in suspected cases of non-small-cell lung cancer: a cross-sectional study**. *Journal of Breath Research* 2015, **9**(2):027101.
118. Fu X-A, Li M, Knipp RJ, Nantz MH, Bousamra M: **Noninvasive detection of lung cancer using exhaled breath**. *Cancer Medicine* 2013, **3**(1):174-181.
119. Tomasz L, Łukasz P, Bogusław B: **Application of an artificial neural network model for selection of potential lung cancer biomarkers**. *Journal of Breath Research* 2015, **9**(2):027106.
120. Sakumura Y, Koyama Y, Tokutake H, Hida T, Sato K, Itoh T, Akamatsu T, Shin W: **Diagnosis by Volatile Organic Compounds in Exhaled Breath from Lung Cancer Patients Using Support Vector Machine Algorithm**.

Sensors 2017, **17**(2).

121. Calderón-Santiago M, Priego-Capote F, Turck N, Robin X, Jurado-Gámez B, Sanchez JC, De Castro MDL: **Human sweat metabolomics for lung cancer screening**. *Analytical and bioanalytical chemistry* 2015, **407**(18):5381-5392.
122. Arasaradnam RP, Wicaksono A, O'Brien H, Kocher HM, Covington JA, Crnogorac-Jurcevic T: **Noninvasive Diagnosis of Pancreatic Cancer Through Detection of Volatile Organic Compounds in Urine**. *Gastroenterology* 2018, **154**(3):485-487.e481.
123. Zhu S, Corsetti S, Wang Q, Li C, Huang Z, Nabi G: **Optical sensory arrays for the detection of urinary bladder cancer related volatile organic compounds (VOCs)**. *Journal of Biophotonics* 2018, Accepted Author Manuscript.(ja):e201800165.
124. Khalid T, White P, De Lacy Costello B, Persad R, Ewen R, Johnson E, Probert CS, Ratcliffe N: **A Pilot Study Combining a GC-Sensor Device with a Statistical Model for the Identification of Bladder Cancer from Urine Headspace**. *PLoS ONE* 2013, **8**(7):e69602.
125. HEERS H, GUT JM, HEGELE A, HOFMANN R, BOESELT T, HATTESOHL A, KOCZULLA AR: **Non-invasive Detection of Bladder Tumors Through Volatile Organic Compounds: A Pilot Study with an Electronic Nose**. *Anticancer Research* 2018, **38**(2):833-837.
126. Cauchi M, Weber C, Bolt B, Spratt P, Bessant C, Turner D, Willis C, Britton L, Turner C, Morgan G: **Evaluation of gas chromatography mass spectrometry and pattern recognition for the identification of bladder cancer from urine headspace**. *Analytical Methods* 2016, **8**(20):4037-4046.
127. Peng G, Hakim M, Broza YY, Billan S, Abdah-Bortnyak R, Kuten A, Tisch U, Haick H: **Detection of lung, breast, colorectal, and prostate cancers from exhaled breath using a single array of nanosensors**. *British Journal Of Cancer* 2010, **103**:542.
128. Khalid T, Aggio R, White P, De Lacy Costello B, Persad R, Al-Kateb H, Jones P, Probert CS, Ratcliffe N: **Urinary Volatile Organic Compounds for the Detection of Prostate Cancer**. *PLOS ONE* 2015, **10**(11):e0143283.
129. Namuangruk S, Sirithip K, Rattanatwan R, Keawin T, Kungwan N, Sudyodsuk T, Promarak V, Surakhot Y, Jungsuttiwong S: **Theoretical investigation of the charge-transfer properties in different meso-linked zinc porphyrins for highly efficient dye-sensitized solar cells**. *Dalton Transactions* 2014, **43**(24):9166-9176.
130. Ma R, Guo P, Cui H, Zhang X, Nazeeruddin MK, Grätzel M: **Substituent effect on the meso-substituted porphyrins: theoretical screening of sensitizer candidates for dye-sensitized solar cells**. *The Journal of Physical Chemistry A* 2009, **113**(37):10119-10124.

131. Uttamlal M, Sheila Holmes-Smith A: **The excitation wavelength dependent fluorescence of porphyrins**. *Chemical Physics Letters* 2008, **454**(4):223-228.
132. Janzen MC, Ponder JB, Bailey DP, Ingison CK, Suslick KS: **Colorimetric Sensor Arrays for Volatile Organic Compounds**. *Analytical Chemistry* 2006, **78**(11):3591-3600.
133. Lipstman S, Muniappan S, Goldberg I: **Interwoven hydrogen-bonded network assembly and supramolecular isomerism of meso-5,10,15,20-tetrakis(4-carboxyphenyl)porphyrin as its dimethylformamide solvate**. *Acta Crystallographica Section C* 2007, **63**(7):o371-o373.
134. Gres SME, Jeffrey GA: **A neutron diffraction refinement of the crystal structure of [beta]-maltose monohydrate**. *Acta Crystallographica Section B* 1977, **33**(8):2490-2495.
135. Suslick KS: **An Optoelectronic Nose:"Seeing" Smells by Means of Colorimetric Sensor Arrays**. *MRS Bulletin* 2004, **29**(10):720-725.
136. Hou C, Dong J, Zhang G, Lei Y, Yang M, Zhang Y, Liu Z, Zhang S, Huo D: **Colorimetric artificial tongue for protein identification**. *Biosensors and Bioelectronics* 2011, **26**(10):3981-3986.
137. Azimi P, Zhao D, Pouzet C, Crain NE, Stephens B: **Emissions of Ultrafine Particles and Volatile Organic Compounds from Commercially Available Desktop Three-Dimensional Printers with Multiple Filaments**. *Environmental Science & Technology* 2016, **50**(3):1260-1268.
138. Kress-Rogers E: **Handbook of biosensors and electronic noses: medicine, food, and the environment**: CRC Press; 1996.
139. Das DK, Goswami P, Barman C, Das B: **Methyl Red: A Fluorescent Sensor for Hg2+ over Na+, K+, Ca2+, Mg2+, Zn2+, and Cd2+**. *Environmental Engineering Research* 2012, **17**(S1):75-78.
140. Persaud K, Dodd G: **Analysis of discrimination mechanisms in the mammalian olfactory system using a model nose**. *Nature* 1982, **299**(5881):352-355.
141. Ghasemi-Varnamkhasti M, Mohtasebi SS, Rodriguez-Mendez ML, Lozano J, Razavi SH, Ahmadi H: **Potential application of electronic nose technology in brewery**. *Trends in Food Science & Technology* 2011, **22**(4):165-174.
142. Dey A: **Semiconductor metal oxide gas sensors: A review**. *Materials Science and Engineering: B* 2018, **229**:206-217.
143. Bicego M, Tessari G, Tecchiolli G, Bettinelli M: **A comparative analysis of basic pattern recognition techniques for the development of small size electronic nose**. *Sensors and Actuators B: Chemical* 2002, **85**(1):137-144.
144. Bishop CM: **Pattern Recognition and Machine Learning (Information Science and Statistics)**: Springer-Verlag; 2006.
145. Jobu K, Sun C, Yoshioka S, Yokota J, Onogawa M, Kawada C, Inoue K, Shuin

T, Sendo T, Miyamura M: **Metabolomics study on the biochemical profiles of odor elements in urine of human with bladder cancer**. *Biol Pharm Bull* 2012, **35**(4):639-642.

146. Rodrigues D, Pinto J, Araújo AM, Monteiro-Reis S, Jerónimo C, Henrique R, de Lourdes Bastos M, de Pinho PG, Carvalho M: **Volatile metabolomic signature of bladder cancer cell lines based on gas chromatography–mass spectrometry**. *Metabolomics* 2018, **14**(5):62.
147. Huang Z, Lin L, Gao Y, Chen Y, Yan X, Xing J, Hang W: **Bladder Cancer Determination Via Two Urinary Metabolites: A Biomarker Pattern Approach**. *Molecular & Cellular Proteomics* 2011, **10**(10):M111.007922.
148. Savitzky A, Golay MJE: **Smoothing and Differentiation of Data by Simplified Least Squares Procedures**. *Analytical Chemistry* 1964, **36**(8):1627-1639.
149. McLachlan GJ: **General Introduction**. In: *Discriminant Analysis and Statistical Pattern Recognition.* 2005 edn.; 2005: 1.
150. MathWorks: **Prediction Using Discriminant Analysis Models**. In: *MATLAB R2018a Documentation.* R2018a edn: MathWorks; 2018.
151. Barker M, Rayens W: **Partial least squares for discrimination**. *Journal of Chemometrics* 2003, **17**(3):166-173.
152. Ballabio D, Consonni V: **Classification tools in chemistry. Part 1: linear models. PLS-DA**. *Analytical Methods* 2013, **5**(16):3790-3798.
153. Willis CM, Church SM, Guest CM, Cook WA, McCarthy N, Bransbury AJ, Church MR, Church JC: **Olfactory detection of human bladder cancer by dogs: proof of principle study**. *BMJ (Clinical research ed)* 2004, **329**(7468):712.
154. Španěl P, Smith D, Holland TA, Singary WA, Elder JB: **Analysis of formaldehyde in the headspace of urine from bladder and prostate cancer patients using selected ion flow tube mass spectrometry**. *Rapid Communications in Mass Spectrometry* 1999, **13**(14):1354-1359.
155. Issaq HJ, Nativ O, Waybright T, Luke B, Veenstra TD, Issaq EJ, Kravstov A, Mullerad M: **Detection of bladder cancer in human urine by metabolomic profiling using high performance liquid chromatography/mass spectrometry**. *The Journal of urology* 2008, **179**(6):2422-2426.
156. Jin X, Yun SJ, Jeong P, Kim IY, Kim W-J, Park S: **Diagnosis of bladder cancer and prediction of survival by urinary metabolomics**. *Oncotarget* 2014, **5**(6):1635.
157. Zhou Y, Song R, Ma C, Zhou L, Liu X, Yin P, Zhang Z, Sun Y, Xu C, Lu X *et al*: **Discovery and validation of potential urinary biomarkers for bladder cancer diagnosis using a pseudotargeted GC-MS metabolomics method**. *Oncotarget* 2017, **8**(13):20719-20728.

158. Horstmann M, Steinbach D, Fischer C, Enkelmann A, Grimm M-O, Voss A: **PD25-03 AN ELECTRONIC NOSE SYSTEM DETECTS BLADDER CANCER IN URINE SPECIMEN: FIRST RESULTS OF A PILOT STUDY**. *The Journal of urology* 2015, **193**(4, Supplement):e560-e561.

159. Zhu S, Nabi G: **238 - Photonic sensor-based detection of urinary volatile organic compounds (VOCs) bladder cancer biomarkers: A prospective study**. *European Urology Supplements* 2019, **18**(1):e315.

160. Aggio RB, Mayor A, Coyle S, Reade S, Khalid T, Ratcliffe NM, Probert CS: **Freeze-drying: an alternative method for the analysis of volatile organic compounds in the headspace of urine samples using solid phase micro-extraction coupled to gas chromatography - mass spectrometry**. *Chemistry Central journal* 2016, **10**:9.

161. Smith S, Burden H, Persad R, Whittington K, de Lacy Costello B, Ratcliffe NM, Probert CS: **A comparative study of the analysis of human urine headspace using gas chromatography-mass spectrometry**. *J Breath Res* 2008, **2**(3):037022.

162. Pelton, A. D., **5 - Thermodynamic Origin of Phase Diagrams**. In *Phase Diagrams and Thermodynamic Modeling of Solutions*, Pelton, A. D., Ed. Elsevier: Amsterdam, 2019; pp 53-84.

Appendix 1 – Publications

Zhu S, Nabi G: **238 - Photonic sensor-based detection of urinary volatile organic compounds (VOCs) bladder cancer biomarkers: A prospective study**. *European Urology Supplements* 2019, **18**(1):e315.

Zhu S, Corsetti S, Wang Q, Li C, Huang Z, Nabi G: **Optical sensory arrays for the detection of urinary bladder cancer related volatile organic compounds (VOCs)**. *Journal of Biophotonics* 2018, **Accepted Author Manuscript.**(ja):e201800165.

Appendix 2 – Patient Recruitment Documents

Study Protocol v. 3.0

Patient Information Sheet (PIS) v.2.0

Informed Consent Form (ICF) v.2.0