Multiple markers for the non-invasive diagnosis and characterisation of prostate cancer

Matthew John Roberts
M.B., B.S. (Qld), B. Sc (Qld)

A thesis submitted for the degree of Doctor of Philosophy at
The University of Queensland in 2017
Faculty of Medicine
UQ Centre for Clinical Research and Centre for Advanced Imaging
Abstract

The early detection of prostate cancer (CaP), the most common internal cancer in men, is limited by an absence of biomarkers that accurately reflect clinically significant disease. Currently, the prostate cancer antigen 3 (PCA3) test in combination with TMPRSS2-ERG fusion gene detection in urine obtained following rectal prostatic massage represent arguably the best standalone biomarkers for CaP but are not widely used in comparison with serum prostate specific antigen (PSA). Limitations in CaP diagnosis and characterisation may be overcome by use of naturally-produced, prostate-specific biofluids, such as ejaculate and post-ejaculate urethral washings (PEUW). Ejaculate is known to contain prostatic cells and relevant molecules in the non-cellular component, while PEUW may contain these as well as biomarkers reflective of systemic changes due to or causing CaP.

The overall aim of this thesis was to use molecular and metabolomic nuclear magnetic resonance (NMR)-based detection methods to study prostatic fluid derived from ejaculate and PEUW to improve CaP diagnosis. Specifically, this thesis sought to optimise ejaculate sample processing for metabolomics studies and evaluate the diagnostic performance of mRNA, microRNA and metabolomic markers in ejaculate to complement serum PSA in detecting clinically significant CaP. Furthermore, this thesis investigated the feasibility and performance of PEUW-based mRNA and metabolomic biomarker performance attributable to the presence of ejaculate in urine to reflect local prostatic and systemic alterations in order to more accurately detect clinically significant CaP. Dependent clinical variables considered were absolute (positive/negative) and clinically significant (present/absent) CaP as well as risk groups (low, intermediate, high) according to the D'Amico criteria.

Using a NMR-based metabolomics enzyme kinetics study design, the addition of tartrate and cooling of ejaculate samples improved the stability of choline and phosphorylcholine concentrations. Sample collection into a sterile urine jar containing 5 mM (on-site) or 10 mM (off-site) tartrate in 20 ml PBS solution cooled to 277 K and cooled during transport until processing would result in at most a 2-3% change in choline and phosphorylcholine to facilitate sample collection off-site without significant effect on choline-based metabolites.

Following prostatic cell RNA isolation, amplification and qPCR for β2-microglobulin (β2M), PSA, PCA3 and Hepsin in ejaculate, adequate RNA for all assays was obtained for 66 patients and determined that a Hepsin:PCA3 ratio in ejaculate together with serum PSA best predicted absolute CaP (AUC= 0.724 vs 0.676) and csCaP (AUC= 0.701 vs 0.680).
Matched mRNA and microRNA expression was possible for a subgroup of patients ($n=20$), with miR-200c ($\text{AUC}=0.788$) and miR-375 ($\text{AUC}=0.758$) performing best on ROC curve analysis. Serum PSA combined with miR-200c and miR-125b improved prediction of absolute CaP ($\text{AUC}=0.869$ vs 0.672; $p<0.05$), improving specificity (67%) at 90% sensitivity compared with PSA alone (11%).

Metabolomic analysis of ejaculate supernatants was complicated by the presence of a variable glucose concentration attributable to the initial buffer used and unbalanced metabolite (choline/phosphocholine) regulation. Following appropriate adjustment, multivariate analysis showed that metabolites best predicted low and intermediate risk CaP with grouping observed between these groups and benign and high risk samples. Lipids/lipoproteins dominated spectra of high grade samples. Overall CaP prediction using metabolites described in previous studies was not validated. However, findings suggest that incorporation of \textit{in vitro} NMR-based metabolomics may translate to \textit{in vivo} metabolite quantification using magnetic resonance spectroscopic imaging (MRSI) to better triage patients for biopsy or monitor active surveillance cohorts.

For 38 PEUW samples, with corresponding ejaculate markers, prostatic cell RNA isolation, amplification and qPCR for $\beta$2M, PSA, PCA3 and Hepsin was possible. PCA3:PSA in ejaculate ($\text{AUC}=0.668$) and PEU ($\text{AUC}=0.569$) were comparable with serum PSA ($\text{AUC}=0.617$) in predicting absolute CaP status. PEUW markers were not predictive of CaP or csCaP, which may be due to predominance of normal prostatic cells over malignant cells and subsequently help identify patients without csCaP.

PEUW NMR-based metabolomics analysis was feasible and demonstrated prostate-specific biomarkers attributable to the presence of ejaculate, both in young volunteers and at-risk patients. As observed for ejaculate, initial metabolomic analysis of PEUW demonstrated best discrimination between low and intermediate risk CaP and benign samples. However, analysis was complicated by the presence of borate-citrate adducts, which has been adjusted in updated sample preparation and data acquisition.

The presented methods may also monitor tumours in active surveillance as well as effects on tumorigenesis by metabolically restricting drugs, such as metformin and atorvastatin, of which this thesis contains a published study protocol for a phase II “window” randomised controlled clinical trial that commenced in March 2015 which has been designed on the findings of this thesis.
This thesis has demonstrated that ejaculate and PEUW are legitimate and valuable sources of CaP-specific biomarkers that require validation in larger cohorts. Prostatic fluid metabolomics serves to improve diagnosis and risk stratification and further investigation is warranted to corroborate these findings in vivo with MRSI in moving towards personalised medicine and improved clinical care for these patients.
Declaration by author

This thesis is composed of my original work, and contains no material previously published or written by another person except where due reference has been made in the text. I have clearly stated the contribution by others to jointly-authored works that I have included in my thesis.

I have clearly stated the contribution of others to my thesis as a whole, including statistical assistance, survey design, data analysis, significant technical procedures, professional editorial advice, and any other original research work used or reported in my thesis. The content of my thesis is the result of work I have undertaken since the commencement of my research higher degree candidature and does not include a substantial part of work that has been submitted to qualify for the award of any other degree or diploma in any university or other tertiary institution. I have clearly stated which parts of my thesis, if any, have been submitted to qualify for another award.

I acknowledge that an electronic copy of my thesis must be lodged with the University Library and, subject to the policy and procedures of The University of Queensland, the thesis be made available for research and study in accordance with the Copyright Act 1968 unless a period of embargo has been approved by the Dean of the Graduate School.

I acknowledge that copyright of all material contained in my thesis resides with the copyright holder(s) of that material. Where appropriate I have obtained copyright permission from the copyright holder to reproduce material in this thesis.
Publications during candidature

Peer-reviewed publications


**Conference abstracts and Presentations**

**Abstracts**


**Invited Oral Presentations (Plenary)**

8. Roberts MJ, Hadway P, Doi SAR, Paterson DL, Gardiner RA. (15th April 2013) Fluoroquinolone resistance in rectal cultures predicts infective complications following prostate biopsy and is not influenced by baseline prevalence – a bias adjusted meta-analysis. Urological Society of Australia and New Zealand (USANZ) 66th Annual Scientific Meeting, Melbourne, Australia (invited presentation)
Oral Presentations

International


Poster presentations


Publications included in this thesis


<table>
<thead>
<tr>
<th>Contributor</th>
<th>Statement of contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matthew J Roberts (Candidate)</td>
<td>Literature search and review (80%)</td>
</tr>
<tr>
<td></td>
<td>Manuscript and figure/table drafting (70%)</td>
</tr>
<tr>
<td>Horst Joachim Schirra</td>
<td>Literature search and review (10%)</td>
</tr>
<tr>
<td></td>
<td>Manuscript and figure/table drafting (15%)</td>
</tr>
<tr>
<td>Martin F Lavin</td>
<td>Manuscript and figure/table drafting including critical revisions (5%)</td>
</tr>
<tr>
<td>Robert “Frank” Gardiner</td>
<td>Literature search and review (10%)</td>
</tr>
<tr>
<td></td>
<td>Manuscript and figure/table drafting (10%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Contributor</th>
<th>Statement of contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matthew J Roberts</td>
<td>Designed experiments (80%)</td>
</tr>
<tr>
<td></td>
<td>Sample collection and performed experiments (60%)</td>
</tr>
<tr>
<td></td>
<td>Data processing (70%)</td>
</tr>
<tr>
<td></td>
<td>Statistical analysis (80%)</td>
</tr>
<tr>
<td></td>
<td>Wrote the paper (75%)</td>
</tr>
<tr>
<td>Jake Hattwell</td>
<td>Sample collection and performed experiments (20%)</td>
</tr>
<tr>
<td></td>
<td>Data processing (10%)</td>
</tr>
<tr>
<td></td>
<td>Statistical analysis (10%)</td>
</tr>
<tr>
<td></td>
<td>Wrote and edited the paper (5%)</td>
</tr>
<tr>
<td>Clement Chow</td>
<td>Designed experiments (5%)</td>
</tr>
<tr>
<td></td>
<td>Sample collection and performed experiments (5%)</td>
</tr>
<tr>
<td></td>
<td>Data processing (5%)</td>
</tr>
<tr>
<td></td>
<td>Critical manuscript revisions (5%)</td>
</tr>
<tr>
<td>Gregory Pierens</td>
<td>Assistance and expert guidance with experimental design and set-up, data acquisition (100%)</td>
</tr>
<tr>
<td></td>
<td>Assistance with interpretation of results (50%)</td>
</tr>
<tr>
<td>Martin Lavin</td>
<td>Assistance with interpretation of results (50%)</td>
</tr>
<tr>
<td></td>
<td>Critical manuscript revisions (10%)</td>
</tr>
<tr>
<td>Robert “Frank” Gardiner</td>
<td>Wrote and edited the paper (5%)</td>
</tr>
<tr>
<td></td>
<td>Critical manuscript revisions (15%)</td>
</tr>
<tr>
<td>Horst Joachim Schirra</td>
<td>Designed experiments (15%)</td>
</tr>
<tr>
<td></td>
<td>Sample collection and performed experiments (15%)</td>
</tr>
<tr>
<td></td>
<td>Data processing (5%)</td>
</tr>
<tr>
<td></td>
<td>Statistical analysis (10%)</td>
</tr>
<tr>
<td></td>
<td>Critical manuscript revisions (70%)</td>
</tr>
<tr>
<td></td>
<td>Wrote and edited the paper (15%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Contributor</th>
<th>Statement of contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Matthew J Roberts</strong> (Candidate)</td>
<td>Designed experiments (30%)</td>
</tr>
<tr>
<td></td>
<td>Patient recruitment, clinical application (10%)</td>
</tr>
<tr>
<td></td>
<td>Understanding of experimental protocol, methodological processes and technique strengths/limitations (100%)</td>
</tr>
<tr>
<td></td>
<td>Data processing and Statistical analysis (70%, 70%)</td>
</tr>
<tr>
<td></td>
<td>Wrote the paper (60%)</td>
</tr>
<tr>
<td><strong>Clement Chow</strong></td>
<td>Sample collection and performed experiments (20%)</td>
</tr>
<tr>
<td></td>
<td>Data processing (10%)</td>
</tr>
<tr>
<td><strong>Horst Joachim Schirra</strong></td>
<td>Statistical analysis (5%), Critical manuscript revisions (20%)</td>
</tr>
<tr>
<td><strong>Renee Richards</strong></td>
<td>Sample collection and performed experiments (10%)</td>
</tr>
<tr>
<td></td>
<td>Data processing (20%)</td>
</tr>
<tr>
<td><strong>Marion Buck</strong></td>
<td>Designed experiments (10%)</td>
</tr>
<tr>
<td></td>
<td>Sample collection and performed experiments (70%)</td>
</tr>
<tr>
<td><strong>Luke Selth</strong></td>
<td>Designed and performed experiments (20%)</td>
</tr>
<tr>
<td></td>
<td>Statistical analysis (10%), Critical manuscript revisions (10%)</td>
</tr>
<tr>
<td><strong>Suhail Doi</strong></td>
<td>Statistical analysis (15%), Critical manuscript revisions (10%)</td>
</tr>
<tr>
<td><strong>Hema Samaratunga, Joanna Perry-Keene, Diane Payton</strong></td>
<td>Critical histopathology review (100%)</td>
</tr>
<tr>
<td></td>
<td>Critical manuscript revisions (5%)</td>
</tr>
<tr>
<td><strong>John Yaxley</strong></td>
<td>Patient recruitment, clinical application (20%)</td>
</tr>
<tr>
<td></td>
<td>Critical manuscript revisions (5%)</td>
</tr>
<tr>
<td><strong>Martin Lavin</strong></td>
<td>Designed experiments (20%), Wrote and edited the paper (10%)</td>
</tr>
<tr>
<td></td>
<td>Critical manuscript revisions (10%)</td>
</tr>
<tr>
<td><strong>Robert “Frank” Gardiner</strong></td>
<td>Designed experiments (20%)</td>
</tr>
<tr>
<td></td>
<td>Patient recruitment, clinical application (70%)</td>
</tr>
<tr>
<td></td>
<td>Wrote and edited the paper (30%)</td>
</tr>
<tr>
<td></td>
<td>Critical manuscript revisions (50%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Contributor</th>
<th>Statement of contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matthew J Roberts</td>
<td>Designed experiments (30%)</td>
</tr>
<tr>
<td>(Candidate)</td>
<td>Patient recruitment, clinical application (10%)</td>
</tr>
<tr>
<td></td>
<td>Sample collection and performed experiments (20%)</td>
</tr>
<tr>
<td></td>
<td>Data processing (80%)</td>
</tr>
<tr>
<td></td>
<td>Statistical analysis (80%)</td>
</tr>
<tr>
<td></td>
<td>Wrote the paper (80%)</td>
</tr>
<tr>
<td>Renee Richards</td>
<td>Sample collection and performed experiments (10%)</td>
</tr>
<tr>
<td></td>
<td>Data processing (5%)</td>
</tr>
<tr>
<td>Clement Chow</td>
<td>Sample collection and performed experiments (15%)</td>
</tr>
<tr>
<td></td>
<td>Data processing (5%)</td>
</tr>
<tr>
<td>Marion Buck</td>
<td>Designed experiments (10%)</td>
</tr>
<tr>
<td></td>
<td>Sample collection and performed experiments (35%)</td>
</tr>
<tr>
<td>John Yaxley</td>
<td>Patient recruitment, clinical application (20%)</td>
</tr>
<tr>
<td></td>
<td>Critical manuscript revisions</td>
</tr>
<tr>
<td>Martin Lavin</td>
<td>Designed experiments (20%)</td>
</tr>
<tr>
<td></td>
<td>Critical manuscript revisions</td>
</tr>
<tr>
<td>Horst Joachim Schirra</td>
<td>Designed experiments (20%)</td>
</tr>
<tr>
<td></td>
<td>Sample collection and performed experiments (20%)</td>
</tr>
<tr>
<td></td>
<td>Data processing (10%), Statistical analysis (15%)</td>
</tr>
<tr>
<td></td>
<td>Research supervision (50%)</td>
</tr>
<tr>
<td></td>
<td>Wrote and edited the paper (10%)</td>
</tr>
<tr>
<td></td>
<td>Critical manuscript revisions</td>
</tr>
<tr>
<td>Robert &quot;Frank&quot; Gardiner</td>
<td>Designed experiments (20%)</td>
</tr>
<tr>
<td></td>
<td>Patient recruitment, clinical application (70%)</td>
</tr>
<tr>
<td></td>
<td>Wrote and edited the paper (10%)</td>
</tr>
<tr>
<td></td>
<td>Research supervision (50%)</td>
</tr>
<tr>
<td></td>
<td>Critical manuscript revisions</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Contributor</th>
<th>Statement of contribution</th>
</tr>
</thead>
</table>
| Matthew J Roberts (Candidate) | Designed experiments (40%)  
Patient recruitment, clinical application (10%)  
Understanding of experimental protocol, methodological processes and technique strengths/limitations  
Data processing (90%), Statistical analysis (80%)  
Wrote the paper (80%) |
| Renee Richards | Sample collection and performed experiments (20%)  
Data processing (5%) |
| Clement Chow | Sample collection and performed experiments (10%)  
Data processing (5%) |
| Suhail Doi | Statistical analysis (15%), Critical manuscript revisions (20%) |
| Horst Joachim Schirra | Designed experiments (5%), Statistical analysis (5%)  
Research supervision (50%), Critical manuscript revisions (15%) |
| Marion Buck | Designed experiments (5%)  
Sample collection and performed experiments (70%) |
| Hema Samaratunga, Joanna Perry-Keene, Diane Payton | Critical histopathology review (100%) |
| John Yaxley | Patient recruitment, clinical application (20%)  
Critical manuscript revisions (5%) |
| Martin Lavin | Designed experiments (30%), Manuscript revisions (5%) |
| Robert “Frank” Gardiner | Designed experiments (20%)  
Patient recruitment, clinical application (70%)  
Wrote and edited the paper (20%), Research supervision (50%), Critical manuscript revisions (55%) |

<table>
<thead>
<tr>
<th>Contributor</th>
<th>Statement of contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matthew J Roberts (Candidate)</td>
<td>Drafting study protocol (40%)</td>
</tr>
<tr>
<td></td>
<td>Manuscript design and initial drafting (80%)</td>
</tr>
<tr>
<td></td>
<td>Trial co-ordination (5%)</td>
</tr>
<tr>
<td></td>
<td>Trial recruitment (10%)</td>
</tr>
<tr>
<td></td>
<td>Laboratory experimental design (40%)</td>
</tr>
<tr>
<td></td>
<td>Statistical considerations (70%)</td>
</tr>
<tr>
<td></td>
<td>Wrote the paper (80%)</td>
</tr>
<tr>
<td>John Yaxley</td>
<td>Trial recruitment (10%)</td>
</tr>
<tr>
<td>Geoff Coughlin</td>
<td>Trial recruitment (10%)</td>
</tr>
<tr>
<td>Troy Gianduzzo</td>
<td>Trial recruitment (10%)</td>
</tr>
<tr>
<td></td>
<td>Critical manuscript revisions (10%)</td>
</tr>
<tr>
<td>Rachel Esler</td>
<td>Trial recruitment (10%)</td>
</tr>
<tr>
<td>Nigel Dunglison</td>
<td>Trial recruitment (10%)</td>
</tr>
<tr>
<td>Suzanne Chambers</td>
<td>Psychosocial advice and oversight (90%)</td>
</tr>
<tr>
<td>Robyn Medcraft</td>
<td>Drafting study protocol (10%)</td>
</tr>
<tr>
<td></td>
<td>Trial co-ordination (80%)</td>
</tr>
<tr>
<td>Clement Chow</td>
<td>Laboratory experimental design (5%)</td>
</tr>
<tr>
<td>Horst Joachim Schirra</td>
<td>Laboratory experimental design (20%)</td>
</tr>
<tr>
<td></td>
<td>Statistical considerations (5%)</td>
</tr>
<tr>
<td></td>
<td>Wrote and edited the paper (5%)</td>
</tr>
<tr>
<td></td>
<td>Critical manuscript revisions (10%)</td>
</tr>
<tr>
<td>Renee Richards</td>
<td>Laboratory experimental design (20%)</td>
</tr>
<tr>
<td></td>
<td>Critical manuscript revisions (10%)</td>
</tr>
<tr>
<td>Nicholas Kienzle</td>
<td>Imaging co-ordination and interpretation (40%)</td>
</tr>
<tr>
<td>Macy Lu</td>
<td>Imaging co-ordination and interpretation (40%)</td>
</tr>
<tr>
<td>Ian Brereton</td>
<td>Imaging co-ordination and interpretation (20%)</td>
</tr>
<tr>
<td>Hema Samaratunga, Joanna Perry-Keene, Diane Payton</td>
<td>Critical histopathology review (100%)</td>
</tr>
<tr>
<td>Name</td>
<td>Contributions</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>---------------</td>
</tr>
<tr>
<td>Chikara Oyama</td>
<td>Laboratory experimental design (5%)</td>
</tr>
<tr>
<td></td>
<td>Statistical considerations (20%)</td>
</tr>
<tr>
<td></td>
<td>Wrote and edited the paper (5%)</td>
</tr>
<tr>
<td></td>
<td>Critical manuscript revisions (20%)</td>
</tr>
<tr>
<td>Suhail Doi</td>
<td>Assistance with trial and experimental design (100%)</td>
</tr>
<tr>
<td></td>
<td>Critical manuscript revisions (5%)</td>
</tr>
<tr>
<td>Martin Lavin</td>
<td>Drafting study protocol (50%)</td>
</tr>
<tr>
<td></td>
<td>Manuscript design and initial drafting (20%)</td>
</tr>
<tr>
<td></td>
<td>Trial co-ordination (15%)</td>
</tr>
<tr>
<td></td>
<td>Trial recruitment (40%)</td>
</tr>
<tr>
<td></td>
<td>Psychosocial advice and oversight (10%)</td>
</tr>
<tr>
<td></td>
<td>Laboratory experimental design (10%)</td>
</tr>
<tr>
<td></td>
<td>Statistical considerations (5%)</td>
</tr>
<tr>
<td></td>
<td>Critical manuscript revisions (45%)</td>
</tr>
<tr>
<td></td>
<td>Wrote and edited the paper (10%)</td>
</tr>
<tr>
<td>Robert “Frank” Gardiner</td>
<td>Drafting study protocol (50%)</td>
</tr>
<tr>
<td></td>
<td>Manuscript design and initial drafting (20%)</td>
</tr>
<tr>
<td></td>
<td>Trial co-ordination (15%)</td>
</tr>
<tr>
<td></td>
<td>Trial recruitment (40%)</td>
</tr>
<tr>
<td></td>
<td>Psychosocial advice and oversight (10%)</td>
</tr>
<tr>
<td></td>
<td>Laboratory experimental design (10%)</td>
</tr>
<tr>
<td></td>
<td>Statistical considerations (5%)</td>
</tr>
<tr>
<td></td>
<td>Critical manuscript revisions (45%)</td>
</tr>
<tr>
<td></td>
<td>Wrote and edited the paper (10%)</td>
</tr>
</tbody>
</table>
Contributions by others to the thesis

The PhD candidate, Matthew J Roberts, undertook data curation, data analysis and interpretation of all chapters constituting this thesis. Drafting of all chapters and publications contributing to this thesis was performed by Matthew J Roberts under the guidance of supervising and contributing authors.

Specific proportions of roles performed by each author/investigator are listed for each manuscript above, while significant and substantial inputs made by others to the research are outlined below.

Part 2 - Methods

This chapter, incorporating the manuscript “Tartrate inhibition of prostatic acid phosphatase improves seminal fluid metabolite stability” (Chapter 3) was designed to determine optimal sample collection and analysis methods. The majority (more than 90%) of experimental design, sample collection, experimental setup, data acquisition and analysis as well as drafting for these chapters were performed by the PhD candidate, Matthew J Roberts. Dr Horst Joachim Schirra provided primary supervision of these experiments and assisted in all components. Prof “Frank” Gardiner provided critical clinical considerations to the experimental design and interpretation of results in addition to critical manuscript revisions. Jake Hattwell assisted with experimental setup, data acquisition for one of the replicates, as well as with manuscript finalisation. Mr Clement Chow assisted with experimental setup and data acquisition of PCR data. Dr Gregory Pierens provided critical technical assisting in NMR data acquisition and interpretation, while Prof Martin Lavin provided input into experimental design, data interpretation and critical manuscript revisions.

Part 3 - Results

Chapter 4 reported on the analysis of ejaculate and includes the manuscripts “Diagnostic performance of expression of PCA3, Hepsin and miR biomarkers in ejaculate in combination with serum PSA for the detection of prostate cancer” (Chapter 4.2) and “Selection and monitoring of active surveillance candidates using NMR-based metabolomics of ejaculate: a preliminary investigation” (Chapter 4.3).

Chapter 5 reported on the analysis of PEUW and includes the manuscript “Prostate-based biofluids for the detection of prostate cancer: a comparative study of the diagnostic performance of cell-sourced RNA biomarkers” (Chapter 5.2), as well as a preliminary
analysis of “NMR-based metabolomics of post-ejaculate urethral washings to predict and characterise prostate cancer using local and systemic metabolite changes.”

Chapter 6 describes how the presented research can be applied clinically using a clinical trial protocol entitled “Can atorvastatin with metformin change the natural history of prostate cancer as characterized by molecular, metabolomic, imaging and pathological variables? A randomized controlled trial protocol” (Chapter 4.4).

The majority (more than 80%) of NMR-based experimental design and setup, data acquisition, as well as overall statistical analysis, interpretation of findings and drafting for these chapters were performed by the PhD candidate, Matthew J Roberts.

Other significant contributions include routine technical work such as sample collection and processing as well as biobank curation by Clement Chow, Renee Richards and Marion Buck. The original conception and design of the overall project was performed by Profs Gardiner and Lavin, with specific application to NMR metabolomics and experimental design in conjunction with Dr Schirra. Dr Selth provided expert advice regarding miRNAs. Patients were recruited by Prof Gardiner and Dr Yaxley, which will be assisted by Dr Coughlin, Dr Gianduzzo, Dr Esler and Dr Dunglison. Pathologists Hema Samaratunga, Joanna Perry-Keene and Diane Payton provided critical histopathological analysis and guidance critical to correct disease classification. Profs Gardiner and Lavin provided appropriate clinical and laboratory guidance with interpretation of data, respectively, while Dr Schirra provided all supervision and assistance in NMR experiments, specifically relating to sample preparation, experimental setup and execution, data processing and statistical analysis. Prof Doi assisted with clinical biostatistics and appropriate analysis for relevant clinical translation. Prof Gardiner and Dr Schirra provided the majority of supervision in experimental and data interpretation as well as critical revision of the manuscript drafts.

**Statement of parts of the thesis submitted to qualify for the award of another degree**

None
Acknowledgements

I am tremendously grateful to my financial supporters of my research candidature, specifically Cancer Council Queensland, Avant Mutual Group Ltd as well as the School of Medicine and Graduate School, The University of Queensland. I would also like to acknowledge financial supporters of my research group such as the Royal Brisbane and Women’s Hospital Foundation, the Prostate Cancer Foundation of Australia (PCFA) and Cancer Council Queensland.

Firstly, I am forever grateful to Professor RA “Frank” Gardiner for his encouragement, mentorship and support throughout my candidature and early career in urology. Prof Gardiner provided me with a unique opportunity and has always looked out for my best interests with support of my enthusiasm for academic urology as well as being a voice of reason during the tough times. A pioneer in academic urology, he is a true surgeon scientist rightfully held in high esteem by urological and non-urological colleagues.

I am thankful for the supervision provided by Dr Horst Schirra throughout my candidature, who taught me to think critically as a scientist with attention to detail in experimental design, statistical analysis and manuscript drafting. His enthusiasm for scientific rigor and perfection has allowed me to become a more complete researcher and are traits that underpin my ever-evolving research philosophy.

I have been very fortunate to closely interact with Professor Martin Lavin, a scientific statesman who, while co-supervising the laboratory works at UQCCR, has been a constant source of wisdom and guidance throughout my candidature. Furthermore, the voluntary mentorship of Associate Professor Suhail Doi has been invaluable, who taught me clinically applicable biostatistics and research integrity.

I would like to thank Dr Renee Richards for her friendship, encouragement and support during my candidature. Furthermore, Mr Clement Chow has been invaluable in providing technical assistance and education.

My clinical mentors Drs John Yaxley, Geoff Coughlin, Rachel Esler, Troy Gianduzzo and Nigel Dunglison have inspired me towards excellence in clinical urology and supported my research endeavours to improve patient outcomes.

My family, in particular my parents Carmel and Michael as well as brothers Jason and Darren, have been a constant source of motivation and inspiration to reach greater heights. I am thankful for the support and understanding constantly provided by my family,
including sister Natasha and her family Kiernan, Max, Grace and Will, brother Jason and
his wife Alison with Evie, Lucy and Isabelle and brother Gareth and his wife Inge with
Sienna and Alina. I extend my sincerest appreciation to Lenore, Bill and Jessica Purvey for
their patience, understanding and assistance in allowing me to concentrate in my
candidature and career in urology.

Most importantly, this journey of my candidature and professional endeavours would not
be possible without the support of my ever-understanding wife Megan who has been a
pillar of strength, motivation and support.
Keywords
prostate cancer, biomarker, ejaculate, seminal fluid, mRNA, microRNA, prostate specific antigen (PSA), prostate cancer antigen 3 (PCA3), nuclear magnetic resonance (NMR) spectroscopy, metabolomics

Australian and New Zealand Standard Research Classifications (ANZSRC) 2008
ANZSRC code: 110199, Medical Biochemistry and Metabolomics not elsewhere classified, 40%

ANZSRC code: 110105, Medical Biochemistry: Nucleic Acids, 30%

ANZSRC code: 111202, Cancer Diagnosis, 30%

Fields of Research (FoR) Classification
FoR code: 1101, Medical Biochemistry and Metabolomics, 60%

FoR code: 1112, Oncology and Carcinogenesis, 30%

FoR code: 1103, Clinical Sciences, 10%
Table of Contents

Publications during candidature ........................................................................... vi
Conference abstracts and Presentations ................................................................. viii
Publications included in this thesis .......................................................................... xi
Contributions by others to the thesis ........................................................................ xviii
Acknowledgements ................................................................................................... xx

Keywords .................................................................................................................. xxii

Australian and New Zealand Standard Research Classifications (ANZSRC) 2008 .. xxii

Table of Contents ..................................................................................................... xxiii

List of Figures .......................................................................................................... xxvi

List of Tables ............................................................................................................ xxx

List of Abbreviations used in the thesis ..................................................................... xxxii

Part 1 - Introduction .................................................................................................. 1

Chapter 1 - Introduction ........................................................................................... 2

  Overview.................................................................................................................. 2

  Aims......................................................................................................................... 5

Chapter 2 – Molecular and metabolomic biomarkers for non-invasive prostate cancer diagnosis and monitoring ................................................................. 6

  2.1 Synopsis.............................................................................................................. 6
  2.2 “NMR-based Metabolomics: Global Analysis of Metabolites to Address Problems in Prostate Cancer” (published manuscript) .................................................. 7
  2.2.1 Introduction.................................................................................................. 9
  2.2.2 Metabolomics: History and Methods .......................................................... 10
  2.2.3 Current Evidence: Metabolomics in Prostate Cancer ............................... 30
  2.2.4 Future Directions ....................................................................................... 38
  2.2.5 Conclusions ............................................................................................... 43

  2.3 Recent developments in prostate cancer metabolomics .................................... 44

  2.4 Conclusions on molecular and metabolomic biomarkers for non-invasive prostate cancer diagnosis and monitoring ............................................. 51

Part 2 – Methods ..................................................................................................... 52
5.2.2 Introduction .............................................................................................................. 123
5.2.3 Subjects and Methods .......................................................................................... 124
5.2.4 Results .................................................................................................................... 127
5.2.5 Discussion .............................................................................................................. 131

5.3 NMR-based metabolomics of post-ejaculate urethral washings to predict and characterise prostate cancer using local and systemic metabolite changes .............. 134
5.3.1 Introduction ............................................................................................................. 135
5.3.2 Materials and Methods ......................................................................................... 136
5.3.3 Results .................................................................................................................... 139
5.3.4 Discussion .............................................................................................................. 145
5.3.5 Conclusions ......................................................................................................... 147

Chapter 6 – From bench to bedside: clinical implementation of prostatic fluid biomarkers ................................................................................................................................. 148

6.1 Synopsis .................................................................................................................... 148

6.2 “Can atorvastatin with metformin change the natural history of prostate cancer as characterized by molecular, metabolomic, imaging and pathological variables? A randomized controlled trial protocol.” (published manuscript) .......................... 149
6.2.1 Abstract ................................................................................................................ 151
6.2.2 Introduction .......................................................................................................... 152
6.2.3 Materials and Methods ......................................................................................... 153
6.2.4 Discussion ............................................................................................................ 160

Part 4 – Conclusions ........................................................................................................ 161

Chapter 7 – Summary and discussion ........................................................................ 162

7.1 Summary .................................................................................................................. 162
7.2 Future directions ...................................................................................................... 164

Bibliography .................................................................................................................... 165

Appendices ..................................................................................................................... 200

Appendix 1 – Chapter 3 ................................................................................................. 200
Appendix 2 – Chapter 4 ................................................................................................. 209
Appendix 3 – Chapter 5 ................................................................................................. 225
Appendix 4 – Publication links ...................................................................................... 230
## List of Figures

### Chapter 2

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Major targets for exploratory analysis in systems biology.</td>
<td>14</td>
</tr>
<tr>
<td>2.2</td>
<td>Sequential steps of processing NMR spectra.</td>
<td>18</td>
</tr>
<tr>
<td>2.3</td>
<td>Illustration of data reduction by “bucketing”.</td>
<td>20</td>
</tr>
<tr>
<td>2.4</td>
<td>Schematic depiction of spectra containing four NMR signals A-D, following two different normalization methods.</td>
<td>21</td>
</tr>
<tr>
<td>2.5</td>
<td>Visualization of a typical multivariate analysis (e.g. PCA, PLS etc.).</td>
<td>26</td>
</tr>
<tr>
<td>2.6</td>
<td>Data included in a PLS analysis.</td>
<td>28</td>
</tr>
<tr>
<td>2.7</td>
<td>Validation by permutation analysis.</td>
<td>29</td>
</tr>
<tr>
<td>2.8</td>
<td>Pertinent physiology of the healthy PZ epithelium.</td>
<td>31</td>
</tr>
<tr>
<td>2.9</td>
<td>Pathophysiology of the PZ epithelium after malignant transformation.</td>
<td>32</td>
</tr>
<tr>
<td>2.10</td>
<td>Graphic representation of stepwise data integration of multiple –omics datasets with O2PLS.</td>
<td>41</td>
</tr>
</tbody>
</table>

### Chapter 3

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Experimental workflow of sample preparation and spectral acquisition using two-factor design.</td>
<td>59</td>
</tr>
<tr>
<td>3.2</td>
<td>Kinetic profiles of phosphorylcholine hydrolysis to choline mediated by prostatic acid phosphatase.</td>
<td>67</td>
</tr>
<tr>
<td>3.3</td>
<td>Kinetic profiles of phosphorylcholine hydrolysis to choline mediated by prostatic acid phosphatase under varying tartrate concentration.</td>
<td>68</td>
</tr>
<tr>
<td>3.4</td>
<td>Apparent first-order rate constants (kapp) obtained from exponential fits of phosphorylcholine hydrolysis and choline production from the three subjects for reactions at 298 K and 279 K.</td>
<td>69</td>
</tr>
<tr>
<td>3.5</td>
<td>Principal Components Analysis of human ejaculate NMR spectra acquired at 0 h and 24 h following donation from three subjects, measured at different temperatures, and with or without the addition of 10 mM tartrate.</td>
<td>71</td>
</tr>
</tbody>
</table>
Figure 3.6: Principal Components Analysis of ejaculate NMR spectra after exclusion of the tartrate signal from the dataset.

Chapter 4

Figure 4.1: Receiver Operating Characteristic curve analysis of serum PSA and Composite1 and Composite2 in predicting prostate cancer status.

Figure 4.2: Sections of 1D NOESY spectra from ejaculate measured at 900 MHz.

Figure 4.3: Principal Components Analysis of ejaculate NMR spectra from men being investigated for prostate cancer (n = 151), prepared with different buffer solutions.

Figure 4.4: Principal Components Analysis after exclusion of choline containing metabolites.

Figure 4.5: Supervised, partial least squares analysis of ejaculate NMR spectra in predicting csCaP (according to D’Amico criteria) following Add-to-Subtract.

Figure 4.6: Supervised, partial least squares analysis of ejaculate NMR spectra in predicting CaP risk (according to D’Amico criteria) following Add-to-Subtract.

Figure 4.7: Attempted adaptations to Add-to-Subtract to improve ejaculate metabolomics analyses.

Figure 4.8: Feature-mswsd box plots demonstrating data spread with variable feature use.

Chapter 5

Figure 5.1: 1D NOESY NMR Spectrum of PEUW, demonstrating the presence of metabolites known to be present in ejaculate, urine and in both.

Figure 5.2: 1D NOESY segment (3.15 – 3.23 ppm) spectral overlay of a urine-based sample set from the Young cohort.

Figure 5.3: Pseudo-2D spectral overlay of serial measurements of PEUW at 298K over time in the region of choline/phosphocholine with hydrolysis reaction curves.

Figure 5.4: Preliminary analysis of 106 PEUW spectra from patients being investigated for CaP.
Figure 5.5: Preliminary sub-group analysis of PEUW spectra classified as negative, low risk (green) compared with intermediate risk (gold) and benign (blue).

Figure 5.6: 1D NOESY PEUW spectral overlay showing sarcosine and citrate in the region 2.74 – 2.5 ppm.

Chapter 6
Figure 6.1: Flow chart of study design, incorporating timing of specimen collection, imaging and capsule distribution.

Appendix 1
Figure A1.1: Representative 1D NOESY spectra of human ejaculate (Subject 3).

Figure A1.2: Time-dependent phosphorylcholine (PCho) hydrolysis to choline (Cho) in ejaculate at 277 K with 1 mM tartrate characterised by NMR spectroscopy.

Figure A1.3: Panel a - Overlay of initial 1D NOESY NMR spectra acquired from ejaculate samples at 298 K with varying tartrate concentrations over the region 2.4 – 4.4 ppm.

Figure A1.4: Relative expression of RNA biomarkers (PCA3, PSA and Hepsin normalized to β2M) in LNCaP cell culture and phosphate buffered saline (PBS - black).

Figure A1.5: Apparent activation energies calculated from linear fits of apparent rate constants of phosphorylcholine hydrolysis for each subject in an Arrhenius plot.

Appendix 2
Figure A2.1 – Comparison plot of CtAv vs. CtSD plots for qPCR of PSA, Hepsin and PCA3.

Figure A2.2 – Flowchart demonstrating loss of samples for consideration in final analysis from initial cohort.

Figure A2.3 – Predictive probability plots illustrating the relationship between the composite scores and the prediction of the binary variable based on the multivariate logistic regression.

Figure A2.4 – Supervised partial least squares multivariate statistical analysis of ejaculate samples with different predictive classifiers against benign samples.
Figure A2.5 – Supervised partial least squares multivariate statistical analysis of ejaculate samples containing only PBS buffer, with different predictive sample classifiers compared with benign samples.

Figure A2.6 – Unsupervised multivariate statistical analysis (principal components analysis) of valid supervised models to test validity of the supervised models.

Figure A2.7 – Box plots comparing lactate bucket intensity against risk categories for all samples and samples collected in PBS only.

Figure A2.8 – Supervised multivariate statistical analysis (partial least squares) of samples according to presence or absence of the TMPRSS2:ERG fusion gene.

Figure A2.9 – Principal components multivariate statistical analysis of samples following sum of choline- and citrate-containing buckets, coloured according to CaP presence.

Appendix 3

Figure A3.1 – Comparison plot of average Ct (CtAv) vs. Ct standard deviation (CtSD) plots for qPCR of PSA, PCA3 and β2M in PEUW.

Figure A3.2 – Flowchart demonstrating sample loss during workflow towards final analysis from initial cohort.

Figure A3.3 – ROC comparison figures for each analysis, comprising PEUW (paired cohort) and Ejaculate (paired cohort) for absolute CaP.

Figure A3.4 – ROC comparison figures for each analysis, comprising PEUW (paired cohort) and Ejaculate (paired cohort) for clinically significant CaP.
List of Tables

Chapter 2

Table 2.1: Comparison of the advantages and disadvantages of NMR, GC-MS and LC-MS. 17
Table 2.2: Summary of multivariate statistical analysis methods used in metabolomics for information recovery. 24
Table 2.3: Examples of software packages used to perform multivariate statistical analysis in metabolomics. 25
Table 2.4: Summary of metabolite changes in prostate cancer. 36
Table 2.5: Localised and systemic metabolite changes in prostate cancer based on metabolomics studies. 50

Chapter 4

Table 4.1: Patient classification schemes, including both CaP status (positive/negative) and clinical risk (as determined by D’Amico or PRIAS criteria). 86
Table 4.2: Demographic information for patients undergoing prostate biopsy (Bx) based on different classification methods used (D’Amico, PRIAS). 90
Table 4.3: ROC analysis summary for available biomarkers for CaP status based on most recent biopsy histology. 92
Table 4.4: ROC analysis summary for available biomarkers using D’Amico-based risk stratification as determined by biopsy and RP histology. 93
Table 4.5: ROC analysis summary for available biomarkers using PRIAS-based risk stratification as determined by biopsy and RP histology. 94
Table 4.6: Demographic information for patients based on biopsy and radical prostatectomy (RP) histology. 107

Chapter 5

Table 5.1: Demographic information for patients based on absolute cancer status (positive vs. negative biopsy; A) and clinically significant CaP classification (B). 130
Table 5.2: Sample collection regime for Young cohort biofluid samples. 137
Table 5.3: Collection totals for mid-stream urine (MSU), post-ejaculate urethral washing (PEUW) and ejaculate (E) samples. 140
Appendix 1

Table A1.1: Calculated apparent first-order kinetics exponential fit parameters for spectral intensities for phosphorylcholyne-choline hydrolysis.

Table A1.2: Descriptive statistics for calculated concentrations of key metabolites identified in the multivariate statistical analysis (Figure 3.6) using tartrate as a concentration standard.

Appendix 2

Table A2.1: Stepwise logistic regression results incorporating serum PSA, Hepsin, PCA3 and Hepsin:PCA3 and additional analysis incorporating miRNAs.

Table A2.2: Figures of Merits of all MVSA models.

Table A2.3: Logistic regression weightings following targeted metabolite quantification using Chenomx.

Table A2.4: Assignments of metabolites identified in 1D NOESY ejaculate spectra.
List of Abbreviations used in the thesis
Abbreviations Used:

1D  one-dimensional
2D  two-dimensional
ADT  androgen deprivation therapy
AMIX  Analysis of Mixtures
ANOVA  analysis of variance
AR  androgen receptor
AUC  area under the curve
β2M  beta-2 microglobulin
bEvo  biochemical evolution
BPH  benign prostatic hyperplasia
CaP  prostate cancer / carcinoma of the prostate
csCaP  clinically significant prostate cancer
Ct / CtAv / CtSD  cycle threshold / average / standard deviation
CT  computerised tomography
D₂O  deuterium oxide
DFTMP  1,1-difluoro-1-trimethylsilanyl methylphosphonic acid
DRE  digital rectal examination
DSS  4,4-dimethyl-4-silapentane-1-sulfonic acid
EPS  expressed prostatic secretion
FID  free induction decay
GC  gas chromatography
GPCho / G-P-Cho  glycerophosphocholine
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>GOI</td>
<td>gene of interest</td>
</tr>
<tr>
<td>GWAS</td>
<td>genome-wide association studies</td>
</tr>
<tr>
<td>HBSS</td>
<td>Hanks Balanced Salt Solution</td>
</tr>
<tr>
<td>HPLC</td>
<td>high performance liquid chromatography</td>
</tr>
<tr>
<td>HR-MAS</td>
<td>high resolution magic angle spinning</td>
</tr>
<tr>
<td>IQR</td>
<td>interquartile range</td>
</tr>
<tr>
<td>ISUP</td>
<td>International Society of Urological Pathology</td>
</tr>
<tr>
<td>$k_{app}$</td>
<td>apparent rate constant</td>
</tr>
<tr>
<td>LC</td>
<td>liquid chromatography</td>
</tr>
<tr>
<td>miRNA</td>
<td>microRNA</td>
</tr>
<tr>
<td>mpMRI</td>
<td>multiparametric magnetic resonance imaging</td>
</tr>
<tr>
<td>MRI</td>
<td>magnetic resonance imaging</td>
</tr>
<tr>
<td>MRS</td>
<td>magnetic resonance spectrometry</td>
</tr>
<tr>
<td>MRSI</td>
<td>magnetic resonance spectroscopic imaging</td>
</tr>
<tr>
<td>MS</td>
<td>mass spectrometry</td>
</tr>
<tr>
<td>MSU</td>
<td>mid-stream urine</td>
</tr>
<tr>
<td>MVSA</td>
<td>multivariate statistical analysis</td>
</tr>
<tr>
<td>NAD</td>
<td>nicotinamide adenine dinucleotide</td>
</tr>
<tr>
<td>NaN$_3$</td>
<td>sodium azide</td>
</tr>
<tr>
<td>NMR</td>
<td>nuclear magnetic resonance</td>
</tr>
<tr>
<td>NOESY</td>
<td>nuclear Overhauser effect spectroscopy</td>
</tr>
<tr>
<td>noCaP</td>
<td>no prostate cancer (negative biopsy)</td>
</tr>
<tr>
<td>NS</td>
<td>not significant</td>
</tr>
<tr>
<td>OPLS / O2PLS</td>
<td>orthogonal projections to latent structures / 2-way OPLS</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>PAP</td>
<td>prostatic acid phosphatase</td>
</tr>
<tr>
<td>PBS</td>
<td>phosphate buffered saline</td>
</tr>
<tr>
<td>PC</td>
<td>principal component</td>
</tr>
<tr>
<td>PCA</td>
<td>principal components analysis</td>
</tr>
<tr>
<td>PCA3</td>
<td>prostate cancer antigen 3</td>
</tr>
<tr>
<td>PET</td>
<td>positron emission tomography</td>
</tr>
<tr>
<td>PEUW</td>
<td>post-ejaculate urethral washings</td>
</tr>
<tr>
<td>phi</td>
<td>Prostate Health Index</td>
</tr>
<tr>
<td>PLS</td>
<td>partial least squares</td>
</tr>
<tr>
<td>PLS-DA</td>
<td>partial least squares discriminant analysis</td>
</tr>
<tr>
<td>PQN</td>
<td>probabilistic quotient normalization</td>
</tr>
<tr>
<td>PRIAS</td>
<td>Prostate Cancer Research International: Active Surveillance</td>
</tr>
<tr>
<td>PSA</td>
<td>prostate specific antigen</td>
</tr>
<tr>
<td>PSMA</td>
<td>prostate specific membrane antigen</td>
</tr>
<tr>
<td>PZ</td>
<td>peripheral zone of the prostate</td>
</tr>
<tr>
<td>RBWH</td>
<td>Royal Brisbane and Women's Hospital</td>
</tr>
<tr>
<td>ROC</td>
<td>receiver operating characteristic</td>
</tr>
<tr>
<td>RG</td>
<td>reference gene</td>
</tr>
<tr>
<td>RP</td>
<td>radical prostatectomy</td>
</tr>
<tr>
<td>Tar</td>
<td>tartrate</td>
</tr>
<tr>
<td>TMAO</td>
<td>trimethylamine N-oxide</td>
</tr>
<tr>
<td>TRUSBx</td>
<td>transrectal ultrasound guided biopsy</td>
</tr>
<tr>
<td>TPBx</td>
<td>transperineal biopsy</td>
</tr>
</tbody>
</table>
Part 1 - Introduction
Chapter 1 - Introduction

Overview

Prostate cancer (CaP) is the most common internal cancer in men and the sixth leading cause of cancer-related death worldwide\textsuperscript{1-3}. Most CaP arises from the acinar epithelium in the peripheral zone (PZ) of the prostate, accounting for 70\% of all CaP diagnoses. The incidence of CaP increases with age and has risen worldwide, with highest incidence observed in highest resource countries\textsuperscript{1}. Early diagnosis and treatment of localised CaP, with radical prostatectomy (RP) or radiotherapy prior to spread outside the prostate, improves overall and cancer-specific survival compared with observation\textsuperscript{4}. Metastatic CaP is predominantly treated with androgen deprivation therapy (ADT) in conjunction with or followed by chemotherapy to improve overall survival, but is incurable. CaP-specific mortality rates are highest in less developed regions and have improved among developed countries\textsuperscript{1}, presumably due to earlier detection and treatment. Thus, biomarkers to accurately diagnose localised CaP are required to enable early treatment and improved outcomes for men.

Biomarkers are essential in aiding medical diagnosis and treatment and are predominantly found in bodily fluids including blood and urine. Serum prostatic specific antigen (PSA) is the main biomarker used to improve diagnosis of localised CaP. However, serum PSA is an imperfect test due to unacceptable sensitivity and specificity in detecting aggressive, clinically significant CaP (csCaP). Furthermore, judicious use of serum PSA has resulted in over diagnosis of indolent, non-aggressive CaP leading to unnecessary morbidity inherent to the diagnostic and treatment process\textsuperscript{5}. These limitations have reduced CaP detection overall, including csCaP, and driven the discovery of biomarkers in serum, such as the Prostate Health Index (phi\textsuperscript{TM}), and urine, such as prostate cancer antigen 3 (PCA3) in combination with TMPRSS2:ERG fusion gene\textsuperscript{6-9}. These tests may improve diagnostic accuracy\textsuperscript{10}, but are currently used as adjunctive tests and are potentially not cost-effective as standalone detection tests\textsuperscript{11}.

Prostatic fluid in ejaculate contributes up to 40\% of due to physiological prostatic smooth muscle contraction and is a promising biofluid for biomarker discovery\textsuperscript{12}. Contributed proteins (such as PSA and prostatic acid phosphatase; PAP) as well as metabolites (such as citrate, spermine and myo-inositol) and ions (such as zinc, calcium, magnesium) physiologically aid spermatozoa in fertilization\textsuperscript{12} but are also altered due to CaP. Indeed
PSA was first discovered in ejaculate, which also harbours PAP, a biomarker used prior to PSA\textsuperscript{13}. Following discovery of malignant prostatic cells in the ejaculates of men with CaP\textsuperscript{14}, small proof-of-concept studies undertaking analysis of genes (PCA3, Hepsin) and microRNAs in the cellular component of ejaculate have demonstrated superior detection to PSA alone\textsuperscript{15,16}. Furthermore, the ability to ejaculate among donors has been shown to correlate with improved survival compared to age-matched controls with CaP\textsuperscript{17}, hypothetically due to adequate erectile function, favourable cardiovascular status and long term survival and treatment benefit for the patient\textsuperscript{18}. Thus, ejaculate is theoretically an ideal biofluid for sourcing biomarkers for improved detection of localised csCaP. Prostatic fluid can also be obtained as expressed prostatic secretions (EPS), by first performing prostatic massage at time of digital rectal examination (DRE) then collecting the first-catch (5ml) after commencing urination. A novel biofluid described in this thesis is post-ejaculate urethral washings (PEUW), which is hypothesised to contain prostatic fluid from remnant ejaculate in the urethra and be a potential physiological alternative to EPS.

Metabolomics is a modern biomarker discovery approach that quantifies small metabolites in biological samples\textsuperscript{19}, most commonly using nuclear magnetic resonance (NMR) or mass spectrometry (MS)\textsuperscript{20,21}. NMR-based metabolomics analysis of ejaculate samples has been shown by multiple investigators to improve diagnosis compared to serum PSA\textsuperscript{22,23}, and represents a highly sensitive and reproducible analysis method with affordable sample-to-sample costs\textsuperscript{20}. Early studies also indicate in vivo monitoring of metabolite profiles in tumours using MRSI to aide diagnosis and active surveillance\textsuperscript{24}.

This thesis investigated molecular and metabolomic detection methods to study prostatic fluid derived from ejaculate and PEUW to improve CaP diagnosis. The thesis is broadly structured as Introduction, Methods, Results and Discussion.

Within Part 1 (Introduction), chapter two of this thesis contains a comprehensive literature review on CaP biomarkers, with particular focus on metabolomics, including the experimental workflow, previous studies and future directions. This literature review is presented as a peer-reviewed book chapter and review article, as well as a literature update since the included manuscripts were published.

Part 2 (Methods) contains chapter three of this thesis, which described a NMR-based metabolomics study to investigate the inhibition of PAP to improve metabolite stability within ejaculate in order to optimise sample collection protocols for future clinical metabolomics studies. This study is presented as a published manuscript.
Part 3 (Results) of this thesis consists of three chapters. Chapter four investigated the utility of ejaculate for non-invasive CaP diagnosis and includes two manuscripts, with one accepted and one submitted for publication. The studies used different analytical platforms to detect CaP biomarkers, including gene detection by PCR in the ejaculate epithelial cell fraction as well as metabolite profiling of ejaculate using NMR spectroscopy.

Chapter five of this thesis investigated the feasibility and performance of PEUW-based biomarkers and includes a published manuscript and preliminary NMR-based metabolomics investigation. First, the presence of prostatic RNA biomarkers and their diagnostic performance in the epithelial cell fraction of PEUW was investigated. Second, PEUW was analysed using NMR spectroscopy with three objectives. Initially, to determine if prostatic metabolite biomarkers are present, both in clinical samples and in a dedicated cohort comparing with mid-stream urine. Then the kinetic behaviour of these markers was characterised to determine if they behave similarly to that observed in ejaculate. Finally, metabolite profiles in PEUW collected from a clinical cohort were analysed to detect and characterise CaP.

Chapter six contains a published clinical trial protocol to demonstrate the translational potential of prostatic fluid markers to characterise and monitor CaP. Specifically, prostatic fluid marker use, as well as magnetic resonance imaging (MRI), is described to detect and monitor CaP in the presence of disease modifying agents (metformin, atorvastatin).

Part 4 (Discussion) includes chapter seven which contains a summary of the thesis findings and discussion regarding future directions for research into diagnosis of CaP.
Aims

The overall aim of this thesis was to use molecular and metabolomic detection methods to study prostatic fluid to improve CaP diagnosis. The specific aims of this thesis were to:

1. Critically appraise the available evidence on CaP biomarkers from prostatic fluid \textit{in vitro} with particular focus on NMR-based metabolomics and imaging \textit{in vivo}, as well as provide a basic outline of a clinical NMR-based metabolomics study.

2. Investigate the performance of tartrate as an inhibitor of PAP, as well as an internal concentration standard by examining ejaculate metabolite profile variation, at different tartrate concentrations and temperatures and evaluate any negative effects on established ejaculate processing methods for molecular (mRNA) analysis.

3. Examine the predictive ability of ejaculate biomarkers, specifically molecular (PCA3 and Hepsin mRNA, novel microRNAs) and metabolomic analysis of ejaculate, together with serum PSA to predict and characterise CaP in a high-risk clinical cohort.

4. Investigate the presence and diagnostic potential of CaP-specific biomarkers in PEUW, including molecular markers in comparison with the same markers in ejaculate and metabolite biomarkers. Metabolite biomarkers will be characterised against mid-stream urine (MSU) in young men, kinetically over time and in a “high risk” clinical cohort to detect and characterise CaP.

5. Describe a phase II “window” clinical trial utilising NMR-based markers \textit{in vitro} (prostatic fluid) and \textit{in vivo} (MRSI) to investigate if atorvastatin and metformin by themselves and together, favourably alter selected parameters in a group of clinically-localized, aggressive tumours.
Chapter 2 – Molecular and metabolomic biomarkers for non-invasive prostate cancer diagnosis and monitoring

2.1 Synopsis

The aim of this chapter is to critically appraise the available evidence on CaP biomarkers from prostatic fluid with particular focus on NMR-based metabolomics. The evolution of NMR-based metabolomics and requirements of a clinical NMR-based metabolomics study will be reviewed, while current approaches in elucidating tumour biology \textit{in vitro} and \textit{in vivo} to improve diagnosis and monitoring in clinical practice will be outlined. This literature review is presented as a peer-reviewed book chapter (section 2.2) and review article (section 2.3), as well as a literature update (section 2.4) since the included manuscripts were published.
2.2 “NMR-based Metabolomics: Global Analysis of Metabolites to Address Problems in Prostate Cancer” (published manuscript)

The published manuscript entitled, “NMR-based Metabolomics: Global Analysis of Metabolites to Address Problems in Prostate Cancer” has been published as a peer-reviewed book chapter in Breast, Cervical and Prostate Cancer (iConcept Press. Tokwawan, Kowloon, Hong Kong; http://www.iconceptpress.com/books/breast-cervical-and-prostate-cancer/).

Since publication, this chapter is the most viewed in this book (viewed 2333 as at 10/02/17), which is the most viewed medical book by this publisher (viewed 38895 as at 10/02/17). This chapter has been downloaded 557 times (as at 10/02/17) from UQ eSpace.

The contributions of the co-authors to this manuscript were as follows: Literature review and summary was performed by Matthew Roberts, the PhD candidate, under the supervision of Dr Horst Schirra and Prof RA “Frank” Gardiner. Prof Martin Lavin provided critical manuscript revisions. Preparation of the manuscript and associated figures was performed by Matthew Roberts, the PhD candidate.

The manuscript text is presented as originally published by the publisher. To maintain a consistent style between chapters in this thesis, abbreviations, placement of figures and tables within the text, as well as numbering of pages, figures and tables has been adjusted. The references have been collated in the section, ‘Bibliography’, with others contained in this thesis.
NMR-based Metabolomics: Global Analysis of Metabolites to Address Problems in Prostate Cancer

Matthew J. Roberts¹,², Horst J. Schirra², Martin F. Lavin¹,³, Robert A. Gardiner¹,⁴

1. University of Queensland Centre for Clinical Research, Brisbane QLD, 4029, Australia
2. Centre for Advanced Imaging, The University of Queensland, Brisbane QLD, 4072, Australia
3. Queensland Institute of Medical Research, Radiation Biology and Oncology, Brisbane, QLD 4029, Australia
4. Department of Urology, Royal Brisbane and Women’s Hospital, Brisbane QLD, 4029, Australia
2.2.1 Introduction

Cancer significantly contributes to the worldwide burden of disease and premature death across many countries. Consequently, current oncology research focuses on discovering and validating new biomarkers to improve early detection. These efforts are of worldwide importance in detecting significant cancer while it is still localised and in lessening associated morbidities and death.

Biomarkers have mostly been sourced from non- or minimally invasive biofluids, such as blood, urine, and biopsy tissue. Traditionally, biomarkers were limited to circulating end-products of altered cellular function in cancer. However, technology advances and emergence of the –omics sciences have improved analysis of genes, gene expression, proteins and metabolites alike – on both an individual and system-wide scale. This field of research, termed “systems biology”, has allowed for molecules at all levels of the cellular hierarchy to be considered as biomarkers. Continuous improvements in sensitivity, resolution and precision of these analytical techniques produces large datasets, allowing for simultaneous characterisation of, ideally all, compounds in a single sample. Subsequent statistical analysis of these datasets and their interpretation with respect to cellular function is the basis of the different -omics technologies, such as genomics, transcriptomics, proteomics and metabolomics.

In this chapter, we will describe principles and processes that are involved in investigating biological or clinical problems with NMR-based metabolomics - an approach that involves the global analysis of metabolites. In writing for the scope of this book, we have broken this chapter into three sections: (1) First we will describe and illustrate the methods commonly used in NMR-based metabolomics, including spectral processing, data treatment and subsequent statistical analysis. (2) Secondly, we will use CaP as a case study to illustrate how NMR-based metabolomics can be applied to a clinical problem. CaP is the second most common type of cancer and the sixth leading cause of cancer-related death worldwide\(^1\text{--}^3\). The diagnosis of prostate cancer is currently problematic for a number of reasons that include lack of sensitive and specific tumour markers as well as limitations due to morbidity inherent with the biopsy diagnosis process. Furthermore, many patients harbour early prostate cancer with insignificant tumours that may not progress to produce clinical problems. (3) Lastly, we will briefly outline the future directions for the role of NMR-based metabolomics, including personalized medicine and integration with other –omics
datasets, in order to create a holistic, systems biology approach to solving clinical problems.

Outlining the processes, applications and potential of metabolomics will be of assistance to biostatisticians and bioinformaticians who may be interested in expanding into this area of research. Similarly, we aim to inspire scientists and clinicians who are interested in applying this approach to a scientific or clinical problem.

2.2.2 Metabolomics: History and Methods

2.2.2.1 What is Metabolomics?

Metabolomics has been highlighted as a technique that is unique and exciting in biomarker discovery\textsuperscript{25-27}. It is the quantification of all small molecular weight metabolites to accurately define the metabolite composition of a biological sample\textsuperscript{19}. The term “metabolomics” is often used interchangeably with “metabonomics”, which is defined as ‘the quantitative measurement of the dynamic multiparametric metabolic response of living systems to pathophysiological stimuli’\textsuperscript{28}. The technical approach to both metabonomics and metabolomics is similar, involving measurement and analysis of metabolite data for a given sample. Conceptually, however, the objective and application of these techniques is slightly different: Metabolomics seeks to describe the composition of complex biological samples, while metabonomics aims to map and understand the change of a biological system in response to external or artificial stimuli. In the current chapter, we will use the term “metabolomics”, while both metabolomics and metabonomics will be discussed. The advent and development of metabolomics/metabonomics has largely been possible due to advances in analytical techniques, such as NMR spectroscopy and MS, as well as chromatographic separation techniques. As the authors have expertise in NMR spectroscopy, the focus of this chapter will be on NMR-based techniques within the field of metabolomics. We refer the reader to several excellent reviews that detail the application of MS in metabolomics and systems biology\textsuperscript{29-31}.

2.2.2.2 Historical Perspective

While “metabolomics” is a recently coined term, the analysis of metabolic end products is a long-practiced ancient scientific process. About 2000-1500 BCE, analysis of urine by human taste or animal behaviour (due to the high urinary glucose concentration) helped diagnose patients with diabetes mellitus\textsuperscript{32}. In 1506, a “urine wheel” by Ulrich Pinder linked crude changes detected by human senses (colour, smell, taste) with various medical
conditions\textsuperscript{33,34}. Although qualitative measures of metabolism have been performed for centuries, the origins of quantitative studies stem from the measurement of insensible perspiration and other hydration losses in medieval Italy\textsuperscript{35}. Technological limitations in analytical chemistry hindered further advances until the early 20\textsuperscript{th} century. At this time, the development of sensitive analytical methods allowed quantification of key compounds/metabolites in urine and other sample types\textsuperscript{36}. This quantitative approach continued to develop with the introduction of various analytical techniques, such as MS\textsuperscript{37,38} and NMR\textsuperscript{39-41}, and with the application of these techniques to metabolic research\textsuperscript{42,43}. The integration of medical science and analytical chemistry at this time led to a greater understanding of metabolic perturbations in medical conditions, e.g. kidney stones among many others\textsuperscript{44}. Further improvements in metabolite profiling in the latter half of the 20\textsuperscript{th} century were aided by advances in chemometrics, the foundation behind data analysis in analytical chemistry\textsuperscript{45}. Appropriate data processing and interpretation was achieved by multivariate statistical methods, which will be outlined in more detail below.

2.2.2.3 Modern Metabolomics and its Varieties/Applications

Metabolomics continues to evolve as a field and is increasingly used in a variety of applications. Initially, biofluids were analysed specifically to quantify metabolic perturbations due to drug toxicity, disease, and other internal and external influences. For example, perturbations in steroid metabolism were used for anti-doping testing during the Los Angeles Summer Olympics in 1984\textsuperscript{46}. In the following section, the diverse applications of metabolomics are briefly described; more comprehensive accounts are available in other sources\textsuperscript{47-51}.

2.2.2.3.1 Biofluid and Excretion Analysis

Metabolomics has been used to quantify endogenous metabolites in many human biofluids, with those most commonly analysed being urine and blood (serum, plasma). Analysis of urine metabolites has shown early promise in diagnosing kidney\textsuperscript{52} and bladder tumors\textsuperscript{53}, as well as more systemic conditions such as type 2 diabetes mellitus\textsuperscript{54}. Indeed, distinct serum metabolite patterns have been characterized for abnormal clinical states including breast cancer\textsuperscript{55}, leukemia\textsuperscript{56}, sepsis and acute lung injury\textsuperscript{57}, coronary artery disease\textsuperscript{58} and obesity\textsuperscript{59,60}. Furthermore, cardiovascular health has been assessed by metabolomic analysis of faeces, linking perturbations in the metabolite profiles of gut flora to the metabolic syndrome and dyslipidemia\textsuperscript{61}. Ejaculate and EPS, have been used to characterize disturbed metabolism in prostate cancer\textsuperscript{22,23,62-64} and infertility\textsuperscript{65-67}. Studies
on cerebrospinal fluid have associated metabolite changes with brain tumours and neurodegenerative disorders, such as multiple sclerosis and Alzheimer's disease\textsuperscript{68}. Salivary metabolomics has been used to investigate oral cancer and pre-malignant changes, such as leukoplakia\textsuperscript{69}. Thus, this minimally invasive approach has enormous potential in providing valuable scientific and clinical information for medical professionals and researchers.

2.2.2.3.2 NMR Spectroscopy of Tissues and \textit{in vivo} Imaging Techniques

High-resolution magic-angle spinning (HR-MAS) NMR spectroscopy can be used to perform non-destructive metabolite profiling of tissue or other solid samples\textsuperscript{70}. That means that after HR-MAS NMR spectroscopy, further testing of tissue samples can be performed such as histopathological evaluation, the current gold standard in disease diagnosis, or genome and protein sequencing. As a result, HR-MAS NMR has been used to investigate a number of disease states\textsuperscript{71-77}.

Metabolomic analysis can also be performed directly \textit{in vivo}, largely owing to advances in magnetic resonance spectroscopic imaging and positron emission tomography (PET). A standard clinical MRI scan uses similar physical concepts to NMR, but takes many scans across a section of living tissue. This data is processed to produce an anatomically correct image based on physical properties of the tissue. A magnetic resonance spectroscopy imaging sequence is able to produce NMR-like spectra for a targeted volume segment in the body, allowing for visualization of metabolite content in that anatomical location. Metabolic alterations measured \textit{in vivo} have been shown to correlate with histopathology\textsuperscript{78,79}. Furthermore, \textit{in vivo} metabolomics is being used to monitor the response to various therapies, such as radiotherapy and chemotherapy\textsuperscript{80,81}.

Different MRI techniques allow the investigation of different phenomena. Dynamic contrast-enhanced MRI uses the uptake and elimination of contrast agents, such as gadolinium, to distinguish between different tissues. The use of dynamic contrast-enhanced MRI in oncology is based on the premise that cancer cells have a higher metabolism, and thus a higher uptake and elimination of gadolinium contrast\textsuperscript{82}. Diffusion weighted imaging, initially used to investigate connectivity between different brain regions\textsuperscript{83}, uses slower water diffusion in cancerous tissues compared with surrounding healthy cells due to a higher nuclear content and cellular density coupled with extracellular changes\textsuperscript{84}. Recently, these MRI techniques have been combined into multiparametric MRI, which has been shown to increase accuracy in cancer detection\textsuperscript{85}. The value of these \textit{in
vivo applications and their role in oncology is commonly described and reviewed in radiology literature\textsuperscript{85-88}. Further understanding of altered metabolism in cancer and identification of abnormal pathways facilitates imaging using PET in combination with computed tomography (CT) via PET/CT\textsuperscript{89,90}. After identifying metabolites, that are either preferentially used or upregulated within particular pathways, nuclear isotopes can be chemically attached either to these metabolites or to metabolite analogues. The emission of positrons from these isotopes can then be measured as gamma rays and superimposed on a CT scan during PET/CT scanning. For example, most cancer cells display heightened glycolysis. Thus, fluoro-deoxy-glucose, containing a radiolabelled positron emitter such as \textsuperscript{18}F, can be administered and taken up by cancer cells, which are highlighted\textsuperscript{91}. In addition, PET is able to distinguish specific cancers, e.g. \textsuperscript{11}C-choline PET is used to detect prostate cancer\textsuperscript{92,93}. The application of PET in other clinical scenarios is diverse, but widespread use is limited by logistical and financial constraints\textsuperscript{94-96}. Recently, a novel method has been proposed to perform in vivo metabolomics during surgery by using real-time MS analysis of the smoke produced from electric cautery to biochemically recognize malignant/diseased tissue in which macroscopic changes are not present\textsuperscript{97,98}. Although major development is required before clinical use, the initial concept is intriguing in its potential to improve surgical accuracy and treatment outcomes following cancer surgery.

2.2.2.4 Integration with other -Oomics Sciences

Metabolites are part of the complex and interconnected cellular hierarchy involving DNA, RNA, proteins and metabolites, as outlined in Figure 2.1. Consequently, providing an understanding of the mutual relationships between genomic, transcriptomic, proteomic and metabolomic data is a major aim of systems biology.

Integration of multi-omics data sets is already providing insight into biological processes. This integration is enabled by the availability of new statistical methods to correlate information contained in multiple large datasets\textsuperscript{99}. Furthermore, ever increasing genome-wide association studies (GWAS) are identifying multiple risk loci associated with various disease states. However, the penetrance of these loci, and therefore their relevance, remains unclear. By integrating metabolomics and other –omics sciences with GWAS, it is expected that identification of loci with a high penetrance or phenotypic manifestations will
Figure 2.1: Major targets for exploratory analysis in systems biology. The flow of information and biochemical processes between various levels in cellular organisation illustrates the progression from genotype to phenotype (solid arrows). The individual levels of cellular organisation are also regulated by complex intrinsic feedback mechanisms (dashed arrows). Adapted from 20.

be unveiled\textsuperscript{100,101}. Integration of –omics data sets has major potential in oncology\textsuperscript{102} and some studies have used a targeted approach in relating datasets obtained by different analytical methods\textsuperscript{99}. A recently published study related metabolomic changes to genomic disturbances in CaP tissue to demonstrate alterations in m-aconitase and acetyl citrate lyase. Phospholipase A2 group VII and choline kinase α were responsible for altered citrate and choline levels, respectively\textsuperscript{103}. Other integrative works investigated colorectal cancer\textsuperscript{104}, heart failure\textsuperscript{105} and other diseases\textsuperscript{106,107}. These studies show impressive proof of concept of this new approach. As a result, multivariate statistical analysis and integration of large and multi-omics data sets are a valuable strategy for further investigation.

Another key frontier in systems biology is the creation of genome-wide \textit{in silico} models of cellular metabolism that are able to incorporate and integrate multi-omic data\textsuperscript{108}. Such reconstructed metabolic networks have already suggested improvements for targeted treatment strategies of cholesterol homeostasis in human cellular models\textsuperscript{109}.

The first genome-scale model of human metabolism published in 2007 was a promising milestone\textsuperscript{110}. The next step in metabolic modelling in oncology will be to construct cancer-specific metabolic models that will incorporate –omics data. This daunting task requires improved curation and annotation of genome databases, as well as integration of high
quality –omics datasets from studies of specific cancer types, or of other disease states to create disease-specific reconstructions. Public accessibility and maintenance of data from publicly funded research is critical to these efforts\textsuperscript{111}, as is data management\textsuperscript{112} and interpretation\textsuperscript{113}.

### 2.2.2.5 Analytical Techniques

Of the many analytical techniques that are used in metabolomics to investigate physiological and pathological states, NMR spectroscopy and gas chromatography (GC) or liquid chromatography (LC) combined with MS are the two most commonly used methods. Both have low running costs, are diverse in sample type and allow for accurate metabolite identification. Other techniques in use are ultra-performance LC-MS, inductively coupled plasma MS, Fourier-transform MS, Fourier-transform infrared spectrometry and thin layer chromatography\textsuperscript{114}. We will briefly describe the processes involved in MS-based metabolomics, with more extensive reviews available elsewhere\textsuperscript{29-31,115-117}. Subsequently, we will focus on NMR-based metabolomics, describing the basic principles and statistical approaches that are currently in use.

#### 2.2.2.5.1 GC-/LC-MS

Mass spectrometry detects ionized compounds in biological samples according to their mass/charge ($m/z$) ratio following chromatographic (e.g. GC or LC) separation and metabolites can be identified in the resulting mass spectrum with reference to internal standards\textsuperscript{118}. LC- and GC-MS are well used techniques in metabolite analysis and have similar sensitivity, with the major difference being that GC requires more sample preparation (derivatization) and higher analysis temperatures, thus LC may be preferred for this reason\textsuperscript{117,119}. We will briefly outline the basic processes in MS-based metabolomics, which are sample preparation, separation (via liquid-/gas-chromatography), ionization, mass analysis and detection, and finally, data processing.

Sample preparation for MS is dependent on the type of sample. Simple biofluid preparation often involves removing macromolecules through protein precipitation and centrifugation or filtration. Similarly, sampling of the exometabolome (metabolites secreted by cells or organisms into the growth medium) is straightforward. In contrast, to obtain intracellular metabolites, tissues or cells need to be extracted in an appropriate solvent system\textsuperscript{117}). As different solvent systems are biased towards particular classes of metabolites (e.g. polar extraction systems yielding predominantly polar metabolites), the
exact choice of solvent depends upon the metabolites of interest. To achieve consistency between samples, internal standards are typically added during/after extraction\textsuperscript{117}.

In GC-MS, derivatization is a further necessary preparation step, which is applied to non-volatile metabolite classes, such as amino and organic acids, sugars, amines and lipids, to render them volatile and thermally stable for GC\textsuperscript{115-117}. Derivatization can introduce bias towards individual metabolites if the derivatising agents are not provided in excess, as the derivatising reactions have different efficiencies with different metabolites\textsuperscript{117}. In addition, metabolites with multiple exchangeable protons will create multiple derivatization products that will show up as separate peaks, thus complicating the final mass spectrum. For GC-MS, electron impact ionization is almost exclusively used\textsuperscript{115}.

LC-MS is rapidly replacing GC-MS as method of choice in metabolomics, as both methods are similarly sensitive, but sample preparation for LC-MS is simpler, because derivatization is not required\textsuperscript{120}. LC typically runs as reverse-phase high-performance LC (HPLC), or recently even as ultra-performance LC, and electrospray ionization is typically used in LC-MS systems\textsuperscript{117}. Electrospray ionization MS can run in positive or negative ionization mode, and because individual metabolites are generally only detected in one of those two modes, both ionization modes need to be run to improve coverage of the metabolome\textsuperscript{116}.

In the resulting mass spectrum, metabolites are quantified by external calibration or by comparison with internal standards\textsuperscript{116}. GC-MS experiments may also require the use of deconvolution software to adequately analyse overlapping chromatographic peaks\textsuperscript{116}. To allow comparable results between experiments, data may also undergo further pre-treatment steps, including spectral alignment and automated picking of metabolite peaks\textsuperscript{121}. The subsequent multivariate analysis of processed MS data is similar to data obtained by NMR-spectroscopy.

2.2.2.5.2 NMR Spectroscopy

NMR spectroscopy is a quantitative technique used to accurately determine metabolite concentrations in samples. Chemical compounds in biological samples are identified by their characteristic peak patterns and signal positions in the NMR spectrum with the aid of online databases\textsuperscript{122} (http://www.metabolomicssociety.org/database). More comprehensive accounts of NMR theory and application in metabolomics are available in dedicated texts\textsuperscript{123}. A comparison of the strengths and limitations of NMR and MS is provided in Table 2.1.
<table>
<thead>
<tr>
<th>Technique</th>
<th>Advantages</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>NMR</td>
<td>• high reproducibility</td>
<td>• low sensitivity</td>
</tr>
<tr>
<td></td>
<td>• high resolution</td>
<td>• peak overlap</td>
</tr>
<tr>
<td></td>
<td>• non-destructive</td>
<td>• libraries of limited use due to complex matrix</td>
</tr>
<tr>
<td></td>
<td>• quantitative</td>
<td>• long acquisition times for</td>
</tr>
<tr>
<td></td>
<td>• low running costs</td>
<td>heteronuclear techniques, e.g.</td>
</tr>
<tr>
<td></td>
<td>• minimal sample preparation/no</td>
<td>13C</td>
</tr>
<tr>
<td></td>
<td>derivatization</td>
<td>• high initial capital cost</td>
</tr>
<tr>
<td></td>
<td>• unbiased metabolite profile</td>
<td>• reduced availability</td>
</tr>
<tr>
<td></td>
<td>• analysis of tissue (HR-MAS)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• translation to in vivo (MRI)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• rapid analysis</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• ability for automation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• structural identification (2D, 3D)</td>
<td></td>
</tr>
<tr>
<td>GC-MS</td>
<td>• high sensitivity</td>
<td>• slow</td>
</tr>
<tr>
<td></td>
<td>• large linear range</td>
<td>• sample unable to be re-used</td>
</tr>
<tr>
<td></td>
<td>• robust</td>
<td>• requires chemical derivatization</td>
</tr>
<tr>
<td></td>
<td>• identification of wide range of metabolites</td>
<td>• potentially multiple derivatization products</td>
</tr>
<tr>
<td></td>
<td>• (wider range with LC-MS)</td>
<td>for metabolites</td>
</tr>
<tr>
<td></td>
<td>• analysis of complex biofluids</td>
<td>• many analytes thermally unstable</td>
</tr>
<tr>
<td></td>
<td>• non-targeted</td>
<td>• metabolite weight limitation</td>
</tr>
<tr>
<td></td>
<td>• established databases</td>
<td>(&lt;1400 Da)</td>
</tr>
<tr>
<td></td>
<td>• widely available and</td>
<td></td>
</tr>
<tr>
<td></td>
<td>comparably low capital cost</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• preferred for targeted analysis</td>
<td></td>
</tr>
<tr>
<td>LC-MS</td>
<td>• high sensitivity</td>
<td>• slow</td>
</tr>
<tr>
<td></td>
<td>• high reproducibility</td>
<td>• limited commercial libraries</td>
</tr>
<tr>
<td></td>
<td>• large linear range</td>
<td>• sample unable to be re-used</td>
</tr>
<tr>
<td></td>
<td>• no chemical derivatization needed</td>
<td>• generation of adducts</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• higher capital cost (HPLC-MS)</td>
</tr>
</tbody>
</table>

Table 2.1: Comparison of the advantages and disadvantages of NMR, GC-MS and LC-MS

114,119,124,125.

2.2.2.6 Processing of NMR Data

Processing of NMR data comprises four steps: Fourier transformation, phase correction, baseline correction and calibration. Fourier transformation transforms the raw real-time
data into the frequency domain, and phase correction corrects the phase of the resulting NMR spectrum. Baseline correction ensures a constant zero baseline across an NMR spectrum, and calibration is needed to ensure a consistent chemical shift scale/axis across all spectra. Inadequate processing introduces artefacts that confound statistical analysis and jeopardize data integrity (see Figure 2.2). Minimum standards for reporting and processing have been outlined, and continue to be a good guide for authors publishing metabolomics research.

![Figure 2.2: Sequential steps of processing NMR spectra. Shown is a spectrum of human ejaculate in the region of citrate (2.45 – 2.65 ppm) as example. Correct baseline position is illustrated with dotted lines. a - Spectrum after Fourier transformation. The phases of the NMR signals are partly dispersive and in need of phase correction. b - Spectrum after phase correction. Globally reduced metabolite peak intensities and negative values for the baseline occur across the entire spectrum. c - Spectrum after baseline correction, but requiring calibration to a chemical shift standard. Chemical shift values are incorrect across the entire spectrum, resulting in incorrect metabolite identification. d - Correctly processed spectrum suitable for data reduction.](image)

Usually, NMR spectra require a phase correction following Fourier transformation in order to achieve pure absorptive line shapes for all peaks in a NMR spectrum. Where possible, phasing should be performed automatically or by the same operator across all samples to ensure consistency. Incorrect phasing can distort peak integrals and thus, the subsequent multivariate statistical analysis (MVSA; see Figure 2.2A/2.2B).

Baseline correction is the third critical processing step in producing consistent, comparable and reliable data in NMR spectroscopy. At a minimum, the y-offset of the entire spectrum is corrected to be zero. However, baseline corrections are often more complex, using
spline, polynomial or other mathematical functions to accomplish a zero baseline over the whole spectrum.

As signal intensities are calculated with reference to zero, inadequate baseline correction will distort spectral peak intensities (Figure 2.2B/2.2C), and compromise the subsequent MVSA.

Importantly, the chemical shift axis of each NMR spectrum must be adequately calibrated using a chemical shift standard, such as (deuterated) 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS). In samples containing a high protein content (such as plasma), DSS cannot be used as internal standard. Thus, other endogenous metabolites that are present across all samples, such as lactate, glucose or formate, are used as internal reference. The alternative is to use DSS as an external standard by either inserting a capillary with DSS in deuterium oxide ($\text{D}_2\text{O}$) into the NMR tube, or inserting the sample in a capillary into a tube containing DSS in $\text{D}_2\text{O}$. The use of 4,4-dimethyl-4-silapentane-1-ammonium trifluoroacetate has also been suggested as chemical shift standard that is not affected by protein binding\textsuperscript{127}. Chemical shift calibration ensures consistent global alignment of spectra in a metabolomics data set which is critical for statistical analysis. In addition, correct spectral alignment is required for reliable metabolite identification. However, even in correctly calibrated spectra, individual peaks can still exhibit differences in chemical shift between individual spectra due to differences in sample pH and ionic strength. These can be corrected post-processing by various automatic peak alignment procedures\textsuperscript{128-133}.

2.2.2.7 Statistical Pre-processing

2.2.2.7.1 Data Reduction

After processing of raw NMR data, further processing steps are needed to prepare data for MVSA, which are usually termed “statistical pre-processing”. Reducing the full resolution data into small segments of equal width, called bins, or “buckets” (Figure 2.3), is the most widespread method of data reduction in chemometrics\textsuperscript{134}. Compared with analysis at full resolution, this method considerably reduces the size of the data matrix in MVSA, and is particularly helpful when peak positions or widths vary slightly due to changes in pH, ionic strength or other factors. However, due to decreasing data resolution, bucketing can complicate metabolite identification following data analysis. Other pre-processing methods that can be used, particularly in targeted metabolomics, include deconvolution, peak-picking, and weighting factors\textsuperscript{126}.
2.2.2.7.2 Normalization

After data reduction, data need to be normalized to produce data that are comparable between samples\textsuperscript{135}. Normalization is a row operation in the data matrix, and different normalization methods are used to obtain the best representation of the data. In total integral normalization, or normalizing to total intensity, the spectral intensity in each bucket is divided by the total intensity of each spectrum. This procedure normalizes differences between spectra due to sample concentration/dilution, e.g. due to different water content between samples. However, total integral normalization is vulnerable to distortions when one or a few intense signals change considerably between spectra.

Another method involves normalization to an internal reference compound. For metabolomics analysis of urine, normalization to creatinine has been widely used\textsuperscript{136,137}. For physiological reasons, urine creatinine is believed to be a suitable indicator of urine concentration as creatinine excretion is constant. However, creatinine normalization has limitations because it will be confounded by any background pathophysiology that alters serum creatinine concentration or creatinine excretion, such as in kidney disease. In these cases, creatinine normalization is not suitable. Furthermore, variations in chemical properties within the sample can distort creatinine alignment, so that other metabolites, especially creatine, will overlap with the creatinine signal and thus impede the proper measurement of creatinine concentration\textsuperscript{123}. 

\textbf{Figure 2.3:} Illustration of data reduction by “bucketing”. a - NMR spectrum of human ejaculate with suitable pre-processing (segment shown). b – Same spectrum segmented/data reduced into buckets of 0.04 ppm width across the region 8 – 0 ppm. The region around the water signal from 5.08 – 4.52 ppm was excluded due to artefacts from imperfect water suppression. Note that the area in each individual bucket is integrated across the spectral width of 0.04 ppm and then normalized, yielding intensities similar to a histogram.
Probabilistic quotient normalization (PQN) is a method that reduces variation caused by large changes in the intensity of one or a few signals across samples, as shown in Figure 2.4. Thus, PQN can overcome the main weakness of total integral normalization. In PQN, which is usually performed after total integral normalization, each variable (bucket) in a spectrum is first divided by the intensity of the same variable in a reference spectrum. Afterwards the full spectrum is divided by the median of these quotients. This procedure is repeated for all spectra in a data set, using the same reference spectrum\textsuperscript{138}.

![Figure 2.4: Schematic depiction of spectra containing four NMR signals A-D, following two different normalization methods. a – raw spectra with marked intensity variation present in the first peak A and identical intensities of remaining peaks between all spectra. b - Integral normalization (normalization to total intensity) reduces variation, and therefore influence of the dominant signal A, but also alters relative intensities of the smaller, previously identical signals. c - Probabilistic quotient normalization partially reduces variation and influence of the larger signal A, while maintaining the original relationships between smaller peaks to allow optimal comparison during MVSA.](image)

Recently, Kohl \textit{et al.} compared many normalization methods, with some derived from genomic data analysis, and recommended more advanced methods, such as quantile normalization for datasets of $n \geq 50$, as well as Cubic Spline Normalization and Variance Stabilisation Normalization\textsuperscript{139}.

### 2.2.2.7.3 Scaling Effects

Following normalization, metabolomic data must be appropriately scaled, or transformed in a column operation in a way that changes how much signals of large and small intensity, respectively, influence the data analysis\textsuperscript{126,135}. The objective is to reduce noise and maximize information content in the data. Inappropriate scaling may lead to results that highlight parts of the data unrelated to a biological factor, thus compromising the analysis and biological interpretation of the data. In metabolomics statistical analysis, three scaling methods are largely used.
Centre scaling, or mean centering, subtracts the mean value of each variable/bucket from the original data of that bucket. This method is the least manipulative, and is also best at minimizing background noise, but large relative variations in small signals may not be detected. Mean centering is usually performed mandatorily. Thus, this method is also sometimes referred to as “no scaling”, as no further scaling is performed after mean centering.

Univariate, or unit variance, scaling divides the raw data obtained after mean centering by the standard deviation of each variable. Univariate scaling gives each variable equal weighting, such that variables with small absolute but large relative variation are highlighted, but this also means that background noise and other unrelated data variation may be overemphasized and thus confound the analysis.

Pareto scaling is performed by dividing each variable by the square root of its standard deviation. This is the recommended scaling method for NMR-based metabolomic data, as it is able to increase the weighting on metabolites with smaller amplitudes, but does not overemphasize the influence of background noise.

2.2.2.8 Statistical Analysis

Statistical analysis of metabolomic data depends on the biological question studied and the design of the particular project, thus, the choice of data analysis methods varies between different projects. In addition, the methods for data analysis are continually evolving. Nevertheless, there is a core set of methods of univariate and multivariate statistical analysis that is in use for metabolomics, and minimum reporting standards for data analysis have been established.

2.2.2.8.1 Univariate Statistical Analysis

The role of univariate analysis in metabolomics is largely of a targeted nature. An example would be where metabolites of interest have been identified by MVSA, and detailed analysis of statistical significance of the individual metabolites is desired. Basic univariate methods can be used to analyse whether or not individual metabolites are significantly different between two classes. However, as with any statistical analysis, the distribution of the data determines the type of analysis used. If the data are normally distributed, t-tests, z-tests, and analysis of variance (ANOVA) may be used. In cases where the distribution is not normal, non-parametric methods such as the Kruskal-Wallis test are used. However, given the high number of variables within a metabolomics dataset, use of multiple
hypothesis testing corrections, such as Bonferroni correction or false discovery rate/Benjamini-Hochberg are absolutely imperative\textsuperscript{143}. This means that, to be significant, $p$-values need to be much smaller (e.g. $p \leq 5 \times 10^{-5}$) after correcting for multiple hypothesis testing compared to standard univariate statistical analysis\textsuperscript{144}.

### 2.2.2.8.2 Multivariate Statistical Analysis (MVSA)

Modern day metabolomics is largely based on data sets incorporating many variables, between several hundred in the case of bucketed data and up to 65,536 if 1D-NMR spectra are used at full resolution. Thus, MVSA methods which simultaneously analyse all these variables are preferred\textsuperscript{144}. MVSA determines whether there are inherent patterns or groupings within the data that correspond to biological states and also which variables are important in discriminating between the different groupings. Thus, this approach is well suited to analysing metabolomics datasets, where the aim is to correlate multiple metabolite changes with alterations in biology.

There are two general classes of MVSA methods: unsupervised methods, which analyse patterns within a data matrix $X$, and supervised methods in which the patterns in $X$ are correlated with other external data (e.g. clinical data) contained in a $Y$ matrix or $Y$ table. The advantages and disadvantages of the most common MVSA methods are summarized in Table 2.2, and will be discussed in the following sections.
<table>
<thead>
<tr>
<th>Approach</th>
<th>Method</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
</table>
| Unsupervised  | Principal components analysis (PCA)   | • Simplifies data  
• Describes variation in original data without bias  
• Groups samples with similar metabolite profiles | • Variation may be unrelated to biological question  
• Influenced by any confounders |
| Supervised    | Partial least squares (PLS)           | • Simplifies data  
• Groups samples with similar metabolite profiles  
• Extracts variation that is correlated with external data/identifiers  
• More directed to the biological question | • Possibility of introducing bias  
• Require rigorous validation |
| Orthogonal projections to latent structures (OPLS) | As PLS  
• Removes orthogonal (unrelated) variation  
• Improved knowledge extraction | | |
| 2-way OPLS (O2PLS) | As OPLS  
• Two-way data correlation between $X$ and $Y$  
• Potential for unsupervised analysis (when analysing two large datasets without external $Y$ table) | | |
| Kernel OPLS   | As OPLS  
• Improved model prediction | | |
| $O_n$PLS      | As $O_2$PLS  
• Simultaneous data correlation from multiple ($n$) matrices | | • Possibility of introducing bias  
• Require rigorous validation  
• Not widely available (commercial/open source) |
| Bi-modal $O_n$PLS | As $O_n$PLS  
• Data correlation between variables (columns) and samples (rows) | | |
| $O_n$PLS path modelling | As $O_n$PLS  
• Linkage of matrices along statistically related paths | | |

Table 2.2: Summary of multivariate statistical analysis methods used in metabolomics for information recovery. The advantages and disadvantages of unsupervised and supervised methods are outlined.
Some examples of software programs, both commercial and free/open-source, that are available to perform MVSA are given in Table 2.3.

<table>
<thead>
<tr>
<th>Software Package</th>
<th>Use in NMR/MS</th>
<th>Reference/Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commercial</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SIMCA</td>
<td>Both</td>
<td><a href="http://www.umetrics.com">http://www.umetrics.com</a></td>
</tr>
<tr>
<td>MATLAB</td>
<td>Both</td>
<td><a href="http://www.mathworks.com">http://www.mathworks.com</a></td>
</tr>
<tr>
<td>MarkerLynx/MassLynx</td>
<td>MS</td>
<td><a href="http://www.waters.com">http://www.waters.com</a></td>
</tr>
<tr>
<td>STATISTICA Data Miner</td>
<td>Both</td>
<td><a href="http://www.statsoft.com">http://www.statsoft.com</a></td>
</tr>
<tr>
<td>Analysis of Mixtures (AMIX)</td>
<td>Both</td>
<td><a href="http://www.bruker.com">http://www.bruker.com</a></td>
</tr>
<tr>
<td>Agilent Mass Profiler</td>
<td>MS</td>
<td><a href="http://metabolomics.chem.agilent.com">http://metabolomics.chem.agilent.com</a></td>
</tr>
<tr>
<td>Professional</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Progenesis CoMet</td>
<td>MS</td>
<td><a href="http://www.nonlinear.com">http://www.nonlinear.com</a></td>
</tr>
<tr>
<td>Free/Open Source</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>Both</td>
<td><a href="http://www.r-project.org">http://www.r-project.org</a></td>
</tr>
<tr>
<td>Metaboanalyst</td>
<td>Both</td>
<td><a href="http://www.metaboanalyst.ca">http://www.metaboanalyst.ca</a></td>
</tr>
<tr>
<td>MAVEN</td>
<td>MS</td>
<td><a href="http://maven.princeton.edu">http://maven.princeton.edu</a></td>
</tr>
<tr>
<td>MZmine</td>
<td>MS</td>
<td><a href="http://mzmine.sourceforge.net">http://mzmine.sourceforge.net</a></td>
</tr>
<tr>
<td>MeltDB</td>
<td>MS</td>
<td><a href="https://meltdb.cebitec.uni-bielefeld.de/">https://meltdb.cebitec.uni-bielefeld.de/</a></td>
</tr>
<tr>
<td>MetabolomeExpress</td>
<td>MS</td>
<td><a href="https://www.metabolome-express.org">https://www.metabolome-express.org</a></td>
</tr>
</tbody>
</table>

Table 2.3: Examples of software packages used to perform multivariate statistical analysis in metabolomics.

2.2.2.8.3 Unsupervised Methods in Multivariate Statistical Analysis

Often, unsupervised analysis methods are initially used in MVSA as they are excellent data exploration tools that can be either used to simplify the data (dimensionality reduction – principal- or independent-components analysis) or to group samples with similar metabolite patterns (clustering – hierarchal, partitional). Although many different methods exist, principal components analysis (PCA) is the most commonly used method of unsupervised analysis in metabolomics.

PCA simplifies the original multivariate data which form a swarm of data points in a high-dimensional statistical space, by projecting them down into a new space with comparatively few dimensions called principal components (PCs). These PCs are latent variables that describe the variation in the original data. The first PC indicates the direction in which most variation occurs in the data. Subsequent PCs are all orthogonal to each
other and sorted in order of descending amount of variation. This arrangement describes
the majority of variation in the data within the first few PCs. PCA is visualized by two types
of plots, the scores and loadings plots (Figure 2.5).

Figure 2.5: Visualization of a typical multivariate analysis (e.g. PCA, PLS etc.). Sheep urine
samples before and after road transport of 12 and 48 hours are shown as an example\textsuperscript{150}. a –
Scores plot – shows any relationships between samples, such as the presence of separate
groups or outliers. This scores plot shows similarity of the pre-transport groups for both
transport durations (open inverted triangle = 12 h, black squares = 48 h), and differences
between both post-transport groups (48 hours = black diamonds, 12 hours = open triangles). b –
Loadings plot – displays the relationship of influential variables that are responsible for the
position of outliers or groups seen in the scores plot. Note that the positions/directions of
groupings/outliers in the scores plot and responsible variables in the loadings plot correspond to
each other\textsuperscript{151}.

The scores plot illustrates the relationship and similarity of samples to each other, and
allows for inspection of groupings and outliers. Outliers can be visually identified on the
scores plot, or statistically defined as being outside the Hotelling’s 95% confidence range
across all components. Further statistical validation can be obtained using the residual
variance of the model, known as distance to model plot\textsuperscript{140,151}. The variables (buckets,
ultimately metabolites) attributable to each component in the model are illustrated in the
loadings plot (see Figure 2.5)\textsuperscript{152,153}.

The inherent advantage of unsupervised methods is that they are unbiased, i.e. they
detect any statistical variation in the data, whether or not it is related to the underlying
biological effect (e.g. differences between cancer and non-cancer samples). This property
is also their most noticeable limitation, because when confounding effects are larger than
the biological effect, unsupervised analyses will predominantly show the effects of these
confounding factors. Distorting variation may also come from uncorrelated background
variation, or from noise. For this reason, robust experimental design that limits
confounding factors and appropriate pre-processing prior to data analysis are vitally important to ensure meaningful results.

**2.2.2.8.4 Supervised Methods in MVSA**

MVSA can be improved by including external data, such as clinical data, in a $Y$ table or $Y$ matrix. This data inclusion then makes it possible to use a different class of MVSA methods, which are called supervised analysis methods. The biggest advantage of supervised methods is that they can identify the variation (and associated variables) in the biological data that is correlated (or co-varies) with the external data, i.e. they improve information recovery and thus interpretation of the biological data. The main supervised methods used in metabolomics-based biomedical research are partial least squares (PLS) and orthogonal projections to latent structures (OPLS)$^{153-155}$.

PLS is a method that seeks to identify correlation between the dataset matrix ($X$) and one or multiple variables (contained in $Y$). $Y$ data may be categorical (class identities, e.g. healthy vs. disease) or continuous (blood pressure, height etc.). If the external variable(s) are qualitative, then the method will discriminate between the corresponding classes and is known as PLS discriminant analysis (PLS-DA)$^{151}$. Figure 2.6 illustrates this distinction. Supervised analyses in metabolomics can be affected by systematic variation that is unrelated to the class, as this affects any correlation found by the analysis method.

OPLS is an improvement on PLS that separates variation in the data into two parts: one that is correlated with the biological factor(s) and one that is unrelated/orthogonal, i.e. OPLS separates $X$ into variation that is predictive of $Y$ and variation that is orthogonal to $Y$$^{58,151,156-158}$. A further development of OPLS is 2-way OPLS (O2PLS). While OPLS only correlates data in $X$ with $Y$, O2PLS is able to correlate $X$ and $Y$ with each other in both directions$^{156}$. In addition, individual variables from an O2PLS analysis can be visualized as a bivariate 1D loadings plot facilitating identification of potential metabolites$^{159}$. Both, OPLS and O2PLS have recently been preferred to PLS, as separation and correlation of predictive variation to the $Y$ table has been shown to optimize discriminant analysis, improving overall knowledge extraction$^{160-162}$. This is because both predictive and orthogonal variation can be examined, which may provide more detailed insight into the factors influencing the biological system$^{163}$. Furthermore, O2PLS can be used to correlate two different data sets with each other, e.g. metabolomic and proteomic datasets in an animal model of prostate cancer$^{99}$. If applied in this way, O2PLS is essentially an
unsupervised analysis that is able to correlate variables of different datasets, providing further insight into related structures and pathways in altered metabolic states.\(^{164}\)

![Figure 2.6](image)

Figure 2.6: Data included in a PLS analysis. The \(X\) table/matrix comprises the metabolomic data following pre-processing. Depending on the objective of the analysis, the \(Y\) table/matrix can include continuous (PLS) or categorical (PLS-DA) data to which the \(X\) data are correlated.

Different extensions of OPLS or O2PLS have been published, including kernel-OPLS, which improves model prediction.\(^{165}\) OnPLS is an extension of O2PLS which determines correlation not only between two, but multiple (\(n\)) matrices, allowing for integration of any number of datasets for a given study.\(^{166}\) Bi-modal OnPLS is an extension to OnPLS that is not only able to analyse orthogonal variation in variables (columns), but also in samples (rows).\(^{167}\) This bi-modal approach should provide more informed data analysis, of both, the variables associated to the biological question, and of confounding factors associated to particular samples. Finally, OnPLS path modelling is a method of linking multiple matrices along a set of paths that flow between data blocks. These paths are assumed to be due to a specific causative mechanism, e.g. changes over time, and are able to extract the minimum number of predictive components that have maximum covariance and correlation.\(^{168}\) Use of these recent extensions of O2PLS is not yet widespread, but highly promising in improving metabolomic data analysis.

One inherent problem of any supervised MVSA method is that, because they attempt to correlate the experimental data \(\mathbf{X}\) with external data \(\mathbf{Y}\), they are prone to introducing
bias in the analysis. This can happen due to overemphasis of spurious correlations in the data that are only coincidental and not caused by biology. As a result, MVSA models have to be rigorously validated when compared to unsupervised analysis methods. The value of a validated supervised model may be higher than that of a model originating from unsupervised methods because supervised methods are more directed toward the biological question. There are several methods of validation. The gold standard is the use of an independent set of data to test the predictive power of the original (training) set of data, combined with external cross-validation of the training data set. An established alternative is permutation analysis in which the data in the \( Y \) table are repeatedly permuted at random and the model recalculated with the permuted \( Y \) data\textsuperscript{169}. If the model is stable and correlations to \( Y \) are only of biological origin, randomization and permutation of \( Y \) data will reduce the fit and predictability of the model (see Figure 2.7A). Weak models in which correlations to \( Y \) are due to chance, rather than biology, will produce permuted models that may provide similar or superior prediction than the original model and are thus invalid (see Figure 2.7B).

![Figure 2.7: Validation by permutation analysis. a – Example of a valid model (original R\(^2\) and Q\(^2\) plotted on right side of panel), with permutations resulting in models that are less predictive (plotted on left side of panel). The x-axis indicates the distance of the permuted model to the original model, and the y-axis indicates R\(^2\) and Q\(^2\). b – Example of an invalid model, with permutations resulting in models with similar or improved predictability.](image)

Measures of validity in this context are the R\(^2\) value, which measures goodness of fit, and Q\(^2\), which measures model prediction ability. However – similar to cross-validation – permutation analysis becomes less trustworthy the lower the ratio between number of samples (\( n \)) and variables (\( k \)) is. The turning point may be near a ratio of \( n/k \) of \(< 0.02 – 0.04\), although this is not applicable for all data sets, and each study has to be evaluated on its own merits\textsuperscript{170}. In situations where validation via permutation analysis is not easily
accessible, cross-validated ANOVA can be used which uses cross-validated predictive residuals using two degrees of freedom for each component, and is more reliable than ANOVA which uses fitted residuals\textsuperscript{171}.

### 2.2.3 Current Evidence: Metabolomics in Prostate Cancer

#### 2.2.3.1 Prostate Cancer Pathophysiology

CaP is the most common internal cancer in men worldwide and is more prevalent and lethal in Western countries\textsuperscript{3}. Continually evolving methods for early CaP detection have improved outcomes due to earlier treatment and a better prognosis for patients. Current methods of detection (serum PSA and/or digital rectal examination) leading to diagnosis (via trans-rectal ultrasound guided biopsy [TRUSBx]) require improvement due to limited diagnostic sensitivity and specificity. Improved methods will help to avoid morbidity in men for whom a diagnosis of CaP remains elusive due to limitations and problems associated with TRUSBx, as is the current situation. Thus, CaP pathogenesis has been extensively studied to facilitate the discovery of new methods for determining the presence of CaP.

The prostate gland sits in the pelvis below the bladder and in front of the rectum. It is a secretory gland that contributes to the seminal plasma component of ejaculate/semen to facilitate sperm motility and egg fertilization \textit{in utero}. The secretory portion of the gland is the PZ, and constitutes 70\% of the gland volume. The epithelium within the PZ secretes prostatic fluid, which contains proteins, such as prostatic acid phosphatase and PSA, and metabolites, such as citrate and polyamines (e.g. spermine)\textsuperscript{172}. Furthermore, prostatic cells have been shown to be present in EPS and ejaculate, which makes both biofluids suitable media for molecular analysis\textsuperscript{14}.

Citrate production, after sequestration, by PZ epithelium results in a higher citrate concentration in EPS when compared with blood plasma\textsuperscript{172}. This process is facilitated by zinc-dependent truncation of the tricarboxylic acid cycle by inhibiting the enzyme m-aconitase, as shown in Figure 2.8.

ZIP1 is the primary transporter for zinc ions in PZ epithelium, and is expressed by the \textit{ZIP1} gene, which has consequently been described as a tumour suppressor gene in CaP\textsuperscript{173}. The expression of ZIP1 and other zinc transporters recently has been described as being regulated by the micro-RNA cluster miR-183-96-182\textsuperscript{174}. Zinc ions inhibit m-aconitase, which converts citrate to isocitrate, the first step of the citric acid cycle. As a result, the
Figure 2.8: Pertinent physiology of the healthy PZ epithelium. Biochemical reactions are shown by solid/open arrows and regulatory interactions by dashed arrows. When healthy, ZIP1 mediated uptake of zinc inhibits isomerization of citrate to isocitrate by m-aconitase. The result is high intracellular concentrations of zinc ions and citrate, which are secreted to aid in fertilization. Adapted from\textsuperscript{20,172}.

preferential sequestration of zinc ions in the PZ epithelium causes citric acid cycle truncation, producing an increased glucose requirement within the PZ epithelium.

The resulting high citrate and Zn\textsuperscript{2+} concentrations in PZ epithelium are reflected in EPS. Citrate is important in ejaculate ion homeostasis, and is the predominant regulator of calcium ions, which are important in the motility, metabolism and fertilization functions of sperm\textsuperscript{175}. Levels of zinc ions are correlated with those of other cations, such as calcium and magnesium, but are considerably higher in concentration. In semen, these cations are largely redistributed in binding to negatively charged seminal vesicle proteins, such as seminogelins, which are vital in regulating sperm function\textsuperscript{176}. In ejaculate, zinc ions are bound mostly to metallothionein, with changes in levels of zinc being paralleled by those of this protein which is mostly derived from the prostate itself\textsuperscript{177}.

2.2.3.1.2 The Malignant Prostate

Malignant transformation of cells is the result of irreversible genetic alterations, most commonly due to mutations. Specific to CaP, malignant transformation impairs Zn\textsuperscript{2+} accumulation, removing zinc-mediated inhibition of m-aconitase. The result is completion of the citric acid cycle and increased ATP production via oxidative phosphorylation. This is
reflected by low zinc and citrate concentrations present in PZ epithelium and prostatic fluid, which have been investigated as potential biomarkers\textsuperscript{172}. Further alterations in gene expression impair normal mitochondrial functioning. Coupled with the relatively rapid division and increased basal metabolic rate in cancer cells, increased glycolysis and lactate fermentation in the presence of oxygen occurring in the malignant state increases glucose uptake, as well as proteolysis and subsequent alanine production. Pyruvate is produced in excess of what can be processed by the tricarboxylic acid cycle, and is converted to lactate. This is known as the Warburg effect, and is seen as a marker of advanced disease in prostate and other cancers\textsuperscript{178,179}. This process is outlined in Figure 2.9.

![Figure 2.9: Pathophysiology of the PZ epithelium after malignant transformation. Impaired zinc uptake reduces inhibition of m-aconitase, resulting in citrate isomerization and completion of the TCA cycle.Alanine is produced secondary to proteolysis and lactate as a consequence of the Warburg effect. Adapted from\textsuperscript{20,172,180}.](image)

Furthermore, increased membraneogenesis accompanying increased cellular proliferation adds to the changes in the metabolite profile with malignant transformation, and requires synthesis of choline and creatine, which have been shown to be elevated in malignant prostate tissues\textsuperscript{181}.

### 2.2.3.2 Individual Biomarkers

A single biomarker that is able to confirm the presence of an altered biological process or indicates progression of a disease is a valuable asset in prompting appropriate management for any medical condition to improve the outcome for a particular patient. For instance, extremely high serum concentrations of the human hormone β-chorionic gonadotropin (β-hCG; a marker normally used in pregnancy) in a male patient with a small
testicular mass strongly indicate the presence of choriocarcinoma. While this is an example of an ideal biomarker, such biomarkers do not currently exist for most scenarios in oncology, particularly in CaP\textsuperscript{182}.

2.2.3.2.1 Serum

The most widely used biomarker for CaP is serum human kallikrein 3, also known as PSA. PSA is a serine protease that is normally secreted in ejaculate to catalyse proteolysis of seminal proteins, such as seminogelin\textsuperscript{183}. PSA is elevated in blood in the presence of CaP as well as with other prostatic conditions, such as bacterial prostatitis and benign prostatic hyperplasia (BPH). Despite not being specific for cancer, PSA is clinically valued and widely used\textsuperscript{21}.

The normal range of serum PSA, based on population studies, is defined as $<1.0$ ng/ml. Serum PSA is not only frequently elevated in other disease states, but can also change in the absence of pathology due to other confounders, such as racial and environmental variables\textsuperscript{184-186}. In biomarker research, often a cut-off point is derived from studies to determine the optimal sensitivity (i.e. the ability of the test to accurately predict true negative patients) and specificity (i.e. the ability of the test to predict accurate true positive patients). A vast body of evidence has shown that a safe cut-off value for PSA does not exist\textsuperscript{187,188}. Safety in this context refers to a level that is low enough to detect the majority of men with cancer, but not so low as to cause extensive and unnecessary investigation of men without cancer. This finding of the absence of a safe cut-off value for PSA has been used as one of the major arguments against population screening for CaP using PSA\textsuperscript{8,189}.

Serum PSA is a clinically valued test when used with discrimination, and various adaptations have been discovered and trialled with varying success, though none has been considered superior to total serum PSA itself\textsuperscript{190,191}. Examples include free to total PSA levels, PSA velocity and doubling time (time course of an increase by a factor of two), PSA density (serum PSA in relation to the prostate volume determined during TRUSBx) and, most recently, the prostate health index which incorporates serum PSA, pro-PSA and percentage free PSA\textsuperscript{192-194}.

Other serum biomarkers in CaP diagnosis vary in type and size, from circulating tumour cells\textsuperscript{195,196}, to microRNAs\textsuperscript{197,198}, with small molecules and ions, such as sarcosine and zinc, having yielded inconsistent results as markers\textsuperscript{199-202}.
2.2.3.2.2 Urine

Ideally, the perfect marker of CaP is sourced from a non-invasive sample/procedure and indicates both the presence and nature of the disease. Currently, the best urinary marker for CaP is PCA3, formerly known as differential display clone 3 [203]. The PCA3 test relies on a patient having had a firm digital rectal examination or prostatic massage just before micturition with the flow of urine flushing dislodged prostatic cells in the prostatic urethra to beyond the external meatus with the void for collection, so there is some licence involved in calling this a urine test. PCA3 is a non-coding RNA which has been shown to be highly expressed in and specific for prostatic tissue [204-206]. PCA3 in urine is expressed as a ratio to PSA RNA, and improves detection compared with serum PSA. Use of recently described PCA3 isoforms may further improve results [207-209]. PCA3 also contributes to and has been recommended for clinical decision making for men with previous negative biopsies but in whom clinical suspicion is high [210], however its role in CaP detection has yet to be established clinically. Inclusion of the TMPRSS2: ERG fusion gene also has been reported to improve detection of CaP with reference to biopsy [211].

2.2.3.2.3 Ejaculate

The main concern with tests using EPS is that both firm DRE / prostatic massage and TRUSBx target the posterior part of the prostate, and neglect anterior and anterolateral aspects of the gland in which up to 30% of tumours are sited [212,213]. In contrast, ejaculate contains a prostatic component, which is the result of global smooth muscle contraction, and thus theoretically reflects the pathological status of the whole gland. Furthermore, the ability to produce ejaculate among donors has been shown to correlate with improved survival compared to age-matched controls with CaP [17], hypothetically due to adequate erectile function and favourable cardiovascular status [18,214]. Thus, men who produce ejaculate are expected to have a more favourable mortality outcome following intervention with curative intent for CaP than men who are impotent since cardiovascular disease is the commonest cause of patient demise in this population. Prostatic tissue and prostatic fluid show similar levels of citrate and Zn\(^{2+}\), further suggesting that prostatic fluid reflects intraprostatic pathophysiological status [215].

2.2.3.2.4 Metabolite Changes in Ejaculate

Historically, changes in citrate and Zn\(^{2+}\) in CaP have been the most pronounced and easily detectable in prostatic and ejaculate [216]. Metabolite profiling and recent metabolomic
analysis of ejaculate and EPS have discovered alterations in other metabolites, summarized in Table 2.4. Disturbed zinc homeostasis removes the inhibition of m-aconitase, resulting in citrate oxidation in the citric acid cycle. This causes luminal Zn$^{2+}$ and citrate depletion. Zinc depletion only occurs in CaP, and has been shown to be a stable indicator of CaP status and progression$^{215}$.

Although citrate levels are altered in other pathophysiological states, such as BPH and prostatitis$^{217}$, reduced citrate concentrations in histologically benign prostatic tissue is considered to precede microscopic evidence of CaP$^72$. In poorly differentiated tumours, these normally abundant metabolites are present in very low concentrations$^{218}$. This metabolite relationship in CaP has also been correlated with the Gleason histological scoring system and is more accurate than serum PSA$^{22}$. Such biochemical changes reflect early neoplastic processes that may not be histologically identifiable, a concept familiar in oncology as the “field effect”$^{172}$. This further supports the role of metabolomics in identifying significant metabolic alterations in pre-malignant tissue.

Other metabolite changes seen in oncology that are not prostate-specific are also present in ejaculate. Disturbed synthesis and intracellular depletion of polyamines, such as spermine, are reflected in prostatic fluid$^{22,23,219,220}$. The prostate contains the highest levels of spermine in the body, and disturbances in ornithine-decarboxylase in polyamine metabolism have been a hypothesized mechanism for spermine depletion$^{220,221,222}$. A role of increased reactive oxygen species production by increased expression of spermine oxidase in CaP has linked inflammation with CaP carcinogenesis$^{223}$. Levels of myo-inositol, a molecule involved in membrane biosynthesis, have also been shown to be reduced in prostatic fluid$^{23,64}$. Some changes in metabolite levels in prostatic tissue are not reflected in prostatic fluid. Choline is upregulated in CaP tissue, both in vitro and in vivo, being hypothesized as another metabolite involved in membrane biosynthesis. The use of choline as a marker in prostatic fluid is compromised by the endogenous conversion of phosphocholine (from the seminal vesicles) to choline catalysed by prostatic acid phosphatase (from the prostate) shortly following ejaculation. This produces a biological artefact in choline concentration. Lactate and alanine are also increased in CaP tissue as part of the Warburg effect. However, spermatozoa utilize fructose from the seminal vesicles and glucose via glycolysis to produce ATP to fuel flagellal movement in utero to aid fertilization, resulting in varying levels of lactate and alanine as metabolic by-products. This illustrates a confounding factor between external cellular components and intraprostatic metabolites.
<table>
<thead>
<tr>
<th>CaP-induced change</th>
<th>Metabolite</th>
<th>Role (normal)</th>
<th>Alteration hypothesis</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increase</td>
<td>Choline</td>
<td>Membrane phospholipid precursor</td>
<td>Increased membraneogenesis</td>
<td>224-227</td>
</tr>
<tr>
<td></td>
<td>Lactate</td>
<td>End product of anaerobic glycolysis</td>
<td>Warburg effect</td>
<td>224, 227,228</td>
</tr>
<tr>
<td></td>
<td>Alanine</td>
<td>End product of anaerobic glycolysis</td>
<td>Warburg effect</td>
<td>224, 227,228</td>
</tr>
<tr>
<td></td>
<td>Omega-6 fatty acids</td>
<td>Cell membrane biosynthesis, fatty acid oxidation</td>
<td>Altered gene Expression</td>
<td>229</td>
</tr>
<tr>
<td></td>
<td>Cholesterol</td>
<td>Membrane biosynthesis, androgen regulated</td>
<td>Increased cell Turnover</td>
<td>230</td>
</tr>
<tr>
<td></td>
<td>Sarcosine</td>
<td>Glycine metabolism, purine synthesis</td>
<td>Cell invasion</td>
<td>231</td>
</tr>
<tr>
<td></td>
<td>(Choline + creatine) / citrate</td>
<td>Metabolite ratio</td>
<td>Increased ratio</td>
<td>232,233</td>
</tr>
<tr>
<td></td>
<td>Choline / citrate</td>
<td>Metabolite ratio</td>
<td>Increased ratio</td>
<td>233</td>
</tr>
<tr>
<td></td>
<td>Choline / creatine</td>
<td>Metabolite ratio</td>
<td>Increased ratio</td>
<td>233</td>
</tr>
<tr>
<td>Decrease</td>
<td>Citrate</td>
<td>Ion homeostasis, pH buffer</td>
<td>m-aconitase activation</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>Spermine</td>
<td>Polyamine synthesis</td>
<td>Oxidative stress, enzyme alteration</td>
<td>23,220</td>
</tr>
<tr>
<td></td>
<td>Myo-inositol</td>
<td>Membrane biosynthesis</td>
<td></td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>Citrate / spermine</td>
<td>Metabolite ratio</td>
<td>Decreased ratio</td>
<td>64</td>
</tr>
<tr>
<td></td>
<td>Citrate / creatine</td>
<td>Metabolite ratio</td>
<td>Decreased ratio</td>
<td>233</td>
</tr>
</tbody>
</table>

Table 2.4: Summary of metabolite changes in prostate cancer. Changes generic to cancer, such as lactate and alanine, are listed together with changes specific to prostate physiology, such as citrate, sarcosine and spermine. Adapted from 20.

2.2.3.3 Metabolomics in Prostate Cancer Diagnosis: Finding the Best Combination

Metabolite concentrations and ratios have aided in distinguishing CaP from benign prostates\textsuperscript{225,226,228,233}. Yet, despite promising preliminary results, there is currently no test available that is accepted as an accurate, stand-alone diagnostic or screening test. With improved data acquisition and processing technology, the concept of using entire
metabolic profiles as a large-scale combination of biomarkers has become feasible. Furthermore, metabolite profiles have been shown to be more sensitive as predictors of CaP, and in predicting metastatic potential\textsuperscript{234-237}. This concept has been demonstrated by metabolomic imaging, in which multivoxel MR spectra of intact prostates were analysed with MVSA. This was able to detect highly significant changes between the global metabolite profiles of benign and malignant prostate tissue without the need to identify specific metabolites\textsuperscript{237}. Similar relationships were demonstrated using freshly frozen CaP tissue when microarray gene expression data and metabolomic data were combined using PLS, providing further insight into mechanisms of metabolite alterations in CaP\textsuperscript{103}. In another study, metabolomic profiling provided an accurate prediction of biochemical recurrence of CaP, that is a rise in serum PSA, following intervention\textsuperscript{75}. This illustrates the potential of metabolomics as a suitable method for monitoring CaP behaviour following clinical interventions\textsuperscript{20}.

The concept of multiple markers in cancer diagnosis has been examined for some time. Specific to CaP, panels of molecular and protein-based markers have been used to improve serum PSA-based CaP detection\textsuperscript{238-240}. The most widely publicized and promising appear to be the combination of PCA3 and TMPRSS2-ERG fusion transcripts in post-massage urine as previously described. Multiple studies have shown improved sensitivity when combining PCA3 and TMPRSS2-ERG compared with PCA3 or serum PSA alone\textsuperscript{208,211}. Multiple mRNA markers (GalNAc-T3, prostate specific membrane antigen; PSMA, Hepsin and PCA3) in malignant prostate tissue have been able to provide optimal detection rates\textsuperscript{21,206,241}. Other researchers have attempted to combine single markers of different origins to improve CaP diagnosis. For example, a multiplex model utilizing gene-, protein- and metabolite-based targets for CaP outperformed any single biomarker\textsuperscript{238}. However, although these studies are promising in improving CaP diagnosis, many are impractical for use in a clinical setting, mostly due to financial and logistical constraints. Thus, in addition to improved accuracy, a further potential benefit of using metabolomics in CaP diagnosis is a reduction in cost and logistical requirements for each sample, although a high initial capital equipment financial outlay is required. Further limitations of this approach are outlined below.

### 2.2.3.4 Limiting/Confounding Factors

As previously discussed, detecting early CaP non-invasively is difficult. This is due to many confounding factors relating to the pathophysiology of CaP, but also due to
concomitant prostatic disease mimicking CaP. In the majority of studies to date, the greatest interference arises from concomitant pathophysiology of the prostate, such as BPH and prostatitis. Serum PSA is known to be elevated in BPH: androgens contributing to BPH development drive PSA synthesis which is mirrored in serum levels. Prostatitis associated with inflammation and prostatic cell lysis, results in increased release of intracellular PSA into the bloodstream, elevating serum PSA. It is for these reasons that serum PSA lacks sensitivity (detecting many false positives) and that there is sustained criticism directed toward serum PSA testing.

Tissue and prostatic fluid levels of citrate were reported to be initially promising in CaP diagnosis compared with serum PSA, but were observed in the past to be depleted in prostatitis. Notwithstanding similar potential dilemmas as those experienced with serum PSA, diminished levels of both zinc and citrate in these samples may provide improved sensitivity in CaP detection, although conclusive evidence in conjunction with PSA elevations is lacking to date. The relationship of citrate depletion in tissues following radiotherapy or hormonal therapy also relates to biochemical recurrence, defined as a rising serum PSA following intervention with curative intent. Furthermore, reduced specificity of serum PSA when compared with metabolite diagnosis may be due to pre-malignant disturbances in metabolic homeostasis that are not histologically visible. The underlying issue of biochemical characterisation preceding histopathology creates ongoing uncertainty, as tissue histopathology is the current gold standard for CaP diagnosis.

Intra- and extracellular citrate levels are also known to increase in BPH, thus citrate estimation in biopsy tissue may be unreliable. In these circumstances, prostatic fluid may be a more appropriate sample to use since prostatic fluid is produced mostly in the peripheral zone, which is also where most CaP is located, whereas BPH develops in the transition (central) zone of the prostate. A small proportion of transition zone tumours may be missed, but in the large majority of cases these are less aggressive and therefore of less significance.

2.2.4 Future Directions

2.2.4.1 Pharmacometabolomics and Theranosis: Towards Personalized Medicine

Pharmacometabolomics seeks to predict the metabolic response to exogenous therapeutic agents prior to or during drug administration, and theranosis is the identification and
monitoring of optimal treatments for patients as guided by diagnostic tests\textsuperscript{246-248}. Both are important aspects, given the recently emerging evidence of inter-individual differences in drug pharmacokinetics, being the ability of an individual to absorb, distribute, metabolize and excrete an administered drug\textsuperscript{100}. The result is reduced therapeutic efficacy, but may also be responsible for toxicity and adverse drug effects. The aetiology of these differences in drug metabolism is diverse and is well understood in only a limited number of circumstances (e. g. Cytochrome p450 enzyme family\textsuperscript{249}). Gender, age, race and concomitant diseases have been suggested as inherent factors, but have yet to be substantiated\textsuperscript{250}. External factors such as dietary and lifestyle habits, as well as toxin exposure, may also have a large influence on therapeutic efficacy. Furthermore, less obvious but important factors may contribute, such as altered gut flora in various circumstances\textsuperscript{251-253}. As such, metabolic phenotypes are diverse and complex due to these many influencing factors\textsuperscript{254}. Although genetic profiling across different disorders is important, metabolite profiling promises to better reflect the phenotype of disease states, and advanced analysis between both methods may help to identify genes with significant penetrance.

As has been illustrated, the carcinogenic changes in various cancers will cause common changes to individual metabolic profiles that can be investigated with metabolomics. In contrast, each individual patient will exhibit inherently different metabolic profiles, while also responding differently to therapeutic interventions\textsuperscript{255}. Thus, the concept of personalized medicine, where treatments are tailored to an individual’s personal metabolic or genetic phenotype, is one that is exciting, and important in advancing medical treatments\textsuperscript{256}. Using metabolite profiles as a representation of metabolic phenotype promises to enable theranosis by providing the most useful information for predicting inter-individual variation that will guide and assess efficacy of treatment outcomes.

Early research focused on genetic predisposition to cancer or alterations in drug-metabolizing enzymes\textsuperscript{257}. Recent research has focused on identifying varying metabolic profiles that indicate significantly affected drug metabolism, with links to altered gut flora homeostasis\textsuperscript{246,258}. This research related individual background urinary metabolic phenotype to biological and therapeutic outcomes of drug metabolism. Other research has identified alterations in metabolite profiles to illustrate pharmacokinetics and early toxicity, important in preventing adverse outcomes from drug toxicity\textsuperscript{259}. Recently, this approach has been applied to surgery, and allows for personalized pre-, intra- and post-operative care to improve patient outcomes\textsuperscript{98,260}. The concept of pharmacometabolomics can also
be applied to outcome prediction, similar to that used in GWAS, with evidence of serum metabolite levels to be predictive of body mass following chemotherapy for breast cancer\textsuperscript{261}.

Theoretically, pharmacometabolomics has advantages over pharmacogenomics in representing the phenotype resulting from multiple genetic effects. However, it is believed that a combination of approaches will provide best prediction and outcomes\textsuperscript{262}. Given current variations in efficacy, toxicity and adverse outcomes of treatments, developing personalized medicine is imperative to provide better medical care to patients while also reducing health budget costs. Thus, the pharmacometabolomic approach is one that has potential to change the therapeutic landscape not only in oncology, but across all fields of medicine.

2.2.4.2 Metabolomics to Elucidate Biological Mechanisms

As outlined, metabolomics has been useful in displaying changes in metabolites in various healthy and pathological states. The analysis of metabolites illustrates the end product of normally functioning or disturbed cellular processes and mechanisms. Thus, analysis of changes in metabolite profiles can lead to insights about the underlying biochemical or biological mechanisms. This has been demonstrated in different areas, including but not limited to, drug toxicity, cancer and plant studies\textsuperscript{99,164,263}.

For example, metabolomics could explain how altered STAT5 signalling as a result of truncated intracellular domains of growth hormone receptor in liver tissue leads to late-onset obesity, as systemic metabolite changes were consistent with globally altered metabolism contributing to obesity\textsuperscript{264}. A similar approach was used for data obtained from a human prostate cancer xenograft model in mice measured by NMR-based metabonomics and proteomics (two dimensional difference gel electrophoresis). Pathway analysis was used to link altered protein expression to changes in amino acids, which contributed to the metabolic phenotype\textsuperscript{99}.

2.2.4.3 Integration with other –omics

As previously outlined, the –omics approach to sample analysis provides data sets that require complex statistical analysis to extract meaningful information. In isolation, each –omics field provides insight into that particular level of cell function, and interactions and influences causing the results are hypothesized based on previous research or logical thinking. Thus, appropriate integration of –omics datasets has become an important step
in providing meaningful information in systems biology\textsuperscript{265,266}. This approach was e.g. used in insulin resistant mice using NMR-based metabonomics and genomics (quantitative trait locus mapping), and showed altered gut metabolites that were linked with genomic alterations\textsuperscript{267}. In both studies, large datasets were used to determine which metabolites were similarly affected by alterations in precursor compounds.

A suggested method that uses O2PLS for integration of large datasets for optimal information recovery is outlined in Figure \textsuperscript{2.10}\textsuperscript{266}, using the example of a study that has data from transcriptomics, proteomics and metabolomics experiments. As O2PLS is able to extract information from two datasets at a time, the method for correlating multiple datasets is naturally a multi-step procedure. In the first step the joint variation between two of the three datasets (e.g. transcript and metabolite data) is extracted by O2PLS. This means using one of the datasets as $X$ matrix in O2PLS and the other one as $Y$ matrix. Note that in this case, O2PLS is effectively run as \textit{unsupervised} analysis, because only two experimental datasets are correlated against each other, without including a set of external metadata. In the second step, the joint variation between transcript and metabolite data obtained in step 1 is then correlated with the third dataset (proteomics) in a second O2PLS, which will yield the variation common to all three datasets. In the final step, the joint variation that is common to all three datasets is deflated from the original datasets in a series of three parallel O2PLS analyses, to produce variation specific to each dataset.

\textbf{Figure 2.10:} Graphic representation of stepwise data integration of multiple –omics datasets with O2PLS. In the first step, joint variation between two –omics datasets (e.g. transcript and metabolite data) is determined. Using O2PLS in a second step, this joint variation is then correlated with the third –omics dataset (e.g. protein data) to determine variation that is common to all three sets of data. The third step removes the joint variation between all three individual datasets to produce variation that is specific to each dataset. This dataset-specific variation may be important in helping to address the biological question. Adapted from \textsuperscript{266}. 


It is trivial to extend this scheme to more than three datasets by adding on further O2PLS steps between steps 2 and 3 that each time introduce a further set of experimental data into the analysis. It should also be noted that it is prudent to repeat the first two steps with different orders of combining the three datasets – e.g. transcript and metabolite data first, then including protein data, versus metabolite and protein data first, then including transcript data, etc. – in order to rule out that potential slight imperfections in the symmetry behaviour of O2PLS might cause secondary effects on the data analysis.

### 2.2.4.4 Use of Computational Modelling

As is widely highlighted, the current approach to metabolomics including valid statistical analysis, metabolite identification, and biological interpretation is highly time-consuming. As such, the quest to develop computerized methods of metabolite analysis and identification is underway\textsuperscript{268,269}. Following metabolite identification, the next step is to determine the relationship and similarities, if any, of the identified metabolites to metabolic pathways, of which some preliminary programming applications have been released to address this issue\textsuperscript{113,270,271}.

Even more promising is the development of genome-scale computer models of metabolic networks. Extensive work has been completed on bacteria such as \textit{Escherichia coli}, with \textit{in silico} simulation reported to mimic experimental changes\textsuperscript{272}. Application of these reconstructed networks is more complicated in eukaryotes, such as human cells, due to the complex cellular and organismic organisation, including intra- and extra-cellular regulation and interactions. Despite these, an initial model human cell was constructed to provide a general baseline in expression and response to biological variables\textsuperscript{108,109}. Furthermore, a reconstruction of healthy liver cells was combined with whole-body pharmacokinetics to investigate multiple levels in biological organisation and provide mechanistic insights into for various drug-induced scenarios\textsuperscript{273}. Alterations to such models to accurately reflect cancer and other pathophysiological states by incorporating known and emerging evidence will better describe the response of these cells\textsuperscript{274}. Depending on the type of model, spatiotemporal processes and interactions within cells that may be undefinable or difficult to quantify are currently difficult to incorporate and apply to an artificial model\textsuperscript{275}. Further development of these reconstructed networks may occur via integration of –omics data sets, and research in this field is continuing\textsuperscript{276}. Computational modelling by incorporating multiple data sets represents a logical and informative, yet challenging, approach to oncology research to guide pharmaceutical development.
strategies. The result will be better informed treatment approaches and improved
treatment outcomes for these patients.

2.2.5 Conclusions

Metabolomics is a novel, modern and robust scientific approach that has shown great advances across many fields in biomedical research. The application of metabolomics to differing fields in medical science, including pathophysiology insight, drug development and *in vivo* imaging make it unique from all other approaches. Further research and collaboration to develop reconstructed networks, via integration of many terabytes of –omics data, is the next frontier in providing valuable insights to advance medical research and treatments in various human disease states.
2.3 Recent developments in prostate cancer metabolomics

Contemporary CaP metabolomics studies have expanded from biomarker discovery to elucidation of tumour aetiology and therapeutic targets. While most mechanistic investigations have occurred using malignant and benign human tissues, as well as cell lines, systemic metabolite disturbances detected in serum have furthered biomarker discovery as well as provided insight into metastatic mechanisms.

Localised metabolite changes

A recent key advancement has shown metabolic aberrations to be most pronounced in early malignant transformation, a notion hypothesised for some time\textsuperscript{20}. Giskeødegård and colleagues used HR-MAS \textsuperscript{1}H NMR spectroscopy to analyse 158 prostate tissue samples from 48 patients\textsuperscript{277}. CaP diagnosis was accurate (sensitivity 86.9\% and specificity 85.2\%) due to reduced spermine and citrate and an increased metabolite combination ratio (total choline+creatine+polyamines)/citrate) differentiation of low from high Gleason grade tumours and were correlated to progressing Gleason score. Other significant metabolites included putrescine, phosphoethanolamine, lactate, alanine, glucose, succinate, glutamate, glutamine, glycine, isoleucine, leucine and valine (glycerophosphoethanolamine, ethanolamine, creatine, taurine, myo-inositol and scyllo-inositol were not significant). A metabolomics approach was also used to characterise advanced CaP using GC/LC-MS in RP tissues, McDunn and colleagues demonstrated that citrate and polyamines, among other amino acid, energetic and lipid metabolites, were altered in CaP\textsuperscript{278}. Aggressive prostate tumours were further subdivided by abundance profiles of metabolites including nicotinamide adenine dinucleotide (NAD) and kynurenine, while metabolites such as ADP, laurate and mannose were able to stratify between organ confinement and extracapsular extension. When added to multiparametric nomograms, metabolites improved prediction of organ confinement (area under the curve [AUC] from 0.53 to 0.62) and 5-year recurrence (AUC from 0.53 to 0.64). These studies further support the role of metabolomics in detecting CaP and determining tumour aggressiveness based on grade and extent of disease.

Two landmark studies from independent investigators used prostate tissue metabolomics with other methods (transcriptomics, immunohistochemistry, histopathology) to investigate the relationship of the TMPRSS2:ERG fusion gene, a clinically used biomarker of aggressive CaP with tumour metabolism\textsuperscript{279,280}. Using HR-MAS \textsuperscript{1}H NMR spectroscopy in 129 samples from 41 patients, Hansen and colleagues demonstrated that the presence of
ERG (36% of samples) was associated with lower citrate and spermine and reciprocally altered gene expression of enzymes involved in citrate and polyamine metabolism. Meller and colleagues found similar metabolite alterations (reduced citrate, cis-aconitate, spermine and putrescine) using GC-/LC-MS in matched (benign/CaP) samples from 106 patients undergoing RP. When CaP grade was considered in conjunction with ERG status, ERG-positive tumours displaying a lower Gleason grade demonstrated more pronounced differences in metabolites and their reciprocal enzymes, including reduced citrate, glycerophosphoethanolamine, putrescine and spermine among increased glutamine and glycine levels, of which citrate and spermine were only observed in higher Gleason grade tumours. Similar relationships were confirmed in vivo using MRSI against 21 samples among 9 patients, with citrate, choline and spermine levels shown to be altered according to ERG status. High grade tumours are associated with altered cholesterol metabolism, potentially improving energy storage. Aspartate, isoleucine, tyrosine and tryptophan were correlated with ERG status, while reduced maltotriose and gluconic acid were seen in ERG positive samples, a relationship also reported for earlier biochemical recurrence. Gene enrichment meta-analysis considering metabolic pathways related to ERG status demonstrated glutathione, polyamine and glycolysis as well as purine and pyrimidine metabolism to be deranged, which was more pronounced in lower Gleason grade tumours. These important studies linked well-described and clinically used genetic changes to metabolic aberrations to improve understanding of metabolic pathways in early CaP development and potentially demonstrate the initiation of tumorigenesis.

**Altered prostatic metabolic pathways**

Further insights into mechanisms of altered metabolic pathways and their consequences on tumour behaviour have been elucidated. Specifically, reduced cellular citrate levels are due to reduced aconitase 2 expression coupled with increased expression of lipogenic enzymes (acetyl-CoA carboxylase alpha and fatty acid synthase). These expression patterns suggest citrate mobilisation to the cytosol for de novo fatty acid synthesis, which may indicate an aggressive phenotype, given association with tumour progression and reduced survival. Concomitantly elevated ATP citrate lyase may also implicate carbohydrate metabolism in malignant transformation. Reduced cellular spermine in CaP is well described, however changes in spermidine are conflicting. Reduced spermidine may be due to increased spermidine synthase and spermidine N(1)-
acetyltransferase\textsuperscript{279}, while elevated spermidine and glutathione has been observed in CaP samples\textsuperscript{280}.

In matched prostate tissues analysed using LC-MS, Ren and colleagues demonstrated multiple altered metabolic pathways\textsuperscript{283}. Dysregulated cysteine and methionine metabolism manifested by elevated choline, S-adenosylhomoserine, S-adenosylmethionine and 5-methylthioadenosine with reciprocally altered methyl transferase gene expression mapped previously reported alterations in sarcosine, glycine and methionine. Upregulation of NAD metabolism, presumably related to oxidative stress, and elevated hexosamine biosynthesis pathway metabolites were also observed. The metabolites carnitine C4-OH, citicoline, choline, GPCho, pantothenic acid, NAD, and sphingosine were significantly increased on MVSA. Sphingosine was determined to be most diagnostic (AUC > 0.8), while also implicated in sphingolipid metabolism due to altered sphingosine-1-phosphate receptor (S1PR\textsubscript{1-5}) expression, specifically impaired S1PR\textsubscript{2} signalling and loss of tumor suppressor gene function. The mechanisms for metabolic changes described in these studies reinforce underlying rationales for current clinical practices, such as 11-choline PET imaging, while also describing potential future imaging and drug targets.

\textit{Hormonal and metastatic metabolite changes}

Metabolomics studies have also helped describe relationships between intracellular processes and tumour progression. Metastatic CaP cell lines demonstrate elevated phosphatidylcholines, phosphatidylethanolamines, glycerophosphoinositols and other metabolites when analysed with MS-based metabolomic and lipidomic profiling\textsuperscript{284}. Choline kinase alpha, involved in lipid metabolism, was upregulated in metastatic cells compared with benign and localised CaP cells, indicating de novo lipogenesis in aggressive, metastatic CaP. While both MYC and AKT1 oncogenes induce FASN expression in immortalized human prostate epithelial cells, dysregulated lipid metabolism with negative enrichment of glycolysis and expression of GLUT1 transporter was present with MYC oncogene overexpression\textsuperscript{285}. In contrast, AKT1 activation results in accumulated aerobic glycolysis metabolites. AKT1/MYC status was not associated with Gleason grade and pathologic staging, indicating difference between architectural and molecular or metabolic phenotypes and potential role of the baseline metabolic state in oncogene-mediated metabolic reprogramming.

Androgen regulation is a key factor in CaP pathogenesis and hallmark of treatment of advanced and metastatic CaP with ADT. Circulating testosterone promotes prostatic
growth and alters cell metabolism following activation of the androgen receptor (AR), of which a splice variant, AR-V7, has been implicated in poor response to ADT in clinical studies. Using an LNCaP model, Shafi and colleagues\textsuperscript{286} demonstrated that inducing AR increased citrate levels, in contrast to AR-V7 which reduced citrate and mirrored metabolic shifts observed in patients resistant to ADT (castrate resistance). Flux analyses suggested that reduced citrate is due to increased utilisation, while AR-V7 produced some citric acid cycle metabolites due to increased dependence on glutaminolysis and reductive carboxylation. In a clinical study, absolute metabolite, lipid and macromolecule concentrations in 23 RP samples including some after receiving ADT (degarelix) were analysed using HR-MAS \textsuperscript{1}H NMR spectroscopy\textsuperscript{24}. Lactate, alanine and total choline concentrations were significantly elevated in CaP tissue compared to BPH, however lactate and total choline were reduced in samples subject to ADT\textsuperscript{24}. While small sample size limited detection of other statistically significant compounds such as citrate, lipids and macromolecules, these are likely to be altered in the high risk, ADT-treated samples. Thus, metabolomics may provide an avenue for \textit{in vivo} characterisation of metastatic disease and monitoring of treatment response.

**Systemic biomarkers**

Analysis of serum in metastatic CaP has demonstrated serum biomarkers related to cholesterol metabolism (deoxycholic acid, glycochenodeoxycholate, docosapentaenoic acid) among others (l-tryptophan, arachidonic acid, deoxycytidine triphosphate, and pyridinoline) to delineate patients with CaP from healthy controls\textsuperscript{287}. When patients were monitored during ADT, these markers remained abnormal in patients who developed hormone resistance in 1 year, yet reverted toward normal levels in those without hormone resistance after 2 years. Similarly dysregulated lipid metabolism is commonly reported among studies investigating circulating markers of localised CaP in urine and serum. Mondul and colleagues showed circulating 1-stearoylglycerol was inversely associated with overall and aggressive CaP, potentially related to overexpression of monoacylglycerol lipase\textsuperscript{288}. Energy/citric acid cycle metabolites, glycerophospholipids and fatty acids have also been shown to be altered in sera collected 1-20 years prior to clinical CaP diagnosis, where serum PSA is significant between subjects\textsuperscript{289}. Metabolite changes in serum acylcarnitines, related to fatty acid synthesis have also been reported, as well as choline and amino acid metabolites to improve diagnostic accuracy\textsuperscript{290,291}. Serum metabolite biomarker panels improve sensitivity and specificity beyond 85\%\textsuperscript{291,292}, with glycine,
sarcosine, alanine, creatine, xanthine, hypoxanthine, and citrate delineated in very high risk CaP (median PSA 31 ± 37 ng/ml) from BPH\textsuperscript{291}.

Analysis of urine has reported to improve diagnosis (sensitivity 88.4%, specificity 93%) and showed many dysregulated metabolic pathways\textsuperscript{293}. Using MetaboAnalyst, upregulation of purine and pyrimidine metabolism as well as downregulation of multiple pathways (histidine, arginine, tyrosine, tryptophan, taurine, alanine, aspartate, glutamate) was observed\textsuperscript{293}. Other investigators have observed implicated deranged pathways including the urea and tricarboxylic acid cycle as well as amino acid and purine metabolism in urine samples\textsuperscript{294}.

While metabolomics analyses of systemic markers in serum and urine have yielded initially promising results, these appear to be conflicting between studies with no reliable validation of their performance described. As such, these continue to be refined until an acceptable, reproducible diagnostic performance is confirmed.

**Disease monitoring in vivo with MRI**

In contemporary clinical practice, multiparametric MRI (mpMRI) has emerged as an important adjunct to serum PSA for improved detection and localisation of CaP tumour foci as well as providing information on tumour extension and metastasis\textsuperscript{6}. While NMR-based tissue studies have demonstrated metabolite disturbances to discern CaP from BPH, \textit{in vivo} NMR as \textsuperscript{1}H MRSI is minimally used due to a reported lack of additional benefit at the expense of increased acquisition time compared with standard mpMRI\textsuperscript{295}. Commonly used MRI hardware and acquisition parameters result in suboptimal MRSI spectra for use in metabolite quantification and ratio calculation. High quality shimming and adjustment for field inhomogeneity\textsuperscript{296} help improve spectral acquisition, while peak overlap between choline, creatine and spermine complicates reproducible quantitation\textsuperscript{297}.

Indeed, Hansen and colleagues corroborated \textit{in vitro} \textsuperscript{1}H HR-MAS NMR findings with \textit{in vivo} \textsuperscript{1}H MRSI, suggesting that MRSI may provide an avenue to stratify patients prior to biopsy or monitor metabolite profiles \textit{in vivo} in active surveillance programs, potentially in place of repeat biopsy\textsuperscript{279}. Furthermore, metabolite profile changes due to ADT could be monitored using \textsuperscript{1}H MRSI, shown \textit{in vitro} using \textsuperscript{1}H HR-MAS NMR\textsuperscript{24}. Multiple clinical studies have combined MRSI with mpMRI to improve diagnostic accuracy and correlation with Gleason score\textsuperscript{298}. MRSI has also been shown to predict extracapsular extension and post-RP treatment failure based on number of voxels with undetectable polyamines\textsuperscript{299}.
Furthermore, voxel resolution is reported to improve tumour localisation compared to expert observer image interpretation\textsuperscript{300}. Increasing utilisation of higher field strength magnets and ongoing refinements in acquisition techniques hopes to reduce signal-to-noise ratio and lipid artefacts to improve metabolic data acquisition and use in clinical studies\textsuperscript{301}. Improved metabolic data acquisition may allow easier incorporation of MRSI into mpMRI protocols to further describe biochemical characteristics of the imaged tumours to guide treatment recommendations or monitoring during active surveillance. These data could also be incorporated into \textit{in vitro} metabolomics biofluid and biopsy diagnostics studies, as well as be correlated with surgical pathology specimens using non-destructive, NMR-based methods.

\textbf{Current status and opportunities}

A major limitation in CaP metabolomics studies to date has been inconsistent reproducibility and validation. As discussed in Section 2.2.3.4, underlying prostatic biology related to BPH can cause localised metabolite changes similar to those seen in CaP, while many serum and urine studies have mostly recruited groups of men with disparate pathology unlike the populations seen in clinical practice. Table 2.5 contains a summary of validated, unverified, putative and unreliable metabolites described in CaP studies.

While verified metabolites are those with consistently reported changes, such as reduced citrate or increased choline and lactate, some metabolites that have been reported by separate studies require further verification, such as sarcosine. Putative markers that appear promising but require further replication include 2-hydroxybehenic acid and sphingosine. Inconsistently reported markers that are currently unreliable include myo-inositol and alanine, which for some, relate to the sample used (e.g. glycine increased in tissue and serum but decreased in urine). Despite reproducible relationships observed in prostate tissue studies, there is a grave lack of reliable prostate-derived biomarkers in prostate-specific biofluids, such as ejaculate or EPS, that can be used for early diagnosis of localised CaP. Further investigation is required to determine accurate biomarkers in prostate-specific biofluids, first confirming validated markers observed in tissue, while exploring unverified markers in need urgent replication and verification and others currently undescribed.
<table>
<thead>
<tr>
<th>Metabolite</th>
<th>Change</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Citrate</td>
<td>Decrease (EPS, tissue, urine)</td>
<td>23,277-280,294</td>
</tr>
<tr>
<td>Spermine</td>
<td>Decrease (EPS, tissue)</td>
<td>23,220,277-280</td>
</tr>
<tr>
<td>Choline</td>
<td>Increase (tissue in vitro/in vivo)</td>
<td>24,224-227,277-279,283</td>
</tr>
<tr>
<td>Lactate</td>
<td>Increase (tissue in vitro/in vivo)</td>
<td>24,224, 227,228,277</td>
</tr>
</tbody>
</table>

### Validated in multiple studies (unverified)

<table>
<thead>
<tr>
<th>Metabolite</th>
<th>Change</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sarcosine</td>
<td>Increase (tissue, serum)</td>
<td>231,280,291</td>
</tr>
<tr>
<td>Glutamate</td>
<td>Increase (tissue)</td>
<td>277,278,280</td>
</tr>
<tr>
<td>Putrescine</td>
<td>Decrease (tissue)</td>
<td>277-280</td>
</tr>
</tbody>
</table>

### Infrequently observed (putative, not replicated, unverified)

<table>
<thead>
<tr>
<th>Metabolite</th>
<th>Change</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-hydroxybehenic acid</td>
<td>Increase (tissue)</td>
<td>280,281</td>
</tr>
<tr>
<td>cis-Aconitate</td>
<td>Decrease (tissue)</td>
<td>280</td>
</tr>
<tr>
<td>Glucose</td>
<td>Decrease (Tissue)</td>
<td>279,280</td>
</tr>
<tr>
<td>Sphingosine</td>
<td>Increase (tissue)</td>
<td>283</td>
</tr>
</tbody>
</table>

### Unreliable based on multiple studies

<table>
<thead>
<tr>
<th>Metabolite</th>
<th>Change</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Myo-inositol</td>
<td>Decrease (EPS, tissue)</td>
<td>23,280</td>
</tr>
<tr>
<td></td>
<td>Increase (tissue)</td>
<td>277</td>
</tr>
<tr>
<td>Cholesterol / Fatty acids</td>
<td>Increase (tissue in vitro/in vivo)</td>
<td>229,230,280</td>
</tr>
<tr>
<td></td>
<td>Decrease (tissue in vitro)</td>
<td>280</td>
</tr>
<tr>
<td>Creatine</td>
<td>Unchanged (tissue)</td>
<td>277</td>
</tr>
<tr>
<td></td>
<td>Reduced (tissue)</td>
<td>280</td>
</tr>
<tr>
<td>Alanine</td>
<td>Increase (tissue in vitro/in vivo, serum)</td>
<td>24,224,227,228,277,278</td>
</tr>
<tr>
<td></td>
<td>Variable (serum)</td>
<td>291</td>
</tr>
<tr>
<td></td>
<td>Reduced (urine)</td>
<td>293</td>
</tr>
<tr>
<td>Glycine</td>
<td>Increase (tissue, serum)</td>
<td>279,280,291</td>
</tr>
<tr>
<td></td>
<td>Decrease (urine)</td>
<td>293</td>
</tr>
</tbody>
</table>

Table 2.5: Localised and systemic metabolite changes in prostate cancer based on metabolomics studies. Described metabolite changes are grouped according to whether they are validated, observed and require verification, infrequently observed and require replication or unreliable (inconsistent).
2.4 Conclusions on molecular and metabolomic biomarkers for non-invasive prostate cancer diagnosis and monitoring

Prostatic fluid-based biomarker discovery has improved understanding of metabolic changes during CaP initiation and progression. Accurate characterisation of underlying metabolic abnormalities in clinical populations using prostatic tissue \textit{in vitro} and MRI \textit{in vivo} provide an ideal target for a mechanistic approach to biomarker discovery. Given known genetic and phenotypic heterogeneity among CaP, it is likely that biomarker panels in combination with selective imaging may provide highest diagnostic accuracy. While prostatic fluid presents an ideal non-invasive biomarker platform, the feasibility and accuracy of a prostatic fluid-based biomarker panel is yet to be investigated within a “high risk”, clinical cohort.
Part 2 – Methods
Chapter 3 – Optimisation of prostatic fluid for metabolomic biomarker discovery

3.1 Synopsis

The aim of this chapter was to optimise biofluid sample collection and processing techniques for future clinical metabolomics studies. PAP activity results in variable ejaculate choline/phosphocholine levels which influence metabolomics analysis so the performance of tartrate as an inhibitor of PAP at different concentrations and temperatures and as an internal concentration standard was investigated to improve ejaculate metabolite stability and quantification. Subsequent effects of tartrate on established ejaculate processing methods for molecular (mRNA) analysis were also analysed. This investigation is presented as a peer-reviewed manuscript (section 3.2).
3.2 “Tartrate inhibition of prostatic acid phosphatase improves seminal fluid metabolite stability.” (published manuscript)


The contributions of the co-authors to this manuscript were as follows: Experiment design, sample collection, experiment execution, data processing and statistical analysis were performed by Matthew Roberts, the PhD candidate, under the supervision of Dr Horst Schirra and Prof RA “Frank” Gardiner. Jake Hattwell and Clement Chow assisted with sample collection, experiment execution, data processing and statistical analysis. Dr Gregory Pierens advised on experimental design and set-up, data acquisition and interpretation of results. Prof Martin Lavin provided critical manuscript revisions. Preparation of the manuscript and associated figures was performed by Matthew Roberts, the PhD candidate.

The manuscript text is presented as originally published by the journal. To maintain a consistent style between chapters in this thesis, abbreviations, placement of figures and tables within the text, as well as numbering of pages, figures and tables has been adjusted. The references have been collated in the section, ‘Bibliography’, with others contained in this thesis.
Tartrate inhibition of prostatic acid phosphatase improves seminal fluid metabolite stability.

Matthew J. Roberts 1,2,3,4, Jake P. N. Hattwell1, Clement W. K. Chow2, Martin Lavin2, Gregory K. Pierens1, Robert A. Gardiner 2,3,4,5, Horst Joachim Schirra 1 #

1 The University of Queensland, Centre for Advanced Imaging, Brisbane, Qld 4072, Australia
2 The University of Queensland, Centre for Clinical Research, Brisbane, Qld 4006, Australia
3 Department of Urology, Royal Brisbane and Women’s Hospital, Brisbane, Qld 4006, Australia
4 The University of Queensland, School of Medicine, Brisbane, Qld 4072, Australia
5 Edith Cowan University, Joondalup, Western Australia
3.2.1 Abstract

**Introduction**: Ejaculate has been suggested as a biofluid suitable to characterise male reproductive organ pathology with metabolomics. However, various enzymatic processes, including phosphorylcholine hydrolysis mediated by prostatic acid phosphatase, cause unwanted metabolite variation that may complicate metabolomic analysis of fresh ejaculate samples. The aim of this study was to investigate the effects of PAP inhibition with tartrate.

**Methods**: Using NMR spectroscopy, the kinetics of phosphorylcholine to choline hydrolysis was characterized in ejaculate samples from three subjects at different temperatures and tartrate concentrations. Principal components analysis was used to characterise the effects of tartrate and temperature on personal differences in metabolite profiles. Potential effects of tartrate on RNA quantification were also determined.

**Results**: Metabolite profiles and the kinetics of phosphorylcholine degradation are reproducible in independent samples from three ostensibly normal subjects. Increasing concentrations of tartrate and refrigerated sample storage (279 K) resulted in greatly reduced reaction rates as judged by apparent rate constants. Multivariate statistical analysis showed that personal differences in metabolite profiles are not overshadowed by tartrate addition, which stabilises phosphorylcholine and choline concentrations. The tartrate signal also served as an internal concentration standard in the samples, allowing the determination of absolute metabolite concentrations in ejaculate. Furthermore, the presence of tartrate did not affect RNA expression analysis by qPCR.

**Conclusion**: Based on these results we recommend as standard protocol for the collection of ejaculate samples, that 10 mM tartrate are added immediately to samples, followed by sample storage/handling at 277 K until clinical processing within 6 hours to remove/inactivate enzymes and isolate metabolite supernatant and other cellular fractions.

**Keywords**: Prostatic acid phosphatase, Nuclear Magnetic Resonance, enzyme inhibition, metabolomics, seminal fluid, prostate cancer
3.2.2 Introduction

Biomarkers are an essential component of modern medicine in aiding diagnosis and treatment of various medical conditions and are predominantly found in bodily fluids including blood, urine and ejaculate. Metabolite changes in ejaculate have been shown to reflect tissue changes\textsuperscript{172}, thus metabolite quantification may allow ejaculate to indicate male reproductive organ pathology. For example, citrate, myo-inositol and spermine have been suggested to assist with prostate cancer diagnosis\textsuperscript{20}, while metabolites associated with oxidative stress have been implicated in other cancers and male infertility\textsuperscript{65,302}.

However, investigation of metabolites or proteins as biomarkers in ejaculate has been limited in comparison with other biofluids, because metabolite profiles in ejaculate exhibit temporal variation as a result of various enzymatic processes. These include small peptide and amino acid changes due to seminogelin proteolysis catalyzed by PSA, a commonly used marker for prostate cancer\textsuperscript{183}. More important for metabolomics studies, hydrolysis of phosphorylcholine to choline increases inorganic phosphate production within minutes and is mediated by PAP, a previously used prostate cancer biomarker\textsuperscript{303,304}. Both of these enzymes enhance function and protection of spermatozoa via improved zinc shuttling and hyaluronidase action, among other mechanisms\textsuperscript{176,305,306}.

The rapid and dynamic changes in ejaculate metabolites over time may complicate reliable quantification, as any single measurement is likely to capture an indeterminable point within the hydrolysis reaction in progress. In practical terms, a varied and unavoidable time delay from sample production to processing and storage may cause unwanted variation in metabolites such as phosphorylcholine and choline, given their known rapid changes after sample collection, to result in analysis not reflecting underlying pathology. Choline-based metabolites have been implicated with prostate cancer \textit{in vivo}\textsuperscript{307}, and choline-based PET imaging is an emerging diagnostic aide clinically\textsuperscript{308}. However, accurate \textit{in vitro} characterization to help predict if these imaging modalities are clinically useful remains unexplored.

Current sample preparation approaches to improve metabolite stability for ejaculate metabolomics studies are varied, including incubation at 37°C\textsuperscript{304,309} and snap freezing with subsequent thawing at 4°C or 37°C\textsuperscript{23,62}, which may jeopardise analysis of other ejaculate components, such as RNA\textsuperscript{310}. 


Because of PAP’s rapid effect in hydrolysing phosphorylcholine, immediate PAP inhibition of ejaculate may permit better characterization of choline-based metabolites compared with other described sample preparation methods. Various compounds that are potentially suitable PAP inhibitors are known, including tartrate, vanadate and molybdate\textsuperscript{311,312}. Ions such as vanadate and molybdate are less attractive for NMR profiling because of paramagnetic line broadening leading to reduced peak signal-to-noise ratio and analytical sensitivity\textsuperscript{313}. In contrast, tartrate has been used previously as a competitive inhibitor of PAP in clinical assays\textsuperscript{312} and is present as a single NMR peak, largely removed in peak position from other metabolites. Furthermore, tartrate quantification during NMR spectroscopy may also serve as an internal concentration standard for metabolomics studies, and allow for determination of absolute metabolite concentrations in ejaculate.

Here we demonstrate the performance of tartrate as an inhibitor of PAP, as well as an internal concentration standard by examining ejaculate metabolite profile variation, at different tartrate concentrations and temperatures. Furthermore, we quantify ejaculate-based mRNA species \textit{in vitro} in the presence of tartrate to evaluate any negative effects of tartrate on established ejaculate processing methods for molecular (mRNA) analysis.

**3.2.3 Materials and Methods**

**3.2.3.1 – Experimental Design**

3.2.3.1.1 – \textit{Reproducibility of kinetic curves and influence of tartrate}

The reproducibility of kinetic curves in ejaculate were investigated using a two-factor design comprising the addition/absence of 1 mM tartrate (\textbf{Tar}) and temperature (298 K = room temperature; and 277 K = refrigerator temperature) as outlined in Figure 3.1a. Duplicate experiments were conducted on ejaculate samples collected on separate days from different subjects for each experimental condition. Samples were collected, as outlined in Section 2.1.3, and measured using NMR spectroscopy, as described in Section 2.2.
Figure 3.1: Experimental workflow of sample preparation and spectral acquisition using two-factor design. a: Reproducibility of ejaculate kinetic profiles was determined by measuring ejaculate kinetic profiles from three healthy subjects on two separate occasions each as well as with and without 1 mM tartrate (Tar). b: The influence of tartrate concentration and temperature on ejaculate kinetic profiles was investigated using cyclically interleaved measurement of ejaculate samples with added tartrate at different concentrations and at separate spectrometers (AV900/AV700) set to different temperatures (298 K; 279 K). The temperature change from 277 K to 279 K was required because the SampleJet refrigerating system, is set to a fixed temperature of 279 K.

3.2.3.1.2 – Kinetic curves with varied tartrate concentration and temperature

The efficacy of tartrate as a PAP inhibitor was investigated using a two-factor design with different temperatures (279 K and 298 K) and tartrate concentrations (0 mM, 1 mM, 3 mM, 5 mM, 10 mM, 20 mM). To save time and increase throughput, the experiments were interleaved by cycling samples and measurements at regular intervals, as outlined in Figure 3.1b. Furthermore, throughput was further increased by measuring the kinetic curves at these two different temperatures simultaneously using two different NMR spectrometers.

Serial aliquots from samples were collected and combined with L(+) – tartrate-phosphate buffered saline (PBS) buffer solutions to produce specific final tartrate concentrations within each sample solution. Aliquots were then measured, interleaved, and simultaneously at 298 K at an AV900 spectrometer and at 279 K at an AV700 MHz spectrometer. The lower temperature of 279 K was chosen, as that is the temperature at which samples are stored in a Sample Jet sample changer, and it is close enough to temperatures in a laboratory refrigerator (277 K). Thus, any changes seen at 279 K and any consequent recommendations would, if anything, be under-estimating the expected effects. For the first two subjects, measurements were taken in triplicate every 12 h after
the first 24 h in order to maintain sample temperature at 279 K for as large a fraction of
time as possible, so that measured kinetic profiles accurately reflected conditions at lower
temperatures over a prolonged period. For the third subject, measurements at 279 K were
taken at regular 30 minute intervals similar to that seen for 298 K, as each sample spent
the majority of time in storage at 279 K for each measurement cycle, with a mandatory
drying time of ~4-5 min at 293 K in the pre-heater before insertion into the magnet. The net
effect of this was an overall slightly faster kinetics than for subjects 1 and 2, which was
deemed to be inconsequential if sufficient inhibition was able to be demonstrated. The
tartrate concentrations were selected on the basis of prior reports of 70, 90 and 95% PAP
inhibition at tartrate concentrations of 1, 5 and 20 mM, respectively.314

3.2.3.1.3 – Subjects/samples

Ethical approval for this study was obtained from the University of Queensland Medical
Research Ethics Committee (Project no. 2006000262) and the Royal Brisbane and
Women’s Hospital (RBWH) Human Research Ethics Committee (HREC/09/QRBW/320,
HREC/09/QRBW/305 together with 1995/088B). The study was conducted in accordance
with the ethical standards outlined in the 1964 Declaration of Helsinki.

Following informed consent, three healthy Caucasian subjects of different ages [26, 29
and 45 (no familial history, PSA=0.96 µg/L)] were recruited and provided serial ejaculate
samples into sterile micro-urine jars containing 20 ml Phosphate Buffered Saline (PBS – in
house preparation). Samples were vigorously mixed with the buffer solution and 100 µl
aliquots were distributed among Eppendorf tubes containing 80 µl of sodium potassium
L(+)-tartrate (Sigma Chemical Company, Saint Louis, MO, USA) in PBS solution at
concentrations according to experimental design (see Figure 3.1). Samples were
prepared to a total volume of 200 µL with 20 µL D2O as lock substance that
contained DSS as internal chemical shift standard and 1,1-difluoro-1-trimethylsilyl
methylphosphonic acid (DFTMP) as internal pH indicator to yield a final concentration of
100 µM for both.

3.2.3.2 – NMR Spectroscopy

NMR spectroscopy of ejaculate samples was performed on Bruker Avance 900 (AV900)
and Avance 700 (AV700) spectrometers operating at 1H frequencies of 900.13 and 700.47
MHz, respectively, (Bruker Biospin, Rheinstetten, Germany), both equipped with a 5 mm
self-shielded z-gradient triple resonance cryoprobe and a SampleJet® sample changer,
chilled to 279 K. Spectra were measured at temperatures of 277 K, 279 K and 298 K according to experimental design (Figure 3.1).

One-dimensional (1D) nuclear Overhauser effect spectroscopy (NOESY) proton ($^1$H) spectra were acquired in high-throughput automation mode via the ICON-NMR interface of Bruker Biospin with the "noesypr1d" pulse sequence ((RD) – 90° - t1 – 90° – τm – 90° – acq, Bruker Biospin pulse program library) in “baseopt” digitization mode with 32 scans (following 8 dummy scans) at 32k resolution and a spectral width of 14 ppm$^{150}$. The transmitter frequency was set to the frequency of the water signal, and water suppression was achieved by continuous wave irradiation during both the NOESY mixing time ($$\tau_m$$) of 0.1 s and the relaxation delay of 3.0 s. Automatic tuning and matching was performed, as well as automatic shimming on the $z$ and $z^2$ shims during spectral acquisition.

Pseudo-2D spectra were acquired manually as 1D NOESY spectra increments acquired every 162.4 s (AV900) and 180 s (AV700) using similar parameters as above with an in-house adaptation of the pulse programs noesygppr1d (AV900) and noesypr1d (AV700).

3.2.3.3 - Spectral Processing

1D NMR spectra were processed with TopSpin Version 3.2 (Bruker BioSpin GmbH, Rheinstetten, Germany). The free induction decays (FIDs) were baseline corrected by a Gaussian function of 0.1 ppm filter width for post-acquisitional water deconvolution according to$^{315}$ and multiplied by an exponential window function with 0.1 Hz line broadening factor, before Fourier transformation to a size of 32,768 (AV900) and 65,536 (AV700) data points. The chemical shift of spectra was calibrated on the DSS peak at 0.0 ppm after exponential multiplication, Fourier transformation and manual phasing, followed by manual baseline correction.

3.2.3.4 – Non-linear curve fitting

For 1D NOESY spectra, peak integrals for phosphorylcholine (3.205 - 3.215 ppm) and choline (3.188 – 3.198 ppm) were calculated with user defined integrals and peak epsilon radius, to allow for variation in peak positions in F2 of 30 data points between experiments, and exported using Dynamics Center Version 2.1.8 (Bruker BioSpin GmbH, Rheinstetten, Germany). For experiments where triplicates were measured, the average integral was calculated.
The averaged integrals ($y$-var) were imported into SigmaPlot 13.0 (Systat Software, San Jose, CA), coupled with the time (in h) since first acquisition ($x$-var, extracted from the timestamps of creation for each FID file) for non-linear curve fitting.

For pseudo-2D-NOESY spectra, the visible region of 3.18 – 3.24 ppm was exported from TopSpin into MATLAB 2012b (The Mathworks Inc., Natick, Massachusetts, United States) and peak integrals for phosphorylcholine (3.205 - 3.21 ppm) and choline (3.188 – 3.198 ppm) were calculated. Time periods between integral peak values were 162.4 s (AV900) and 180 s (AV700).

To characterise the decay of phosphorylcholine in initial pseudo-2D NOESY and serial 1D NOESY spectra, integral values at time $t$ were fitted to the following exponential equation in SigmaPlot:

$$f(t) = a \cdot e^{(-b\cdot t)} + c$$

The resulting fit parameters are listed in Table A1.1. Where poor fits were obtained, as determined by a low $r^2$ value or large standard errors for fitted parameters, a two-parameter fit omitting $c$ was used. If that yielded a poor fit as well, the corresponding fit was denoted in Table A1.1e. Similarly, to characterise the kinetics of choline production, peak integral values at time $t$ in initial pseudo-2D NOESY and serial 1D NOESY spectra were fitted in SigmaPlot to the exponential equation

$$f(t) = a \cdot (1 - e^{(-b\cdot t)}) + c$$

and then displayed as intensity ratio ($f(t)/f(t=\infty)$, i.e. scaled by the fitted peak integral at $t=\infty$, the highest value of the curve. Each combination of volunteer and temperature was scaled using the curve with the highest final intensity at $t=\infty$ (usually the uninhibited kinetics). The rationale for this was that all samples were aliquots from the same ejaculate sample and should thus increase to the same final choline concentration. For most inhibited samples the reaction speed is so slow that this point is never reached. The resulting exponential fit parameters are also listed in Table A1.1.

3.2.3.5 – Multivariate analysis

1D NOESY spectra corresponding to samples containing 0 mM and 10 mM tartrate acquired at different temperatures and time points were automatically data-reduced with AMIX (version 3.6.6, Bruker Biospin, Rheinsetten, Germany) into spectral integral regions (“buckets”) of 0.001 ppm width over the range $\delta = 10.0 – 0.25$ ppm (with exclusion of the
water region from 5.08 – 4.52 ppm), and metabolite peaks were then aligned with the spectral alignment algorithm “icoshift”\textsuperscript{131}, first by initial automatic alignment based on the lactate doublet at 1.32 ppm, then by a combination of alignment on the choline singlet at 3.19 ppm and subsequent manual alignment of metabolite peaks corresponding to user-defined intervals representing approximate peak widths. The resulting dataset was manually data-reduced to buckets of 0.01 ppm width (in the range 9.996 – 0.256 ppm). Quantile normalization was performed in order to correct for improper influence of large signals across the dataset\textsuperscript{139}. A second data matrix was constructed in a similar way, but excluding the tartrate signal from δ = 4.4334 – 4.24 ppm after peak alignments, but before quantile normalisation. Both data matrices were imported into SIMCA P+ 12.0 (Umetrics, Umeå, Sweden) for MVSA.

Both metabolite data matrices (\(X\)-matrix) underwent Pareto-scaled Principal Components Analysis (PCA) as an unsupervised analysis\textsuperscript{151}. The multivariate models were optimized using cross-validation, with model quality assessed using the \(R^2\)X and \(Q^2\) figures of merit. \(R^2\)X measures the variance of the \(X\) variables, expressed as the sum of squares fraction for the selected component, to quantify “goodness of fit”. \(Q^2\) uses cross validation to quantify the predictive accuracy, or “goodness of prediction” of the model.

3.2.3.6 – Univariate statistical analysis

To demonstrate the usefulness of tartrate as internal concentration standard, absolute concentrations of significant metabolites in the PCA were calculated based on their signal intensity and corresponding proton ratios relative to that of tartrate. The equation used to determine the relative molar concentration of each metabolite was:

\[
[\text{metabolite}] = [\text{Tartrate}] \cdot \frac{\text{bucket\_intensity(metabolite signal)}}{\text{bucket\_intensity(tartrate signal)}} \cdot \frac{\text{num\_protons(tartrate signal)}}{\text{num\_protons(metabolite signal)}}
\]

with bucket intensities taken from the icoshifted but not normalized bucket table with 0.01 ppm bucket width. This formula was applied to each metabolite identified in the MVSA for each tartrate-containing spectrum in the dataset where the tartrate peak was excluded (\(n = 66\)). As the citrate signal exhibited sample dependent shifts, the full chemical shift range of the citrate signal of δ = 2.486 – 2.706 ppm was used to quantify citrate.

3.2.3.7 – Application to mRNA extraction and processing

Various groups are exploring the use of RNA (PSA, PCA3 and Hepsin) and microRNA markers in ejaculate as biomarkers for prostate cancer\textsuperscript{15,16}, while also investigating
metabolomic profiles in this context (using sodium azide [\( \text{NaN}_3 \)] as our preservative of choice). The influence of tartrate on ejaculate-derived mRNA expression and processing was investigated in vitro using LNCaP cell culture. Cells were grown to 90% confluence in T175 flasks with RPMI media (In-house) supplemented with 10% Fetal Calf Serum (Invitro Technologies) and 0.01% Penicillin/Streptomycin (Gibco). Cells were harvested in 20 ml PBS with a cell scraper and split evenly into 4 separate tubes. The tubes were spun down at 1200 rpm (290×g) for 300 s and the supernatant was decanted. The resulting cell pellets were subsequently re-suspended in PBS solution with additional 20 mM sodium potassium tartrate (Sigma-Aldrich), 0.05% Na\( \text{N}_3 \) (Sigma-Aldrich), both 0.05% Na\( \text{N}_3 \) and 20 mM sodium potassium tartrate and no additives. The cells were then incubated in the respective solutions for 600 s to simulate the maximum time taken for specimens to be processed after collection from an on-site clinic.

In 4 fresh tubes, 10 ml of isopynic gradient (1.08 g/ml Percoll (GE Healthcare)/PBS buffer) was added to each of them and the cells in the respective different buffers was layered gently on top. The tubes were then spun down for 0.5 h at 2200 rpm (974×g) at 277 K. Following centrifugation, a layer of cells should be seen at the sample/Percoll interface. The cell layer which is at the sample/Percoll interface was removed into fresh tubes and re-suspended in 25 ml of cold PBS. The cells underwent two cycles of centrifugation (2500 rpm; 1258×g) at 277 K for 600 s and the supernatant was removed. The cell pellet was then re-suspended in 1 ml Trizol (Life Technologies)/ 200 mg of pellet. RNA was isolated with Trizol reagent and synthesized according to first strand cDNA synthesis protocol.

All qRT-PCR were performed using the Corbett Rotorgene-6000 (Corbett Research, Australia), Rotorgene-6000 Series Software Version 1.7 (Corbett Research, Australia) and Quantitect® SYBR® Green PCR Kit (Cat. No. 204143, Qiagen). qRT-PCR reaction were prepared in triplicates and each contained 7.5 \( \mu \)l of qPCR master mix, 0.5 \( \mu \)l of each 10 \( \mu \)M forward and reverse primer and 5 \( \mu \)l of diluted cDNA (1:10 dilution). Cycling condition for PCA3, PSA, Hepsin and beta-2 microglobulin (\( \text{β2M} \)) primers were as follows: 368 K for 900 s, followed by 45 cycles of 368 K for 20 s, 331 K for 20 s and 345 K for 20 s. Data were generated with the Rotorgene-6000 Series Software and relative gene expression levels were calculated as described by Pfaffl et al.\(^{316}\), with \( \text{β2M} \) used as the reference gene (\( \text{RG} \)), and plotted using Excel.
3.2.3.8 Data Deposition

All primary NMR spectra and categorical metadata were deposited in the MetaboLights database (http://www.ebi.ac.uk/metabolights) with the accession number MTBLS285.

3.2.4 Results

3.2.4.1 – NMR spectroscopy of human ejaculate over time

3.2.4.1.1 – Establishing reproducibility of kinetic curves

Because NMR spectrometers can only measure one sample continuously at a time, we performed initial 1D NOESY (example in Appendix 1 – Figure A1.1a) and pseudo-2D experiments (example in Figure A1.2b) to establish whether metabolite profiles and the kinetics of phosphorylcholine degradation are reproducible in independent samples (Figure 3.1a). The pseudo-2D spectra showed time dependent changes in the signal intensities of phosphorylcholine and choline (Figure A1.2a, b) which were plotted as time-dependent decay/growth curves in Figure 3.2 (plotted as intensity ratios, which were calculated as peak integral at time \( t \) : peak integral at \( t=0 \) for phosphorylcholine, and as peak integral at time \( t \) : fitted peak integral at \( t=\infty \) for choline), together with exponential fits to the raw data. The decay/growth curves for the phosphorylcholine and choline peak integrals, respectively, were reproducible when separate samples were measured at different temperatures (277 K/ 298 K) or with/without 1 mM tartrate using a simple two-factor experiment (Figure 3.2). The spectral profiles for ejaculate samples in our dataset were comparable to those reported previously in NMR studies of this biofluid, and an NMR spectrum for a representative ejaculate sample is shown in Figure A1.1a.

3.2.4.1.2 – Kinetic curves with varied tartrate concentration and temperature

After establishing in the initial experiments that kinetic parameters were reproducible in independent samples, we used a more detailed two-factor experimental design with different temperatures and tartrate concentrations to explore the influence of both parameters (Figure 3.1b). To save time and increase throughput, experiments were interleaved by using the sample changer to cycle samples and measurements at regular intervals. In addition, measurements at 279 K and 298 K were performed on two separate NMR spectrometers in parallel. Initial choline concentration \((t=0)\) differed between samples at the same temperature, as without the addition of tartrate, a different amount of phosphorylcholine has already been hydrolysed to choline when each NMR measurement
starts. This resulted in the kinetic curves for choline starting at different amounts of choline. Similar kinetic curve patterns were observed for each subject (Figure 3.3) when compared to the curves obtained in the preliminary experiments (Figure 3.2). Increasing concentrations of tartrate and reduced storage temperature (279 K) resulted in a slower reaction, reflected by the kinetic curves (Figure 3.3), by the fitted apparent first-order rate constants (Figure 3.4), and by the ratio of change in signal after 24 h at a given tartrate concentration compared to the uninhibited sample (Appendix 1 – A1.3b). Overall the change from initial concentration of choline or phosphorylcholine was lowest at a tartrate concentration of 20 mM and the reduced storage temperature of 279 K (Figure 3.3 right half, Figure 3.4), signifying maximal inhibition of PAP.
Figure 3.2: Kinetic profiles of phosphorylcholine hydrolysis to choline mediated by prostatic acid phosphatase. The raw data points and calculated exponential fit curves are shown. First and third column of panels: The intensity of phosphorylcholine (PCho) is for samples in Figure 3.1a displayed as peak integral at time $t$ : peak integral at $t=0$ and at varying temperatures (298 K and 277 K) and in the absence (left panels) or presence (right panels) of 1 mM tartrate (Tar). Second and fourth columns: Time-dependent intensity ratios of choline (Cho), displayed as measured peak integral at time $t$: peak integral at $t=\infty$, for samples in Figure 3.1a. Starting values for choline differ because the measurement of the first spectrum of each curve occurs at different times after the reaction was initiated by sample donation, and so a different amount of PCho has been converted to Cho before the first measurement. Kinetic curves at 298 K (faster kinetics) were followed for a shorter amount of time than curves at 279 K (slower kinetics) to minimize NMR measurement time. Black curves: First sample, grey curves: second sample.
Figure 3.3: Kinetic profiles of phosphorylcholine hydrolysis to choline mediated by prostatic acid phosphatase under varying tartrate concentration. The raw data points and calculated exponential fit curves are shown. First and third column of panels: The intensity of phosphorylcholine (PCho) is for samples in Figure 3.1b displayed as peak integral at time \( t \): peak integral at \( t=0 \) and at varying temperatures (298 K; 279 K) and in the absence (left panels) or presence (right panels) of tartrate (Tar) at varying concentrations as indicated (mM). Second and fourth columns: Time-dependent intensity ratios of choline (Cho), displayed as measured peak integral at time \( t \) : peak integral at \( t=\infty \), for samples in Figure 3.1b. As in Figure 3.2, sreaction time elapsed before spectra acquisition. Kinetic curves for subject 1 and 2 at 298 K (faster kinetics) were followed for a shorter amount of time than curves at 279 K (slower kinetics) to minimize NMR measurement time. Different tartrate concentrations are denoted by different symbols: Black circles: no tartrate, open circles: 1 mM, black triangles: 3 mM, open triangles: 5 mM, black squares: 10 mM and open squares: 20 mM. The discrepancy between using a low temperature of 277 K in Figure 3.2 and 279 K in this Figure is due to 279 K being the fixed temperature at which samples are stored in a Sample Jet sample changer.
Apparent rate constants ($k_{\text{app}}$) were fitted to each exponential curve and are presented in Figure 3.4 and Table A1.1. The apparent rate constants for the hydrolysis of phosphorylcholine and the appearance of choline were reduced at 279 K compared with 298 K, as well as with increasing tartrate concentration (Figure 3.4; Table A1.1).

![Figure 3.4](image.png)

**Figure 3.4:** Apparent first-order rate constants ($k_{\text{app}}$) obtained from exponential fits of phosphorylcholine hydrolysis and choline production from the three subjects for reactions at 298 K (left panel) and 279 K (right panel). The $k_{\text{app}}$ decreases with increasing tartrate concentration. Lower $k_{\text{app}}$ were observed for samples measured at 279 K compared to 298 K.

3.2.4.1.3 – Influence of tartrate concentration and temperature on metabolomic analysis of ejaculate

After establishing that addition of tartrate to ejaculate samples inhibits the phosphorylcholine to choline hydrolysis by PAP, we characterized what effects the addition of tartrate had on the subsequent multivariate analysis of ejaculate metabolite profiles. 1D NOESY spectra of samples with 0 mM and 10 mM tartrate measured at 279 K and 298 K at 0 and 24 hours after sample preparation were spectrally aligned, Quantile normalized and analyzed with Pareto-scaled PCA.

PCA modelling of the data resulted in a model with 12 principal components ($n=66$, $k=982$, $R^2_X = 0.915$; $Q^2 = 0.764$). The samples were observed to cluster into groups corresponding to the presence/absence of tartrate (principal component 1) and inter-individual differences (principal component 2 – Figure 3.5a). The presence of tartrate (principal component 1, Figure 3.5a, b) expectedly displayed an inverse relationship with
choline (3.19, 3.20 ppm). Clustering in principal component 2 as a result of inter-individual
differences, was caused by different concentrations of glycerophosphorylcholine,
phosphorylcholine and citrate between subjects (Figure 3.5a, b). Changes in metabolite
peaks due to time and temperature were responsible for cluster separation in principal
components 3 and 5, respectively (Figure 3.5c, e). Changes due to time were due to
elevated phosphorylcholine in contrast to choline and lactate at 0 and 24 h, and with
tartrate loadings associated to be inversely associated with early (t=0) uninhibited samples
(Figure 3.5d). Temperature changes were predominantly as a result of temperature-
dependent differences in peak integrals of fructose and glycerophosphorylcholine inversely
associated with lactate and phosphorylcholine as well as peak widths of tartrate, citrate
and spermine (Figure 3.5f). Figure 3.5c also shows the effect of tartrate inhibition
strikingly: In the scores plot the distance between samples at 0 h and 24 h is
systematically larger for samples without tartrate than for samples containing 10 mM
tartrate.

Given that 10 mM tartrate was added to more than half of the ejaculate samples in this
PCA, it is trivial that the presence/absence of tartrate is the dominant factor seen in this
multivariate model. However, as the tartrate signal does not overlap with any other major
metabolite signals it can be easily excluded from the analysis to reveal the underlying
biological effects without direct interference from tartrate. Indeed, PCA modelling of the
same data with exclusion of the tartrate signal resulted in a statistical model with 9
components (n=66, k=982, R²X = 0.871; Q² = 0.772). As before, samples were observed
to cluster in distinct groups in principal components 1-3 (Figure 3.6a, c), with principal
components 1 and 3 corresponding with inter-individual metabolite differences (Figure
3.6a, c), such as glycerophosphorylcholine, choline, citrate, and spermine (Figure 3.6b, d),
previously observed in principal component 2 (Figure 3.5a, b). Sample separation in
principal component 2 was caused by an inverse relationship between choline and
phosphorylcholine due to the PAP-inhibiting effects tartrate and sample measurement
time, with tartrate containing samples measured at 24 h similar to uninhibited samples at
0 h, in general (Figure 3.6a, b). Clustering in principal component 4 was due to
temperature differences (Figure 3.6e, f), with broader peaks observed at lower
temperature, seen in principal component 5 of the previous analysis (Figure 3.5f).
Figure 3.5: Principal Components Analysis of human ejaculate NMR spectra acquired at 0 h (circles/triangles) and 24 h (squares/inverted triangles) following donation from three subjects (subject 1: red, subject 2: blue, subject 3: gold), measured at different temperatures (298 K – squares/circles; 279 K – triangles/inverted triangles), and with (filled symbols) or without (open symbols) the addition of 10 mM tartrate. Panels a, c: scores plots; panels b,d: loadings plots. Sample clustering is observed due to the presence of tartrate and subsequent inhibition of phosphophorylcholine hydrolysis to choline (principal component 1) and inter-individual differences in metabolite (glycerophosphorylcholine, phosphorylcholine and citrate) variation (principal component 2, panels a – b). Metabolite peak changes as a result of time (phosphorylcholine, choline and lactate) and temperature (Lactate, spermine, citrate, fructose, glycerophosphocholine and phosphorylcholine) and caused clustering in principal components 3 (panels c, d) and 4 (panels e, f). Metabolite abbreviations: Cho = choline; PCho = phosphorylcholine; G-P-Cho = glycerophosphorylcholine; Tar = tartrate; Cit = citrate.
3.2.4.1.1 – Univariate statistical analysis

The presence of the 10 mM tartrate signal does not just inhibit PAP, it also acts effectively as a concentration standard, and thus its presence enables us to calculate the absolute molar concentrations of other metabolites of interest. Comparing these concentrations directly is able to overcome the well-established shortcomings of most data normalisation methods, specifically that in the presence of unbalanced regulation (= a small number of high-concentration metabolites whose levels are systematically changing) the normalised levels of other minor metabolites can appear to increase when their absolute concentrations are actually decreasing and vice versa\textsuperscript{318}. To evaluate whether this is the case in ejaculate, we calculated the mean molar concentrations of the key metabolites identified in the MVSA models in Figure 3.6 for the six experimental groups comprising the tartrate-containing spectra in this analysis (three subjects, two temperatures) (Appendix 1 - Table A1.2). The mean absolute concentrations confirm that there are individual differences in glycerophosphorylcholine, citrate, lactate, spermine, phosphorylcholine and fructose between the subjects. However the interpersonal differences were opposite in the univariate analysis to what would be expected from the MVSA when comparing subjects. The univariate analysis also confirms that there are changes in the levels of phosphorylcholine, choline, citrate, spermine, fructose and lactate associated with temperature. These changes are similar in the multi- and univariate analysis except for fructose. These results demonstrate that there are multiple cases of unbalanced regulation present in the ejaculate data that can be detected by the use of tartrate as internal concentration standard. Interestingly, phosphorylcholine is – next to choline – one of the two major metabolites who dominate the NMR spectra and who are thus prime candidates for causing unbalanced regulation.

3.2.4.1.1 – Application to mRNA extraction and processing

As several groups are investigating the potential of RNA and microRNA biomarkers in ejaculate\textsuperscript{12,15,16,319,320} and ejaculate-related fluids\textsuperscript{321}, we sought to assess if the addition of tartrate, as well as sodium azide, a sample preservative, may change RNA expression and impact on subsequent interpretation of the results. We determined the expression of PSA, PCA3 and Hepsin in an \textit{in vitro} LNCaP cell culture model in the presence of tartrate and/or sodium azide. Following qPCR and relative RNA quantitation normalized to β2M, PCA3, PSA and Hepsin expression was observed to be identical in the presence tartrate, sodium azide or a combination of both (Figure A1.4). This gene expression
Figure 3.6: Principal Components Analysis of ejaculate NMR spectra after exclusion of the tartrate signal from the dataset. Panels a, c, e: scores plots; panels b, d, f: loadings plots. Sample clustering is due to inter-individual variation in principal components 1 and 3 (panels a, c), both absence/presence of tartrate and time in principal component 2 (panel a), and altered peak alignment due to temperature differences, as well as tartrate inhibition of phosphorylcholine hydrolysis to choline in principal component 4 (panels e, f). Symbols and metabolite abbreviations are as in Figure 3.5.
consistency in the presence of tartrate and sodium azide supports these agents being used as experimental adjuncts without jeopardizing the diagnostic accuracy of RNA-based tests.

3.2.5 Discussion

In this paper we established a method for reliably and reproducibly inhibiting PAP in samples of ejaculate by adding sodium tartrate to a concentration of 20 mM. This has many benefits for the use of ejaculate for clinical investigation, most notably in the early detection and surveillance of prostate cancer. This method maintained the inherent biological variation between samples despite tartrate adding an additional signal to the NMR spectrum (Figures 3.5a, b; 3.6a, b). In fact, the position of the tartrate singlet at 4.32 ppm does not significantly overlap with NMR signals of other metabolites, and can easily be excluded from the MVSA. However, the presence of the tartrate signal offers the opportunity of acting as internal concentration standard in each spectrum, allowing determination of absolute metabolite concentrations as demonstrated in Appendix 1 - Table A1.2 and detection of unbalanced regulation as shown above. While absolute metabolite concentrations in the NMR sample can also be calculated from other internal standards such as DSS and DFTMP, the presence of tartrate allows for the back calculation of metabolite concentration in the original ejaculate. Usually that calculation is imprecise as an unknown and variable amount of ejaculate is added to 20 ml of PBS at time of sample collection. If the PBS contains a known tartrate concentration, then the final tartrate concentration in the NMR sample allows calculation of the dilution factor between ejaculate and NMR sample and thus enables calculation of ejaculate sample volume and absolute metabolite concentrations in ejaculate.

Tartrate is a competitive inhibitor of PAP. This has two consequences: Firstly, the amount of inhibition is dependent on the tartrate concentration; secondly it is impossible to fully inhibit PAP with tartrate. One consequence of this is that there is residual PAP activity, observed even in samples with 20 mM tartrate at 279K, resulting in time dependent changes in choline and phosphorylcholine concentrations of 3-5% over 24 h. Furthermore, the time delay from sample production to data measurement was a factor in obtaining consistent initial (time=0) measurements for un- or weakly inhibited samples, shown by a lower relative final choline concentration compared to inhibited samples (Figures 3.2, 3.3) and increased final relative phosphorylcholine concentration at 24 h with 3 mM tartrate (Figure A1.3). This variation was unavoidable and would require exact sample preparation
and spectrometer calibration times across multiple spectrometers which is logistically not possible. However, at a tartrate concentration of 10 mM and with storage of samples at 279 K, levels of choline/phosphorylcholine changed by less than 8% over 24 h (Figure 3.3b), so that this effect has in practice been negated during typical clinical sample preparation procedures which are more likely to occur within 6 h rather than the pessimistically assumed 24 h here, allowing these metabolites to become more useful in metabolite profiling of ejaculate. For example, being able to accurately quantify choline-based metabolites may allow ejaculate to be used to indicate men who may be most suitable for prostate cancer detection or surveillance using choline-based PET imaging.

When used as a marker for CaP, PAP is elevated in the serum of men with advanced disease\textsuperscript{322}. If a comparable elevation were also present in ejaculate due to a large intraprostatic tumour burden, this could lead to a less efficient inhibition of PAP by tartrate in CaP patients. However, this is unlikely to be an issue in early disease detection when cancer burden is small and clinically unapparent. In the absence of available data, if an over-estimated 2-fold increase in ejaculate PAP in early prostate cancer were considered, a resulting 4-6% change may be seen in choline and phosphorylcholine concentrations over 6 h, the expected maximum time period for clinically collected samples, compared to the 2-3% change over 6 h observed above in healthy subjects.

Our results were obtained from three volunteers. In principle one could argue that this is a small number of study participants. However, the requirement for a large number of subjects in a human study is usually based on the need to reduce statistical uncertainty and statistical trends unrelated to biological effect when testing a hypothesis based on a cause-effect relationship in the face of biological diversity. In contrast, the objective of our study was to investigate the effects of inhibiting the ubiquitous enzyme PAP. This objective is more similar to determining the optimal or best intervention for an individual subject, similar to a N-of-1 study commonly used in clinical research. Secondly, it should be noted that the analysis of each subject included multiple repeats that all showed consistent results. Indeed, we showed that the kinetic behaviour was reliably reproducible and consistent within and across the three subjects. Thus, we were able to use multiple replicate samples to determine the effects of tartrate and temperature on the ejaculate metabolome of both volunteers. Scientifically, this is equivalent to validating experimental effects in three different cell culture methods or animal models in which the underlying biological principle is the same. Indeed, the effects of tartrate and temperature on PAP enzyme activity were consistent across all three subjects, as would be expected when
effectively performing a kinetic assay of the same enzyme in three subjects, and are thus expected to translate to samples from other patients. Finally, our study design and methods were based on a landmark study\textsuperscript{304}, which used 3 patients, each providing 2 samples, to characterise the biochemical processes in human ejaculate. Our study used 13 samples and yielded equivalent results, whilst we also explored the role of tartrate and temperature inhibition of PAP in detail.

The measurement of kinetic rate constants at three different temperatures and with varying inhibitor concentrations opens the door to several analysis methods of classical enzyme kinetics, including the calculation of the activation energy for the PAP-catalysed phosphorylcholine hydrolysis. However, such an analysis needs to be interpreted cautiously as the kinetic curves did not exactly follow first-order kinetics and the calculated rate constants are thus only apparent first-order rate constants that illustrate successful PAP inhibition. In addition, it is known that individuals might possess multiple isoforms of PAP\textsuperscript{303}, which contributes to this deviation from classical first-order kinetics, and prevents subsequent thermodynamic calculations based on the presence of a single catalytic enzyme. While testing for these multiple enzyme isoforms is beyond the scope of this manuscript, we have nevertheless undertaken an estimation of the activation energy of PAP for the three volunteers. The rationale is that if different people have different PAP isoforms (or isoform compositions) with different catalytic properties, then that should be reflected in different apparent activation energies. However, our estimation shows that the apparent activation energy of PAP is highly similar for all three volunteers (Figure A1.5), meaning that if there are different isoforms they must have similar catalytic capabilities, as has indeed been observed in the similar behaviour of the kinetic curves and rate constants under inhibition between the three volunteers (Figures 3.2-3.4). This analysis supports the above argument that our findings from these volunteers are biochemically generalizable.

3.2.6 Conclusion

In summary, we found that the addition of tartrate and cooling of ejaculate samples improves the stability of choline and phosphorylcholine levels. Although definitely not a substitute for immediate processing, for samples collected off-site with a delay to processing, sample collection in a sterile urine jar containing 10 mM tartrate in 20 ml PBS solution cooled to 277 K and cooled during transport until processing may suffice, as it would result in at most a 2-3% change in choline and phosphorylcholine concentrations over 6 h, the expected maximum time period for clinically collected samples. These
metabolite changes were characterised at 279 K and, if anything, under-estimate the expected effects. Where samples are collected on-site with near-immediate processing available, collection in a jar containing 5 mM tartrate in 20 ml PBS solution cooled to 277 K is also appropriate. This reduced tartrate concentration would be expected to result in a 3-4% change in choline and phosphorylcholine levels over 6 h. The further addition of 0.05% (w/v) sodium azide as antimicrobial preservative should be considered. This method should be implemented for ejaculate samples collected for use in multiple scientific experiments where sample integrity across multiple biomarker types is of utmost importance. Furthermore, this method may improve the clinical application of ejaculate NMR metabolomics and metabolite quantitation in biomarker discovery.
Part 3 – Results
Chapter 4 – Investigating the utility of ejaculate for non-invasive prostate cancer diagnosis

4.1 Synopsis

The aim of this chapter was to determine the feasibility and diagnostic accuracy of ejaculate as a vehicle for non-invasive CaP detection and characterisation. First, PCA3 and Hepsin mRNA and microRNAs (miRNAs) contained in the epithelial cell fraction of ejaculate were examined together with serum PSA (section 4.2). Second, metabolite profiling of ejaculate was performed using NMR-spectroscopy (section 4.3). Both studies examined the accuracy of acquired biomarkers to detect and characterise CaP and are presented as accepted (section 4.2) and submitted (section 4.3) peer-reviewed manuscripts. The preceding analyses for ejaculate metabolomics experiments are outlined in section 4.4.
4.2 “Diagnostic performance of expression of PCA3, Hepsin and miR biomarkers in ejaculate in combination with serum PSA for the detection of prostate cancer.” (published manuscript)

The manuscript entitled, “Diagnostic performance of expression of PCA3, Hepsin and miR biomarkers in ejaculate in combination with serum PSA for the detection of prostate cancer” has been published by The Prostate (2015; 75(5):539-49)

Since publication, the manuscript has been received 8, 11 and 14 citations for Web of Science, Scopus and Google Scholar, respectively (as at 10/02/17).

The contributions of the co-authors to this manuscript were as follows: Experiment design, patient recruitment, experiment execution, data processing, statistical analysis and clinical interpretation were performed by Matthew Roberts, the PhD candidate, under the supervision of Dr Horst Schirra and Prof RA “Frank” Gardiner. Renee Richards, Clement Chow and Marion Buck assisted with sample collection and biobank curation, experiment execution, data processing and data analysis. Suhail Doi provided guidance for statistical analysis and data interpretation. Hema Samaratunga, Joanna Perry-Keene, Diane Payton and John Yaxley provided advice on histological and clinical interpretation. Prof Martin Lavin provided critical manuscript revisions. Preparation of the manuscript and associated figures was performed by Matthew Roberts, the PhD candidate.

The manuscript text is presented as originally published by the journal. To maintain a consistent style between chapters in this thesis, abbreviations, placement of figures and tables within the text, as well as numbering of pages, figures and tables has been adjusted. The references have been collated in the section, ‘Bibliography’, with others contained in this thesis.
Diagnostic performance of expression of PCA3, Hepsin and miR biomarkers in ejaculate in combination with serum PSA for the detection of prostate cancer

Matthew J. Roberts¹,²,³, Clement W. K. Chow¹, Horst Joachim Schirra³, Renee Richards¹, Marion Buck⁴, Luke A. Selth⁵, Suhail A. R. Doi⁶, Hema Samaratunga⁷, Joanna Perry-Keene⁸, Diane Payton⁸, John Yaxley², Martin F. Lavin¹, Robert A. Gardiner, ¹, ²#

¹ The University of Queensland, Centre for Clinical Research, Brisbane, Qld 4006, Australia
² Department of Urology, Royal Brisbane and Women’s Hospital, Brisbane, Qld 4006
³ The University of Queensland, Centre for Advanced Imaging, Brisbane, Qld 4006, Australia
⁴ Department of Environmental Health Sciences, University Medical Centre Freiburg, 79106 Freiburg, Germany
⁵ Dame Roma Mitchell Cancer Research Laboratories and Adelaide Prostate Cancer Research Centre, The University of Adelaide, South Australia, Australia and Freemasons Foundation Centre for Men’s Health, The University of Adelaide, South Australia, Australia
⁶ School of Population Health, The University of Queensland, Brisbane, Queensland, Australia
⁷ Aquesta Pathology, Brisbane, Qld 4066, Australia
⁸ Anatomical Pathology, Pathology Queensland, Brisbane, Qld 4006, Australia
4.2.1 Abstract

**Background and Methods:** Here, we report on the evaluation of the diagnostic performance of ejaculate-derived PCA3, Hepsin and miRNAs to complement serum PSA to detect prostate cancer. cDNA was prepared from 152 candidate specimens following RNA isolation and amplification for PSA, PCA3 and Hepsin qPCR, with 66 having adequate RNA for all 3 assays. Small RNA sequencing and examination of CaP-associated miRNAs miR-200b, miR-200c, miR-375 and miR-125b was performed on 20 specimens. We compared findings from prostate biopsies using D'Amico and PRIAS classifications and in relation to whole gland histopathology following RP. Multivariate logistic regression modeling and clinical risk (incorporating standard clinicopathological variables) were performed for all ejaculate-based markers.

**Results:** While Hepsin alone was not of predictive value, the Hepsin:PCA3 ratio together with serum PSA, expressed as a univariate composite score based on multivariate logistic regression, was shown to be a better predictor than PSA alone of prostate cancer status ($AUC = 0.724 \text{ vs } 0.676$) and risk, using D'Amico ($AUC = 0.701 \text{ vs } 0.680$) and PRIAS ($AUC = 0.679 \text{ vs } 0.659$) risk stratification criteria as classified using prostate biopsies. It was also possible to analyse a subgroup of patients for miRNA expression with miR-200c ($AUC = 0.788$) and miR-375 ($AUC = 0.758$) showing best single marker performance, while a combination of serum PSA, miR-200c and miR-125b further improved prediction for prostate cancer status when compared to PSA alone determined by biopsy ($AUC = 0.869 \text{ vs } 0.672; p < 0.05$), and risk (D'Amico/PRIAS) as well as by RP histology ($AUC = 0.809 \text{ vs } 0.690$). For prostate cancer status by biopsy, at a sensitivity of 90%, the specificity of the test increased from 11% for PSA alone to 67% for a combination of PSA, miR-200c and miR-125b.

**Conclusion:** These results show that use of a combination of different types of genetic markers in ejaculate together with serum PSA are at least as sensitive as those reported in DRE urine. Furthermore a combination of serum PSA and selected miRNAs improved prediction of prostate cancer status. This approach may be helpful in triaging patients for MRI and biopsy, when confirmed by larger studies.

**Keywords:** prostate cancer: PCA3, Hepsin and PSA: mRNA and microRNA: ejaculate
4.2.2 Introduction

The two-step process (PSA and biopsy) of early detection of CaP is fraught with imprecision. Initial limitations relate to deficiencies of the standard first-line test, serum PSA, which lacks a lowermost cut-point and has a continuum of CaP risk with increasing values\textsuperscript{323}. However, despite its deficiencies, published evidence has indicated that total serum PSA concentration remains the single most significant clinically-used predictive factor for identifying men at increased risk for CaP\textsuperscript{191} and is the parameter on which adaptations such as free:total PSA, PSA velocity, PSA density and phi are based\textsuperscript{193}.

An abnormal serum PSA level and/or an abnormal DRE lead to the second step in detecting CaP, namely prostatic biopsy. Since histological evidence of tumour is required for diagnosis and the majority of tumours arise posteriorly in the peripheral zone, TRUSBx has become entrenched as the standard method for obtaining prostatic tissue for diagnosis, with \textgreater 10 cores now generally recommended\textsuperscript{324}. However, TRUSbx targets the anterior prostate poorly so that up to approximately one third of CaPs may be missed by this approach\textsuperscript{213}, particularly in large glands\textsuperscript{212}. Transperineal biopsy (TPbx) is theoretically a more accurate technique but is used less often due to time and logistical constraints\textsuperscript{325}. Consequently, TRUS directed biopsy remains the standard approach, at least in the first instance. These limitations become even more significant when it is appreciated that the majority of men undergoing TRUSbxs for suspected CaP do not have CaP detected and may have further biopsies, as do men with low-risk CaP who elect to be managed by an active surveillance regimen\textsuperscript{213}.

Of the non-PSA based first line tests, the PCA3 urine test is most widely used\textsuperscript{326} and involves analyzing the first part of a specimen of voided urine after milking the prostate by firm DRE or prostatic massage to provide prostatic fluid from the posterior part of the gland. This approach is required if voided urine is to be used since there are otherwise too few cells present in a freely voided specimen. Although the PCA3 test has been reported to improve detection of CaP compared with PSA in a pre-screened population, its role in initial assessment of patients suspected of having CaP has yet to be established as a first-line, stand-alone investigation\textsuperscript{327}. Addition of other RNA markers to the ‘PCA3 urine test’, such as the TMPRSS2:ERG fusion gene, has been reported in some, but not all cases, to improve cancer prediction\textsuperscript{10,211,328,329}.

Increasingly, mpMRI is being employed for CaP detection, although it is yet to be approved for routine clinical use\textsuperscript{330}. An attraction is that MRI permits avoidance of many
unnecessary biopsies and facilitates targeted rather than multiple ‘blind’ biopsies to locate significant malignant lesions\textsuperscript{331}. However, mpMR imaging misses up to 30\% of CaPs and up to 10\% of significant tumours\textsuperscript{330} so, together with logistical and cost implications, its acceptability as a further investigation for every man suspected of harbouring CaP is limited. Consequently, the need for an inexpensive, non-invasive test to determine whether or not CaP is present and, better still, if a lesion is significant, remains an imperative.

Approximately 40\% of ejaculate is derived from prostatic epithelia so that its contents are most likely to contain disease-specific biomarkers. Following our observation that abnormal prostatic cells in ejaculate correlated with and predated biopsy diagnosis of CaP, in some cases by years\textsuperscript{14}, we identified 4 mRNA markers that detected 100\% of CaPs in tissues\textsuperscript{206,241}. PCA3 and Hepsin have proven to be the most predictive\textsuperscript{332}. In preliminary analyses, a combination of PCA3 and Hepsin with serum PSA consistently provided a combined sensitivity and specificity >75\%\textsuperscript{332}. Our results with ejaculate paralleled those of other investigators who examined a range of markers, mostly in urine, so we have continued to evaluate ejaculate, with respect to other marker expression to complement PCA3 and Hepsin expression and improve sensitivity and specificity. We recently profiled and compared ejaculate miRNAs from men with subsequently proven CaP by biopsy and men suspected of having CaP but biopsy results were negative\textsuperscript{16}. Using deep small RNA sequencing followed by quantitative RT-PCR validation, we showed that 5 miRNAs were significantly elevated in ejaculate from CaP patients and that each of these miRNAs outperformed PSA as a diagnostic tool\textsuperscript{16}. These observations are in keeping with pooled results from different studies which suggest that miRNA analysis can significantly improve the overall accuracy of CaP diagnosis\textsuperscript{333}. Our focusing on ejaculate as relevant for prostate cancer research has been supported by others who have also recognized its diagnostic potential\textsuperscript{334,335}.

The aim of this study was to examine the predictive ability of PCA3 and Hepsin mRNA from ejaculate together with serum PSA to identify those patients who presented for prostatic biopsy but who, with the benefit of hindsight, may have been spared this procedure and treatment with curative intent had discriminating test(s) been available clinically. We also determined the potential of miRNA, previously identified by this group as elevated in CaP, in combination with the other markers to improve diagnosis. In order to maximise accuracy, we updated information from repeated biopsies to exclude those patients whose initial TRUSbxs had missed their tumours. In addition, we included more
accurate RP findings from a subgroup of patients, none of the men in this study having had mpMRI examinations to aid in locating dominant lesions.

4.2.3 Materials and Methods

4.2.3.1 – Patients

Men \((n = 152)\) suspected of having CaP on the basis of an elevated serum PSA and/or an abnormal DRE provided specimens of ejaculate collected immediately in sterile micro-urine jars containing 20 ml Hanks Balanced Salt Solution (HBSS) (Gibco), subsequently replaced by PBS for the last 35 patients (due to glucose content in HBSS affecting other biomarker discovery techniques) between January 2007 and February 2013. All participants have been followed closely to monitor initial biopsy findings, in particular to identify false negative findings subsequently corrected by further biopsy results, and to relate biopsy histopathology findings to those from RP for men who proceeded to surgery \((n = 61)\). Ejaculate specimens were collected prior to TRUSbx or over one month following TRUSbx or TPbx. All specimens were brought without cooling to the hospital by the patients and transferred to the laboratory and processed within 2 h of production.

Ethics approval was obtained from the University of Queensland Medical Research Ethics Committee (Project no. 2006000262) and the RBWH Human Research Ethics Committee (HREC/09/QRBW/320, HREC/09/QRBW/305 together with 1995/088B).

4.2.3.2 – Clinical data

Clinical details for each patient, including serum PSA, were obtained through Queensland Health pathology software, AUSLAB, supplemented by reports outside the hospital from the private pathology firms Queensland Medical Laboratories (Siemens Centaur ® assay) and Sullivan Nicolaides Pathology (Abbott Architect ® assay).

All TRUSbx and TPBx histology specimens were reported and reviewed by DP, JP-K and MLTHS using the 2005 International Society of Urological Pathology (ISUP) classification\(^{336}\) with standard biopsy (number of cores, number of positive cores, percentage, Gleason score) and RP (Gleason score, pathological stage, gland size, margin status, extracapsular status and pathological stage) parameters recorded. Risk stratification for biopsies were those recommended in the American Urological Association Guidelines based on D’Amico et al.\(^{337}\) for biopsies, and the Prostate Cancer Research International: Active Surveillance (PRIAS) inclusion criteria (low-risk category)\(^{338}\), and
those outlined in Table 4.1. Our analyses of the PCR findings were based on histology that was most representative of tumour grade both TRUSbx and TPBx, as well as RP histopathology because of the well-known disparity between many TRUSbx and RP histopathology findings.

Patients were classified into two clinical groups according to the three different classification schemes outlined in Table 4.1. They cover either:

1. a distinction between any presence of CaP and no evidence of malignancy (noCaP), (Scheme 1) based on pathological evidence of CaP on either initial biopsy histology, or on updated histology for those men who had further biopsies or tissue sampling (e.g. via transurethral resection) as indicated clinically, or

2. a distinction between high/intermediate-risk CaP and low-risk CaP combined with noCaP based on D’Amico (Scheme 2) or PRIAS (Scheme 3). Schemes 2 and 3 thus distinguish between the types of CaP requiring medical intervention (scheme 2: D’Amico high/intermediate risk; scheme 3: PRIAS high/intermediate risk) and those prostatic conditions that do not warrant definitive cancer treatment (scheme 2: D’Amico negative/low risk; scheme 3: PRIAS negative/low risk).

<table>
<thead>
<tr>
<th>Classification Scheme</th>
<th>Classification categories</th>
<th>Serum PSA</th>
<th>Biopsy</th>
<th>RP</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Cancer status based on biopsy</td>
<td>noCaP</td>
<td>Serum PSA not used for classification</td>
<td>Absence of CaP</td>
<td>Presence of CaP</td>
</tr>
<tr>
<td></td>
<td>CaP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Risk status based on D’Amico</td>
<td>Negative/ Low risk</td>
<td>Any Serum PSA AND absence of CaP on Biopsy</td>
<td>≤ 10 ng/ml</td>
<td>Gleason ≤ 3+3</td>
</tr>
<tr>
<td></td>
<td>High risk</td>
<td></td>
<td>&gt; 10 ng/ml</td>
<td>Gleason ≥ 3+4</td>
</tr>
<tr>
<td>3. Risk status based on PRIAS</td>
<td>Negative/ Low risk</td>
<td>Any Serum PSA AND absence of CaP on Biopsy</td>
<td>≤ 10 ng/ml</td>
<td>Gleason ≤3+3 in &lt;50% of no more than 2 cores and cT1c/cT2 stage</td>
</tr>
<tr>
<td></td>
<td>High risk</td>
<td>&gt; 10 ng/ml</td>
<td>All other CaP</td>
<td>All other CaP</td>
</tr>
</tbody>
</table>

*Table 4.1*: Patient classification schemes, including both CaP status (positive/negative) and clinical risk (as determined by D’Amico or PRIAS criteria). noCaP = no evidence of malignancy; High risk = requires immediate treatment; Negative/Low risk = no immediate treatment is required. *a* The latter group is ineligible for Active Surveillance and deferred active treatment or further investigation is advised. The table includes two stratifications for serum PSA together with biopsy and RP findings.
4.2.3.3 – Radical Prostatectomy

Because there are often disparities between biopsy and RP histopathology (with the latter considered to be the ‘gold standard’), we also analysed our data using RP histopathology as the benchmark as opposed to biopsy.

4.2.3.4 – Specimen processing

Ejaculate reconstituted in either 20 ml Hanks, subsequently changed to PBS to avoid confounding effects of glucose in parallel metabolomics studies, was layered over 10 ml isotonic Percoll (GE Healthcare-Pharmacia) and centrifuged at 974×g for 30-60 min at 4°C and samples were collected in 1 ml aliquots, snap-frozen on dry ice and stored at -80°C. The epithelial cell layer at the Percoll interface (which appeared as a discrete band in the middle of the test tube with supernatant above and with sperm and non-cellular components at the bottom was then pipetted, washed with 25 ml PBS or Hanks and centrifuged at 1258×g for 10 min at 4°C. Cells were subsequently subjected to total RNA isolation.

4.2.3.5 – RNA isolation and amplification

Total RNA was isolated using TRizol reagent (Invitrogen) followed by a clean-up inclusive DNase treatment with the RNeasy kit (Qiagen). Amplification was performed with the SenseAmp kit (Gensisphere). For the cDNA synthesis, Superscript III (Invitrogen) and random hexamer primers were used.

4.2.3.6 – cDNA synthesis and qPCR

qPCR was performed using the QIAGEN Quantitect SYBR green qPCR Mastermix (QIAGEN, Germany) on a Corbett Rotorgene machine 3000/6000 (Corbett Research, Australia). 200-500 ng of total RNA were reverse transcribed into cDNA using Superscript III (Invitrogen) in the presence of random primers (Promega). The synthesized cDNA was diluted 10-fold and 5 µl was used for the assay in the presence of 7.5 µl Quantitect SYBR green mastermix (Qiagen) and 5 pmol gene specific forward and reverse primers. Each reaction was carried out in triplicates for both patient samples and calibrator. Reaction conditions were 95°C for 15 min followed by 45 cycles of 20 s at 95°C, 20 s at 58°C and 20 s at 72°C. Data for each cycle was acquired at the 72°C for 20 s step.

The primers (Sigma-Aldrich, Australia) for qPCR were: PCA3 (Fwd: 5’-GGAAGGACCTGATGACAGGAGTGGTACG-3’, Rev: 5’-CACAGGGCGAGGCTCATCG-3’);
Hepsin (Fwd: 5’-TGTACCCAGTGACGTACG-3’, Rev: 5’-CGTCCCTTCGTTTGCCTCAA-3’ and PSA (Fwd: 5’-GCATCAGGAACAAAGTG-3’, Rev:5’-CCTGAGGAAATCGATTCTTCA-3’).

All runs underwent standardized processing (including standard curve fitting, dynamic tube, slope correct, using Rotor-Gene 6000 Series Software Version 1.7 (Corbett Research, Australia). Samples with atypical melt curves as well as quantitation curves that did not pass the threshold value for any single target (but had normal melt curves) were excluded (n = 55), as were samples with only one technical replicate yielding results (n = 8), or samples that yielded insufficient RNA for cDNA synthesis (n = 11). A single threshold value was manually set for each individual gene across all samples and subsequently calculated cycle threshold (Ct) values were exported to Excel (Microsoft Corporation, Redmond, CA). Based on average Ct (CtAv) and standard deviation (CtSD) values observed across all tests (see Appendix 2 – Figure A2.1), CtAv > 35 and CtSD > 1.5 were defined as the cutoff points for exclusion (n = 12) leaving 66 samples for further data analysis.

A method for calculating relative gene expression, previously described by Pfaffl\textsuperscript{316}, was applied in the analysis of the real-time PCR results. The Pfaffl equation is as follows:

\[
\text{Expression (R)} = \frac{E_{\text{GOI}}^{\Delta \text{Ct}(\text{Calibrator-sample})}}{E_{\text{Ref}}^{\Delta \text{Ct}(\text{Calibrator-sample})}}
\]

“\(\Delta \text{Ct} (\text{Calibrator - sample})\)” in this equation estimates the difference in amplification (Ct values) between the target gene transcription of the sample reaction (unknown) and the calibrator reaction (uniform template quantity to standardize all the runs)\textsuperscript{316}. This method also considers the reaction efficiency (\(E\)) of the gene of interest (GOI) and the endogenous RG, without a requirement for use of a standard curve in every run by assuming the reaction efficiency between different runs was consistent and normalized by the calibrator used. In accordance with the well-established method for determining PCA3, the RG used was PSA\textsuperscript{208}.

4.2.3.7 – Inclusion of miRNAs

Recent investigations by our group confirmed ejaculate as a source of prostate cancer specific miRNA biomarkers\textsuperscript{16}. Small RNA-seq (TruSeq Workflow, Illumina; San Diego, CA, USA) was used to profile RNA collected from the non-sperm cellular fraction of ejaculate, processed as described above and in Selth et al\textsuperscript{16}. CaP-associated miRNAs (miR-200b, miR-200c, miR-375, miR-125b) were examined for 48 patients, a subset for whom PCA3
and Hepsin were also appropriately quantified in the current study (\(n=20\)). The previously generated miRNA data set were thus incorporated with the new data in a stepwise logistic regression for multivariate predictability.

4.2.3.8 – Data analysis

Data were collated in Excel. Hepsin and PCA3 scores were calculated with reference to PSA in ejaculate. A Hepsin:PCA3 ratio was also calculated, following preliminary analysis that indicated the two markers (relative to PSA-PCR) to be inversely associated (see Results).

The qPCR results were analysed using the three clinical classifications outlined in Table 4.1 and are reported for their ability to distinguish:

1. CaP versus negative biopsy/no CaP (noCaP)
2. Based on D’Amico, those requiring treatment (D’Amico high/intermediate risk) versus those not requiring treatment (D’Amico negative/low risk)
3. Based on PRIAS, those requiring treatment (PRIAS high/intermediate risk) versus those eligible for active surveillance, thus not requiring treatment (PRIAS negative/low risk)

First, clinical characteristics were reported based on presence or absence of CaP. Data was tested for normality with the Shapiro-Wilk test. Where normality was accepted, significance testing was performed using Student’s \(t\)-test (following \(F\)-testing indicating equal variance), and where normality was rejected, the non-parametric Mann-Whitney \(U\) test was used. The predictive ability of each clinical marker was determined and compared using Receiver Operating Characteristic (ROC) analysis, calculated using the DeLong method\(^{339}\) and binomial exact Confidence Interval. Stepwise logistic regression (\(p\) to enter 0.3; \(p\) to remove 0.35) was performed to ascertain the independent predictors by the CaP criterion (which was expressed as categorical variables: 1 = present, 0 = missing). This was done separately after inclusion of PCR markers (composite 1) and after inclusion of both PCR and miR markers (composite 2) with the PSA. The beta coefficients of selected variables were used to create the composite scores for composite1 and composite2 and their diagnostic performance evaluated and compared to PSA alone. Since stepwise regression selects only independent predictors of outcome into the model, we verified that our final models were properly specified by using a model specification (link) test, making it unlikely that any correlated variables were included in the model. All reported \(p\)-values <
Statistical analysis was performed using MedCalc for Windows, Version 12.7 (MedCalc Software; Ostend, Belgium)

4.2.3 Results

4.2.3.1 – Clinical characteristics

Of the 152 men who provided specimens between January 2007 and February 2013, 54 had non-cancerous histology and 98 had cancer detected by biopsy. Across the six year follow up period, 18 patients (12%) demonstrated CaP on further biopsies (Table 4.2), which were performed because of a clinical suspicion that CaP had been missed. The median (interquartile range; IQR) time elapsed between the first and second biopsy was 12.51 (9.27 – 21.76) months.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>n</th>
<th>noCaP</th>
<th>CaP (on Bx)</th>
<th>CaP (treated by RP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>66</td>
<td>64 (59 – 71)</td>
<td>59 (55 – 68)NS</td>
<td>56 (54 – 65.5)</td>
</tr>
<tr>
<td>Serum PSA</td>
<td></td>
<td>5.8 (3.25 – 6.95)</td>
<td>6.5 (5.13 – 10.25)*</td>
<td>6.25 (4.35 – 9.8)</td>
</tr>
<tr>
<td>PCA3</td>
<td></td>
<td>0.99 (0.38 – 1.73)</td>
<td>1.69 (0.73 – 7.77)NS</td>
<td>1.95 (0.78 – 8.69)</td>
</tr>
<tr>
<td>Hepsin</td>
<td></td>
<td>0.47 (0.14 – 1.96)</td>
<td>0.37 (0.11 – 1.74)NS</td>
<td>0.33 (0.18 – 1.40)</td>
</tr>
<tr>
<td>Hepsin/PCA3</td>
<td></td>
<td>1.13 (0.25 – 2.53)</td>
<td>0.21 (0.06 – 1.52)NS</td>
<td>0.21 (0.02 – 1.69)</td>
</tr>
<tr>
<td>miR-200b</td>
<td>20</td>
<td>-6.04 (-9.31 – -5.12)</td>
<td>-4.76 (-6.28 – -2.35)NS</td>
<td>-5.93 (-6.85 – -4.12)</td>
</tr>
<tr>
<td>miR-200c</td>
<td></td>
<td>-3.86 (-4.87 – -3.02)</td>
<td>-2.24 (-3.36 – -1.45)*</td>
<td>-3.16 (-4.09 – -2.34)</td>
</tr>
<tr>
<td>miR-375</td>
<td></td>
<td>-5.32 (-6.22 – -4.36)</td>
<td>-3.45 (-5.25 – -2.54)NS</td>
<td>-4.94 (-5.58 – -3.72)</td>
</tr>
<tr>
<td>miR-125b</td>
<td></td>
<td>-4.90 (-5.44 – -3.71)</td>
<td>-3.88 (-5.45 – -1.69)NS</td>
<td>-5.20 (-6.10 – -3.25)</td>
</tr>
</tbody>
</table>

**Table 4.2:** Demographic information for patients undergoing prostate biopsy (Bx) based on different classification methods used (D’Amico, PRIAS). * p < 0.05, NS = non-significant following Mann-Whitney U-test (two-tailed) based on raw values for "No CaP" versus "CaP (on Bx)".

- **D’Amico (per Bx histology)**
  - Negative/Low risk: 21 (32%) vs 8 (12%) vs 3 (5%)
  - High risk: 37 (56%) vs 25 (38%)²

- **PRIAS (per Bx histology)**
  - Negative/Low risk: 21 (32%) vs 7 (10%) vs 2 (3%)
  - High risk: 38 (58%) vs 26 (39%)²

This group is ineligible for Active Surveillance and deferred active treatment or further investigation is advised. ² This group contains patients who were classified as Intermediate or High risk (D’Amico) or ineligible for Active Surveillance (PRIAS). ³ All patients were subsequently classified as high risk based on RP histology (n = 28).
Of the 98 patients with a prostate biopsy positive for CaP, 61 underwent RP (Table 4.2). Of the remaining 37 patients, 17 proceeded to receive radiation based therapy (External Beam Radiation Therapy or brachytherapy), one commenced androgen deprivation therapy for metastatic disease, 7 were lost to follow up, and the remainder \((n=12)\) are undertaking conservative management in the form of active surveillance \((n=7)\) or watchful waiting \((n=5)\). When patients were classified into D'Amico or PRIAS risk groups as outlined in the methods section, the group assignments produced identical statistical results reflecting comparability of these stratification schemes.

Following relative expression calculations, inclusion of PSA, PCA3 and Hepsin was possible for 66 patients. These were due to strict exclusion criteria (see Materials and Methods section and Figure A2.2), with 63 having unsatisfactory q-PCR analyses (55 with atypical melt or quantitation curves and 8 having only single samples), 11 having insufficient RNA for cDNA synthesis and 12 with \(Ct\) values \((CtAv\) or \(CtSD\)\) outside defined cut-offs. The relationships observed for the entire cohort with respect to age and serum PSA were preserved in this group.

Demographic information on the cohort including median (IQR) age, serum PSA, PCA3, Hepsin, Hepsin/PCA3 and miR’s (miR-200c, miR-200b, miR-375, miR-125b) for each group are presented in Table 4.2. The categorization schemes outlined in Table 4.1 are also indicated in terms of CaP diagnosis via biopsy. Serum PSA was significantly higher in men with CaP on biopsy \([6.5 \text{ (IQR 5.13 – 10.25) ng/ml}]\) compared with those in the no CaP group \([5.8 \text{ (IQR 3.25 – 6.95) ng/ml}; p < 0.05]\).

4.2.3.2 – Ejaculate PCA3 and Hepsin expression performance using the biopsy criterion

For biopsy-based CaP status and risk classification (Table 4.3), PCA3 and Hepsin/PCA3 displayed varied performance depending on the classification method, with best performance observed for CaP status being \(AUC\) of 0.625 (95% CI 0.498 – 0.742) for PCA3 and 0.650 (95% CI 0.522 – 0.763) for Hepsin/PCA3 and both were inferior to serum PSA with an \(AUC\) of 0.676 (95% CI 0.550 to 0.786). Performance of Hepsin by itself was consistently poor across all classifications \((AUC\) range 0.517 – 0.559), further supporting its use as a RG in this context. For prediction of CaP status, a logistic regression based composite score combining serum PSA and Hepsin:PCA3 demonstrated best discriminative performance \([AUC 0.724 (95\% \text{ CI 0.600 – 0.827}), \text{ Figure 4.1A}]\). Similar estimates were observed for predicting CaP risk using both D’Amico (Table 4.4) and PRIAS (Table 4.5) based risk classifications.
<table>
<thead>
<tr>
<th>n</th>
<th>Biomarker</th>
<th>AUC</th>
<th>95% CI</th>
<th>Sp (95% CI) a</th>
<th>LR+</th>
</tr>
</thead>
<tbody>
<tr>
<td>66</td>
<td>Serum PSA b, c (cut-off &gt;3.45ng/ml)</td>
<td>0.676</td>
<td>0.550 to 0.786</td>
<td>28.57 (4.76 - 47.62)</td>
<td>1.28 (91% Sn)</td>
</tr>
<tr>
<td></td>
<td>Hepsin (cut-off ≤7.1428)</td>
<td>0.517</td>
<td>0.391 to 0.642</td>
<td>9.52 (0.00 - 28.57)</td>
<td>1.03 (93% Sn)</td>
</tr>
<tr>
<td></td>
<td>PCA3 (cut-off &gt;0.1827)</td>
<td>0.625</td>
<td>0.498 to 0.742</td>
<td>19.05 (0.00 - 33.33)</td>
<td>1.07 (87% Sn)</td>
</tr>
<tr>
<td></td>
<td>Hepsin:PCA3 b (cut-off ≤4.296)</td>
<td>0.650</td>
<td>0.522 to 0.763</td>
<td>19.05 (0.00 - 38.10)</td>
<td>1.13 (91% Sn)</td>
</tr>
<tr>
<td></td>
<td>Composite1 b (cut-off &gt;3.0492)</td>
<td>0.724</td>
<td>0.600 to 0.827</td>
<td>38.10 (9.52 - 61.90)</td>
<td>1.47 (91% Sn)</td>
</tr>
<tr>
<td>20</td>
<td>miR-125b c (cut-off &gt;-6.3274)</td>
<td>0.606</td>
<td>0.366 to 0.813</td>
<td>22.22 (0.00 - 66.67)</td>
<td>1.17 (91% Sn)</td>
</tr>
<tr>
<td></td>
<td>miR-200b (cut-off &gt;-7.0847)</td>
<td>0.717</td>
<td>0.475 to 0.892</td>
<td>44.44 (0.00 - 88.89)</td>
<td>1.64 (91% Sn)</td>
</tr>
<tr>
<td></td>
<td>miR-200c c (cut-off &gt;-4.4428)</td>
<td>0.788</td>
<td>0.550 to 0.936</td>
<td>33.33 (0.00 - 55.56)</td>
<td>1.50 (100% Sn)</td>
</tr>
<tr>
<td></td>
<td>miR-375 (cutoff &gt;-5.7315)</td>
<td>0.758</td>
<td>0.517 to 0.918</td>
<td>44.44 (0.00 - 66.67)</td>
<td>1.80 (100% Sn)</td>
</tr>
<tr>
<td></td>
<td>Composite2 c (cut-off &gt;-2.4592)</td>
<td>0.869</td>
<td>0.705 to 1.00</td>
<td>66.67 (5.00 - 100.00)</td>
<td>2.7 (91% Sn)</td>
</tr>
</tbody>
</table>

Table 4.3: ROC analysis summary for available biomarkers for CaP status based on most recent biopsy histology. a Sp: Specificity (95% CI – BC a bootstrap interval with 1000 iterations) at 90% Sensitivity; b Composite 1 based on PSA and Hepsin/PCA3; c composite2 based on PSA, miR200c and miR125b. LR+ = positive likelihood ratio.

**Figure 4.1:** Receiver Operating Characteristic curve analysis of serum PSA and Composite1 (serum PSA, Hepsin:PCA3; panel A; p > 0.05) and Composite2 (serum PSA, miR-125b, miR-200c; panel B; *p = 0.044) in predicting prostate cancer status (based on biopsy result). The area under the curve (AUC) is indicated.
<table>
<thead>
<tr>
<th>Biomarker</th>
<th>N</th>
<th>Biopsy</th>
<th>RP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>AUC</td>
<td>95% CI</td>
</tr>
<tr>
<td>Serum PSA</td>
<td>66</td>
<td>0.680</td>
<td>0.554 to 0.790</td>
</tr>
<tr>
<td>Hepsin</td>
<td></td>
<td>0.551</td>
<td>0.423 to 0.674</td>
</tr>
<tr>
<td>PCA3</td>
<td></td>
<td>0.566</td>
<td>0.438 to 0.687</td>
</tr>
<tr>
<td>Hepsin:PCA3</td>
<td></td>
<td>0.563</td>
<td>0.435 to 0.685</td>
</tr>
<tr>
<td>Composite1</td>
<td></td>
<td>0.701</td>
<td>0.576 to 0.807</td>
</tr>
<tr>
<td>miR-125b</td>
<td>20</td>
<td>0.542</td>
<td>0.308 to 0.763</td>
</tr>
<tr>
<td>miR-200b</td>
<td></td>
<td>0.646</td>
<td>0.404 to 0.843</td>
</tr>
<tr>
<td>miR-200c</td>
<td></td>
<td>0.719</td>
<td>0.476 to 0.894</td>
</tr>
<tr>
<td>miR-375</td>
<td></td>
<td>0.667</td>
<td>0.424 to 0.858</td>
</tr>
<tr>
<td>Composite2</td>
<td></td>
<td>0.844</td>
<td>0.655 to 1.00</td>
</tr>
</tbody>
</table>

Table 4.4: ROC analysis summary for available biomarkers using D’Amico-based risk stratification as determined by biopsy and RP histology. a Sp: Specificity (95% CI – BCa bootstrap interval with 1000 iterations) at 90% Sensitivity; b Composite 1 based on PSA and Hepsin/PCA3; c composite2 based on PSA, miR200c and miR125b. LR+ = positive likelihood ratio.
### Scheme 3 – PRIAS risk based on Biopsy

<table>
<thead>
<tr>
<th>Biomarker</th>
<th>AUC</th>
<th>95% CI</th>
<th>Sp (95% CI)</th>
<th>LR+ (sensitivity)</th>
<th>AUC</th>
<th>95% CI</th>
<th>Sp (95% CI)</th>
<th>LR+ (sensitivity)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>N</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>66</td>
<td></td>
<td>Serum PSA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(cut-off &gt;3.48; &gt;3.36 ng/ml)</td>
<td>0.659</td>
<td>0.532 to 0.772</td>
<td>25.00</td>
<td>(3.57 - 39.29)</td>
<td>1.19 (89% Sn)</td>
<td>0.619</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hepsin</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(cut-off &gt;0.0871; ≤5.8126)</td>
<td>0.559</td>
<td>0.432 to 0.681</td>
<td>21.43</td>
<td>(0.00 - 39.29)</td>
<td>1.14 (89% Sn)</td>
<td>0.520</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PCA3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(cut-off &gt;0.1651; &gt;0.1744)</td>
<td>0.547</td>
<td>0.420 to 0.670</td>
<td>14.29</td>
<td>(0.00 - 25.00)</td>
<td>1.14 (97% Sn)</td>
<td>0.643</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hepsin:PCA3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(cut-off ≤3.8556; ≤6.0477)</td>
<td>0.542</td>
<td>0.415 to 0.666</td>
<td>17.86</td>
<td>(0.00 - 28.57)</td>
<td>1.12 (92% Sn)</td>
<td>0.651</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Composite1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(cut-off &gt;3.0729 ; &gt;2.8446)</td>
<td>0.679</td>
<td>0.552 to 0.788</td>
<td>32.14</td>
<td>(8.43 - 50.00)</td>
<td>1.36 (92% Sn)</td>
<td>0.665</td>
</tr>
<tr>
<td>20</td>
<td></td>
<td>miR-125b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(cut-off ≤-1.3024; &gt;-6.3274)</td>
<td>0.505</td>
<td>0.276 to 0.732</td>
<td>9.09</td>
<td>(0.00 - 27.27)</td>
<td>1.10 (100% Sn)</td>
<td>0.606</td>
</tr>
<tr>
<td></td>
<td></td>
<td>miR-200b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(cut-off &gt;-14.1977; &gt;-7.0847)</td>
<td>0.596</td>
<td>0.357 to 0.806</td>
<td>0.00</td>
<td>(0.00 - 0.00)</td>
<td>0.89 (89% Sn)</td>
<td>0.717</td>
</tr>
<tr>
<td></td>
<td></td>
<td>miR-200c</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(cut-off &gt;-4.4824; &gt;-4.4428)</td>
<td>0.646</td>
<td>0.404 to 0.844</td>
<td>27.27</td>
<td>(0.00 - 45.45)</td>
<td>1.37 (100% Sn)</td>
<td>0.788</td>
</tr>
<tr>
<td></td>
<td></td>
<td>miR-375</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(cut-off &gt;-5.7583; &gt;-5.7315)</td>
<td>0.606</td>
<td>0.366 to 0.813</td>
<td>36.36</td>
<td>(0.00 - 54.55)</td>
<td>1.57 (100% Sn)</td>
<td>0.758</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Composite2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(cut-off &gt;-10.8483; &gt;-2.4592)</td>
<td>0.758</td>
<td>0.522 to 0.993</td>
<td>18.18</td>
<td>(0.00 - 36.36)</td>
<td>1.22 (100% Sn)</td>
<td>0.869</td>
</tr>
</tbody>
</table>

Table 4.5: ROC analysis summary for available biomarkers using PRIAS-based risk stratification as determined by biopsy and RP histology.  

*Sp*: Specificity (95% CI – BCa bootstrap interval with 1000 iterations) at 90% Sensitivity;  

*Composite 1 based on PSA and Hepsin/PCA3;  

*Composite2 based on PSA, miR200c and miR125b. LR+ = positive likelihood ratio.
4.2.3.3 – Ejaculate PCA3 and Hepsin expression performance using RP histology criterion

Risk classification according to RP histology was possible for a subgroup of 49 patients (Table 4.2), in which D’Amico and PRIAS classifications were identical. Within this group, the composite score (Composite1) combining Hepsin:PCA3 with serum PSA [AUC 0.665 (95% CI 0.516 – 0.793)] demonstrated superior performance to serum PSA or other individual biomarkers (Tables 4.4 & 4.5).

4.2.3.4 – Incorporating diagnostic potential of miRNAs in ejaculate

Integration of serum PSA, Hepsin and PCA3 results with miRNAs previously quantified by our group was performed as a preliminary feasibility investigation. Small RNA for this investigation was only available for 20 patients with biopsy- and RP-based classifications (11 with CaP). For biopsy-based classification, miR-200c and miR-375 showed best single marker performance in this group (Tables 4.3-4.5). Serum PSA, miR-200c and miR-125b were most discriminatory when combined as a composite score (Composite2) derived from logistic regression (Table A2.1, Figure 4.1B). This discrimination was demonstrated for CaP status based on biopsy [AUC 0.869 (95% CI 0.705 – 1.000)] and CaP risk determined by RP histology [AUC 0.869 (95% CI 0.644 – 0.976)] with a specificity estimate of 67% at 90% sensitivity (Tables 4.3-4.5). Furthermore, a consistent positive likelihood ratio of ≥ 2.7 was observed for Composite2 estimates across all classification methods (Tables 4.3-4.5).

4.2.4 Discussion

A pervading concern with any CaP diagnostic study is the true status of the biopsy negative (no CaP) patients. This issue also extends to cases where less dominant lesions are biopsied with the more significant tumours missed, in what is often multifocal disease within the prostate. Although this problem of inaccurate pathological status is likely to become less of an issue in the future with MRI referencing, MRI was not available for any of these patients. Thus, the best that could be done in this study was to ensure that the number of cores was 12 or more and that an adequate period had elapsed before analysing the data so that any clinical suspicion of a ‘missed tumour’ could be addressed. For this cohort, the monitoring period was a minimum of 13 months that extended to 88 months for the earliest donors. Forty-one patients had second and third TRUSbxs or TPBxs or subsequent transurethral resection of the prostate and, as a consequence, 18
men whose initial biopsies were negative had their tumour status changed to CaP following further biopsies.

Limitations with current diagnostic tests for CaP necessitate identification of new molecular biomarkers for prognosis and patient follow-up\textsuperscript{340}. An automated assay for PCA3, Progensa\textsuperscript{TM} PCA3, relying on urine collection after prostate massage was shown to have excellent clinical value based on repeat biopsies\textsuperscript{207}. Nevertheless, the detection rate was $<50\%$ in patients with PCA3 scores $>300$. In a large study of 1962 cases with PSA $>2.5\text{ng/ml}$, PCA3 ($\text{AUC} 0.706$) performed better than PSA ($\text{AUC} 0.569$) but when these two markers were combined AUC improved to 0.720\textsuperscript{341}. As many as 50\% of CaP patients possess a fusion between the androgen-regulated gene, TMPRSS2, and the ETS family member, TMPRSS2:ERG\textsuperscript{342}. While a correlation between ERG expression and clinical and pathological parameters has not been observed, specificity and recurrence of ERG in CaP suggests that it may be a useful adjunct diagnostic biomarker\textsuperscript{343}. In another study a strong link was recorded between ERG activation, young patient age and low grade cancer\textsuperscript{344}. Although a combination of PCA3 and \textit{phi} has recently been reported to only moderately enhance diagnostic power for CaP with first or repeated prostatic biopsies\textsuperscript{10}, combining TMPRSS2-ERG, PCA3 and \textit{phi} may be complementary, reflecting other predictors of cancer aggressiveness\textsuperscript{329}.

In this study we employed a combination of gene, miRNA and protein markers in an attempt to improve the efficiency of detection of CaP. The performance of this combined set of markers in ejaculate for the 66 selected patient specimens gave comparable results to those obtained with DRE-urine by other investigators but, unexpectedly, PCA3 alone was not more predictive than serum PSA for this patient cohort.

For the combination of serum PSA, PCA3 and Hepsin markers, findings for risk of significant CaP were comparable for D'Amico and PRIAS-based stratifications with \textit{AUC}s of 0.701 and 0.679, respectively, using logistic regression. Interestingly, these results were not bettered in the smaller RP group. There was a very strong concordance of biopsy and RP histology, consistent with repeated biopsying of at-risk patients, but RP patients were a selected subgroup with more aggressive tumour patients being directed to androgen deprivation therapy, most commonly in association with radiation therapy. This was further reflected in the performance of serum PSA for the current cohort (68\%), which better predicts CaP than commonly reported for screening, or targeted populations. However, the finding for the 20 patients' specimens that also had miR-200c and miR-375 measurements
is particularly notable, with these miRNAs providing best single marker performance. Furthermore, combining serum PSA with miR-200c and miR-125b improved significant CaP prediction, providing an AUC of 0.869 (95% CI 0.705 – 1.000) for a 90% sensitivity and 67% specificity, which was maintained when RP histology was incorporated (AUC 0.869; 95% CI 0.644 – 0.976) and an estimated likelihood ratio ≥ 2.7. This performance is non-inferior, and potentially improved, when compared to PCA3 in contemporary cohorts (< 34% specificity at 90% sensitivity) and on systematic review (36% specificity and positive likelihood ratio 1.4 at 90% sensitivity)\textsuperscript{345,346}. The robust nature of miRNAs in comparison with mRNAs makes them an attractive quarry for further investigation since their integrity is more likely to be retained in the enzymate-rich milieu that is ejaculate.

Although ideally the time interval between production and specimen processing in the laboratory should be only a few minutes, it was nominally up to 2 hours as there was reliance on patients delivering their specimens to the laboratory. This variable delay which had the potential for cellular and mRNA degradation may help to explain the 86 specimens which did not meet the stringent standards imposed for inclusion for analyses. On the other hand, cells present in ejaculate, a medium in which sperm cells are nurtured, are expected to be less susceptible to time-dependent changes in cell degradation and lysis than, for example, those forced into the urethra with DRE and passed in urine. In order to minimize potential confounding factors, such as cell lysis that may result in RNA degradation, and enzyme mediated changes in the metabolite and protein/peptide composition of the samples, a room on campus is now provided in which patients may produce specimens without delay, and adaptations have been made to the sample collection medium by adding tartrate, an inhibitor of endogenous enzymes (such as prostatic acid phosphatase), to the fluid in the collection jar. Once the reliability of diagnostic performance is established in this setting, sample collection protocols can be tailored for practical use as a screening test with incorporation of optimization methods (such as that presented in Section 3.2).

The extremely strict criteria used for our laboratory results caused the number of specimens to be significantly reduced but, for those fully complying with the requirements selected, the results are directly comparable with those of others who have reported on an expanded ‘PCA3 urine test’ to include the fusion gene TMPRSS2:ERG with or without the $ph$\textsuperscript{10,211} as well as other candidates such as EN2\textsuperscript{346}. Further, the cross-referencing of a subgroup of those CaP patient biopsy specimens with RP pathology illustrates the high level of fidelity for the findings from this study.
While others have focused elsewhere, we have continued to examine ejaculate in research into the early detection of CaP since, as the closest body fluid to the prostate itself, assaying is comparable with ‘in-vitro biopsying’ of the whole gland. Specifically, prostatic smooth muscle contracts globally with ejaculation producing liquid from all parts of the prostate, while DRE/prostatic massage (prior to urine voiding) is directed to the posterior part of the gland and so may miss anteriorly sited malignancies. Undeniably, there have been difficulties in obtaining specimens which may extend to the clinic, particularly for older men. This may not be the case for younger patients if willingness to provide ejaculate for subfertility management is any indication. A further factor of relevance in CaP diagnosis is life-expectancy, particularly in relation to cardiovascular disease with the onset of erectile dysfunction (potentially limiting ejaculate specimen production) increasingly recognized as a strong indicator of impending morbidity with significant mortality implications\textsuperscript{18,347}. Supporting this close relationship, we have found that ejaculate donors for our research studies who were subsequently diagnosed with CaP had statistically equivalent mortality expectations to the general population\textsuperscript{17}.

4.2.5 Conclusions

These results with ejaculate mirror those from other investigators with other body fluids in that they are not sufficiently discriminating to be recommended as a stand-alone test for identifying significant CaP or for use in triaging patients. However, they do provide a basis to which other markers can be added to address these requirements and contribute to the improved selection of patients for biopsies and treatment. We intend for this study’s findings to serve as a template for evaluating the addition of other candidate markers using the same specimens (and corresponding post-ejaculate urines) in our quest to produce the best selection of markers for profiling patients’ clinical status and tumours.
4.3 “Selection and monitoring of active surveillance candidates using NMR-based metabolomics of ejaculate: a preliminary investigation” (published manuscript)

The manuscript entitled, “Selection and monitoring of active surveillance candidates using NMR-based metabolomics of ejaculate: a preliminary investigation” has been accepted for publication by *Prostate International* (in press).

The contributions of the co-authors to this manuscript were as follows: Experiment design, patient recruitment, experiment execution, data processing, statistical analysis and clinical interpretation were performed by Matthew Roberts, the PhD candidate, under the supervision of Dr Horst Schirra and Prof RA “Frank” Gardiner. Renee Richards, Clement Chow and Marion Buck assisted with sample collection and biobank curation, experiment execution, data processing and data analysis. John Yaxley provided advice regarding clinical interpretation. Prof Martin Lavin provided critical manuscript revisions. Preparation of the manuscript and associated figures was performed by Matthew Roberts, the PhD candidate.

The manuscript text is presented as originally published by the journal. To maintain a consistent style between chapters in this thesis, abbreviations, placement of figures and tables within the text, as well as numbering of pages, figures and tables has been adjusted. The references have been collated in the section, ‘Bibliography’, with others contained in this thesis.
Selection and monitoring of active surveillance candidates using NMR-based metabolomics of ejaculate: a preliminary investigation

Matthew J. Roberts MBBS\textsuperscript{1,2,3,4}, Renee S. Richards PhD\textsuperscript{2}, Clement W. K. Chow BSc (Hons)\textsuperscript{2}, Marion Buck BSc\textsuperscript{5}, John Yaxley FRACS\textsuperscript{4}, Martin F. Lavin PhD\textsuperscript{2}, Horst Joachim Schirra Dr. sc. nat.\textsuperscript{3} \#, Robert A. Gardiner FRACS\textsuperscript{2,4,6} \#

\textsuperscript{1} The University of Queensland, School of Medicine, Brisbane, Qld 4006, Australia
\textsuperscript{2} The University of Queensland, Centre for Clinical Research, Brisbane, Qld 4006, Australia
\textsuperscript{3} The University of Queensland, Centre for Advanced Imaging, Brisbane, Qld 4072, Australia
\textsuperscript{4} Department of Urology, Royal Brisbane and Women’s Hospital, Brisbane, Qld 4006
\textsuperscript{5} Department of Environmental Health Sciences, University Medical Centre Freiburg, 79106 Freiburg, Germany
\textsuperscript{6} Edith Cowan University, Joondalup, Western Australia
4.3.1 Abstract

**Background:** Diagnosis and monitoring of localized prostate cancer requires discovery and validation of non-invasive biomarkers. NMR-based metabolomics of ejaculate reportedly improves diagnostic accuracy, but requires validation in a high-risk clinical cohort.

**Methods:** Ejaculate samples of 151 men being investigated for prostate cancer were analysed with $^1$H-NMR spectroscopy. After adjustment for buffer (Add-to-Subtract) and endogenous enzyme influence on metabolites, metabolite profiling was performed with multivariate statistical analysis (principal components analysis, partial least squares) and targeted quantitation.

**Results:** Ejaculate metabolites best predicted low- and intermediate-risk prostate cancer with differences observed between these groups and benign samples. Lipids/lipoproteins dominated spectra of high grade samples with less metabolite contributions. Overall prostate cancer prediction using previously described metabolites was not validated.

**Conclusions:** Despite being unable to validate previous reports and after validation of discriminatory reliability with optimized sample collection, metabolomics of ejaculate *in vitro* may assist diagnosis and monitoring of either low or intermediate grade prostate cancer, but offers less benefit in high-risk patients. Further investigation in active surveillance cohorts, and/or in combination with *in vivo* magnetic resonance spectroscopic imaging may further optimize localized prostate cancer outcomes.

**Keywords:** biomarker, metabolomics, Nuclear Magnetic Resonance, prostate cancer, seminal fluid, serum PSA
4.3.2 Introduction

Accurate CaP diagnosis to prolong life with minimal morbidity is a daily challenge for urologists. While early treatment of localized csCaP with curative intent reduces mortality and metastases\(^{348,349}\), harms associated with over detection and treatment of indolent CaP driven by injudicious use of serum PSA and prostate biopsy have reduced overall CaP detection\(^{8,9}\). Limitations of serum PSA have driven advancements in multiparametric magnetic resonance imaging and biomarkers in serum (e.g. Prostate Health Index) and urine (PCA3, TMPRSS2:ERG fusion gene)\(^{6,7}\). However, due to cost-effectiveness concerns, these are used as adjunctive tests rather than as standalone detection tests despite their improved diagnostic accuracy\(^{10,11}\).

Prostatic fluid, produced as ejaculate after physiological prostatic smooth muscle contraction, contains the clinical biomarkers PSA and PAP\(^{12,13}\). Malignant prostatic cells in ejaculates of men with CaP have been shown to express genes (PCA3, Hepsin) and microRNAs that improve detection compared to serum PSA\(^{14-16}\). Metabolomics is a modern biomarker approach that quantifies small metabolites\(^{19}\), most commonly using NMR spectroscopy or mass spectrometry\(^{20,21}\). NMR-based metabolomics is highly sensitive and reproducible with affordable sample-to-sample costs\(^{20}\). Ejaculate metabolite profiles improve PSA-based diagnosis\(^{22,23}\), but require clinical validation.

This study investigates the feasibility of ejaculate analysis using NMR-based metabolomics for the prediction of csCaP in a high-risk clinical cohort and compares metabolite profile CaP diagnosis against prostate biopsy and RP histology.

4.3.3 Subjects and Methods

Ethical approval was obtained from the University of Queensland Medical Research Ethics Committee (Project no. 2006000262) and the Royal Brisbane and Women’s Hospital Human Research Ethics Committee (HREC/09/QRBW/320, HREC/09/QRBW/305 and 1995/088B).

4.3.3.1 – Patients and Clinical data

Male patients (\(n=154\)) attending either the Royal Brisbane and Women’s Hospital Urology outpatient department or local private consulting rooms for investigation of elevated PSA and/or abnormal digital rectal examination between January 2007 and February 2013 were
enrolled in this prospective cohort study. Following informed consent, patients provided ejaculate specimens on site or at home prior to or at least one month after prostate biopsy.

Patient data collected included age, serum PSA and detailed prostate biopsy and RP histology records. Biopsy and RP specimens were reported by uropathologists according to the 2005 ISUP classification\textsuperscript{350}. Patients were monitored for biopsy progression, such as CaP detection following initial false negative biopsy or upgraded Gleason score with further biopsy or RP (n=61).

Risk stratification (low, intermediate, high risk) was performed according to the D'Amico criteria recommended in the American Urological Association Guidelines\textsuperscript{337} and used to determine csCaP presence (intermediate, high risk requiring treatment). Here, low risk included those not requiring treatment (negative biopsy or low D'Amico risk). Given established disparity between biopsy and RP histopathology, risk classification accuracy was optimised using whichever histopathology best described tumour characteristics.

4.3.3.2 – Specimen processing

Ejaculate specimens were deposited directly into sterile micro-urine jars containing 20 ml HBSS (Gibco) for the first 117 patients used initially for cytology and RNA analyses, which was thereafter replaced by PBS (in-house preparation) because glucose in HBSS (Figure 4.2) interfered with preliminary metabolomics analysis. All specimens were provided to the laboratory without cooling as soon as logistically possible by the patients and were processed in the laboratory within 2 h of production. Specimens were combined with 20 ml HBSS or PBS, layered over 10 ml isotonic Percoll (GE Healthcare-Pharmacia) and centrifuged at 974×g for 30-60 min at 4°C. Isolated supernatants, herein referred to as ejaculate, were snap-frozen on dry ice in 1 ml aliquots and stored at –80°C.

4.3.3.3 – Sample preparation

Ejaculate samples were thawed on wet ice and distributed in 100 µl aliquots. 80 µl of PBS solution were added along with 20 µL D\textsubscript{2}O as lock substance that contained DSS as internal chemical shift standard and DFTMP as internal pH indicator leading to final concentrations of DSS and DFTMP of 100 µM, resulting in 200 µL total sample volume. Samples were transferred to sterile 3 mm NMR tubes (Bruker Biospin, Rheinstetten, Germany).
4.3.3.4 – NMR spectroscopy

NMR spectra of ejaculate samples were measured on a Bruker Avance 900 spectrometer operating at a $^1$H frequency of 900.13 MHz (Bruker Biospin), equipped with a 5 mm self-shielded z-gradient triple resonance cryoprobe and SampleJet sample changer. 1D NOESY spectra were acquired at 298 K with the “noesypr1d” pulse sequence, accumulating 200 transients (following 8 dummy scans) at 32k data points with a spectral width of 14 ppm$^{150}$. The transmitter frequency was set to the water resonance, which was suppressed by continuous wave irradiation during the NOESY mixing time of 0.1 s and relaxation delay of 3.0 s. Tuning/matching, shimming, and data acquisition were performed automatically with the ICON-NMR interface for high-throughput automation. Samples were measured in one batch per sample collection buffer and ordered randomly within these batches.

4.3.3.5 – Spectral Processing

NMR spectra were processed in TopSpin 3.2 (Bruker Biospin). The free induction decays were baseline corrected by a Gaussian function (0.1 ppm filter width) for post-acquisition water deconvolution$^{315}$, followed by multiplication with an exponential window function (0.1 Hz line broadening), and Fourier transformation to 64k points. Subsequently, the spectra were manually phased, manually baseline corrected with a cubic spline curve, and referenced to DSS at 0.0 ppm. For all further data manipulation, the spectra were truncated to $\delta=10.0-0.25$ ppm, exported into MATLAB 2015b (The Mathworks Inc., Natick, Massachusetts, United States), and scaled according to the Bruker $\text{NC\_proc}$ parameter.

4.3.3.6 - “Add-to-Subtract” glucose exclusion

Preliminary analysis revealed glucose at sometimes dominant levels in most samples (Figure 4.2). As HBSS contains 1 g/L D-glucose and ejaculate volumes were varied, the exogenous glucose concentration and its influence on subsequent MVSA was unpredictable. Thus, we used the “Add-to-Subtract” method$^{351}$ to exclude glucose signals from the NMR spectra: First, we added 1 $\mu$l of 1 M D-glucose in PBS to each sample and repeated measurement with identical experimental parameters, leading to a total of 302 spectra for 151 patients (151 original, 151 with additional glucose). Secondly, using Topspin’s multiple display, we determined the corresponding scaling factor between spectrum 2 and 1 for each sample that ensures elimination of the glucose signal upon subtraction. Then the exported spectra 1 and 2 for each sample were aligned using
Figure 4.2: Sections of 1D NOESY spectra from ejaculate measured at 900 MHz. Panel A – 1D NOESY spectrum of ejaculate collected in HBSS, with additional peaks due to exogenous glucose present. Panel B – 1D NOESY spectrum of ejaculate collected in PBS. Metabolite abbreviations; Ala = alanine, Arg = arginine, Asn = asparagine, Cho = choline, Cit = citrate, DFTMP = 1,1-difluoro-1-trimethylsilyl methylphosphonic acid, DSS = 4,4-dimethyl-4-silapentane-1-sulfonic acid, Fru = fructose, Gln = glutamine, Gluc = glucose, Glu = glutamate, Ile = isoleucine, Lac = lactate, Leu = leucine, Lys = lysine, PCho = phosphocholine, Pyr = pyruvate, Spe = spermine, Val = valine.

“icoshift” on the glucose peaks at 3.37–3.44 ppm and then along 10 equal segments. Finally, for each sample spectrum 2 was scaled with the scaling factor recorded in Topspin and subtracted from spectrum 1. The resulting difference spectra were stored in a separate matrix.

4.3.3.7 – Spectral alignment and data reduction

The peaks of all difference spectra were aligned at full resolution using “icoshift”, initially on the lactate doublet at 1.32 ppm and subsequently on manually defined segments. No shifting artefacts were identified. Using an in-house MATLAB script, the aligned difference
spectra were data reduced to buckets of 0.01 ppm width over the range 10.0–5.08 and 4.52–0.25 ppm, excluding the water signal region.

4.3.3.8 – Multivariate statistical analysis

Metabolite data ($\mathbf{X}$) matrices containing original and difference (Add-to-Subtract) data were quantile normalized with the “affy” package\textsuperscript{352} in R version 3.2.2\textsuperscript{139} and imported into SIMCA P+ 12.0 (Umetrics, Umeå, Sweden) for MVSA together with clinical data variables ($\mathbf{Y}$-matrix). $\mathbf{X}$-matrices were Pareto-scaled before unsupervised PCA\textsuperscript{151}. To determine which metabolite signatures were associated with clinical data (cancer/risk status; $\mathbf{Y}$-matrix), supervised PLS was performed\textsuperscript{151}. Multivariate model quality was judged by the $R^2$ (“goodness of fit”) and $Q^2$ (“goodness of prediction”) figures of merit (Table A2.2). PLS models were validated by 200-fold response permutation. Traditional statistical analysis, including the non-parametric Mann-Whitney $U$ test, logistic regression and Receiver Operating Characteristic analysis were performed in MedCalc 12.7 for Windows (MedCalc Software; Ostend, Belgium).

4.3.3.9 – Targeted metabolite profiling

Ejaculate metabolites were quantified using Chenomx NMR suite\textsuperscript{353} using DSS as internal concentration standard. Logistic regression on the ejaculate metabolite concentrations was performed in MedCalc, similar to that described in\textsuperscript{23}.

4.3.4 Results

4.3.4.1 – Clinical cohort demographics

From 151 patients who provided ejaculate samples, 80 were initially diagnosed with CaP and an additional 18 patients diagnosed during the follow-up period. Within these 98 patients, 82 met csCaP criteria. 61 patients underwent RP for localized CaP in which 59 were determined to be high risk per the D'Amico criteria, with 6 upgraded from negative/low risk. Primary Gleason pattern 4 or higher or tertiary pattern 5 was present in 34 patients based on RP histology. The demographic information (Table 4.6) demonstrated that serum PSA was higher in those with high D'Amico risk.

4.3.4.2 – Unsupervised multivariate statistical analysis

The ejaculate samples were analysed with $^1$H NMR spectroscopy. 1D NOESY spectra were measured, aligned, and data reduced to 0.01 buckets. For initial PCA, buckets
corresponding to ethanol, resulting from sample preparation, were excluded, as were spectra that were outliers due to broad resonances (n=2). PCA yielded a model (Table A2.2 M1) with 6 PCs, in which samples clustered per the buffer solution used (PC1/PC2; Figure 4.3a-c), with higher glucose levels in samples prepared in HBSS. In higher PCs, sample variation was observed due to inter-sample differences of lipids/lipoproteins, phosphocholine, choline and citrate, as well as spermine (data not shown), which were unrelated to CaP in this analysis.

The “Add-to-Subtract” method\textsuperscript{351} was used to remove glucose signals from NMR spectra. Following measurement of a “baseline spectrum” (spectrum 1), glucose was added in high concentration to the sample in the same NMR tube and a second spectrum was measured (spectrum 2). Spectrum 2 was subtracted from spectrum 1 with an appropriate scaling factor to remove glucose signals but preserve signals of compounds in the resulting difference spectrum. The method assumes that introduction of glucose does not change sample conditions, preserving sample matrix, line shapes and signal frequencies.

<table>
<thead>
<tr>
<th>Biopsy</th>
<th>Age (years)</th>
<th>Serum PSA (ng/ml)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>61 (55 – 66)</td>
<td>6.5 (4.33 – 9.23)</td>
</tr>
<tr>
<td>CaP status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive (n = 98)</td>
<td>60.5 (55 – 65)</td>
<td>6.4 (4.5 – 11)</td>
</tr>
<tr>
<td>NS</td>
<td>62 (55.75 – 68.25)\textsuperscript{NS}</td>
<td>6.5 (3.6 – 7.95)\textsuperscript{NS}</td>
</tr>
<tr>
<td>Negative (n = 53)</td>
<td>61 (55 – 66)</td>
<td>6.75 (4.5 – 11.9)</td>
</tr>
<tr>
<td>NS</td>
<td>62 (55.75 – 68.25)\textsuperscript{NS}</td>
<td>6.5 (3.6 – 7.95)\textsuperscript{NS}</td>
</tr>
<tr>
<td>D’Amico risk</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High (n = 82)</td>
<td>61 (55 – 66)</td>
<td>6.0 (3.6 – 8.13)\textsuperscript{*}</td>
</tr>
<tr>
<td>Negative/Low (n = 69)</td>
<td>61 (55 – 67)\textsuperscript{NS}</td>
<td>6.0 (3.6 – 8.13)\textsuperscript{*}</td>
</tr>
<tr>
<td>RP</td>
<td>n = 61</td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>57 (54 – 64)</td>
<td>6.2 (4.13 – 9.4)</td>
</tr>
<tr>
<td>RP – Primary/tertiary pattern</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gleason ≥ 4 (n = 34)</td>
<td>57.5 (55 – 64)</td>
<td>7.6 (4.9 – 13)</td>
</tr>
<tr>
<td>NS</td>
<td>57 (51.25 – 63.5)\textsuperscript{NS}</td>
<td>5.4 (3.23 – 6.83)\textsuperscript{**}</td>
</tr>
<tr>
<td>Gleason 3 (n = 27)</td>
<td>57 (51.25 – 63.5)\textsuperscript{NS}</td>
<td>5.4 (3.23 – 6.83)\textsuperscript{**}</td>
</tr>
</tbody>
</table>

Table 4.6: Demographic information for patients based on biopsy and radical prostatectomy (RP) histology. Median and interquartile range are shown for age and serum PSA. All comparisons were made using the Mann-Whitney U-test (two-tailed). * p < 0.05; ** p < 0.01; NS = not significant.
Figure 4.3: Principal Components Analysis of ejaculate NMR spectra from men being investigated for prostate cancer (*n* = 151), prepared with different buffer solutions (Hanks Balanced Salt Solution: filled square, phosphate buffered saline: empty square). Panels a-c: Initial sample clustering is observed due to the difference in buffer solutions and resulting sample glucose content (principal component 1) and inter-sample differences in metabolite (citrate, choline, lipids/lipoproteins and phosphocholine) variation (principal component 2). Panels d-f: After Add-to-Subtract elimination of glucose, the previously observed effects of different buffer solutions are no longer apparent (panel a). No clustering was present according to CaP status (blue squares = benign; red triangles = CaP). Panels a, b, d, e: scores plots; panels c, f: loadings plots. Metabolite abbreviations: Cho = choline; Cit = citrate; Gluc = glucose; Lip = lipids/lipoproteins; PCho = phosphorylcholine.
PCA of the difference spectra (Figure 4.3d-f, Table A2.2 M2) showed no sample grouping due to differences in buffer used (Figure 4.3d). The predominant drivers of sample variation were lipids/lipoproteins (PC1), an inverse relationship between choline and phosphocholine as well as citrate. An association with csCaP was suggested by the presence of lipids/lipoproteins, although separation between clinical groups was not observed in any PC.

Given that the inverse relationship observed between phosphocholine and choline is due to PAP-mediated hydrolysis, a reaction which was not inhibited in these samples\textsuperscript{354}, choline-based metabolites (choline, phosphocholine and glycerophosphocholine) were excluded to remove their effect of unbalanced regulation on the MVSA. However, PCA (Figure 4.4, Table A2.2 M3) showed no obvious clustering, with most variation due to lipids/lipoproteins, citrate and serine (Figure 4.4a,b). Fructose and spermine were other significant sources of variation in PC3/PC4 (Figure 4.4c,d).

4.3.4.3 – *Supervised multivariate statistical analysis*

In the unsupervised PCAs, which determine sources of variation potentially independent of underlying biology, no sample clustering into clinical groups was seen, prompting the need for supervised MVSA. First, the presence of csCaP according to the D’Amico criteria based on biopsy was used as the predictive variable in PLS analysis (Figure 4.5, Table A2.2 M4) and demonstrated lipids/lipoproteins to be associated with variation for csCaP, which were mostly limited high-risk patients. Furthermore, there was potential subgrouping among the D’Amico risk groups (Figure 4.5b).

Based on these results and reports that maximal metabolite disturbances are observed in low- and intermediate-risk tumours, we analysed with PLS a subgroup of 11 samples correlating to these grades confirmed by RP histology only (Figure 4.6a-c, Table A2.2 M5). The single low-risk sample was separated from the intermediate-risk samples due to reduced lactate, pyruvate and lipids/lipoproteins and increased citrate, myo-inositol, spermine and fructose (Figure 4.6a,c). Within these low/intermediate-risk samples, separation was seen in accordance with primary Gleason pattern 4, associated with higher levels of lipids/lipoproteins, lactate and pyruvate as well as lower levels of citrate, spermine and myo-inositol (Figure 4.6b,c). These relationships were observed when classifiers based on all low/intermediate-risk patients, determined by biopsy or RP, were performed (Figure 4.6d,e; Table A2.2 M6). When benign samples were considered with risk group
Figure 4.4: Principal Components Analysis after exclusion of choline containing metabolites demonstrated that lipids/lipoproteins, citrate and serine were influential metabolites (panels a, b) as well as fructose and spermine (panels c, d). No clustering was present according to CaP status (blue squares = benign; red triangles = CaP). Panels a, c: scores plots; panels b, d: loadings plots. Metabolite abbreviations: Cit = citrate; Fru = fructose; Lip = lipids/lipoproteins Ser = serine; Spe = spermine.

Analysis of only the samples collected in PBS, unaffected by any external glucose (Figure A2.5, Table A2.2 M11-17), showed similar relationships to those seen for the full cohort. Specifically, valid models were obtained for separation between low- and intermediate-risk samples (Table A2.2 M12, Figure A2.5c,d; limited by sample size) and low-risk and benign samples (Table A2.2 M13, Figure A2.5e,f). Findings were confirmed with PCA (Table A2.2 M15-17, Figure A2.6) and driven by lactate levels (Figure A2.7). The presence of the TMPRSS2:ERG fusion gene, detected in the epithelial cell fraction of ejaculate, used as Y variable was weakly but non-predictively associated with lipid/macromolecule resonances (Figure A2.8, Table A2.2 M18).
Figure 4.5: Supervised, partial least squares analysis of ejaculate NMR spectra in predicting csCaP (according to D’Amico criteria) following Add-to-Subtract. Minimal separation is seen according to csCaP (blue squares = benign, red triangles = csCaP; panel a). When coloured according to risk subgroups (blue squares = benign; green dots = negative/low risk; yellow diamonds = intermediate risk; red triangles = high risk/cancer present; panel b), potential intragroup clustering was seen due to pyruvate, serine and lipids/lipoproteins (high/intermediate risk) and TMAO (negative/low risk).

Panels a, b: scores plots; panel c: loadings plot. Metabolite abbreviations: Lip = lipids/lipoproteins; Pyr = pyruvate; Ser = serine; TMAO = trimethylamine N-oxide
Figure 4.6: Supervised, partial least squares analysis of ejaculate NMR spectra in predicting CaP risk (according to D’Amico criteria) following Add-to-Subtract. Separation between low- (green dots) and intermediate- (yellow diamonds) risk patients based on RP histology due to elevated lactate, lipids/lipoproteins and pyruvate and reduced citrate, fructose, myo-inositol and spermine in intermediate-risk samples (panels a,c). These relationships were observed when expanded to all low and intermediate risk patients (panels d, e). Discrimination within the low/intermediate group was observed due to primary Gleason pattern 4 (filled square = present; empty square = absent) and higher lipids/lipoproteins, lactate and pyruvate as well as reduced citrate, myo-inositol and spermine (panels b,c). Panels a, b, d: scores plots; panels c, e: loadings plots. Metabolite abbreviations: Cit = citrate; Fru = fructose; Lac = lactate; Lip = lipids/lipoproteins; Myo = myo-inositol; Pyr = pyruvate; Spe = spermine.
4.3.4.4 – Targeted metabolite profiling

Ejaculate metabolite quantification with subsequent logistic regression showed that citrate or myo-inositol were not significant predictors of CaP status (Table A2.3). Significant metabolites for CaP status (choline, leucine) and csCaP (leucine, valine) did not significantly improve diagnosis compared with serum PSA metabolite predictability.

4.3.5 Discussion

In this paper, we present the largest validation study of ejaculate-based metabolite prediction of CaP using high resolution NMR spectroscopy, having analysed ejaculate metabolite profiles from 151 men being investigated for CaP. Undue influence of exogenous glucose contained in the HBSS buffer used for RNA analyses was successfully excluded by applying Add-to-Subtract and revealed inherent variation due to enzyme-dependent changes in choline-based metabolites. Ejaculate metabolites best predicted low- and intermediate-risk CaP with differences observed between these groups and benign and high-risk samples. Metabolites previously reported to determine CaP, such as citrate, spermine and myo-inositol, showed minimal predictive ability in this clinically applicable cohort. These findings were confirmed with targeted metabolite quantification.

Well described prostatic metabolite changes due to CaP, specifically reduced citrate and polyamines (e.g. myo-inositol, spermine) and increased intracellular lactate, choline and creatine\(^20\), were not predictive in this study due to underlying the following clinical and biological factors: Clinically, the study population presented here contains patients suspected of harbouring CaP, encountered in daily urological practice (Table 4.6). In earlier reports where ejaculate metabolites significantly improved CaP detection, CaP-positive samples were compared with healthy controls or men unlikely to have CaP, suggested by marked discrepancies in serum PSA between groups. Our population contained heterogeneous disease states, inclusive of all tumour grades with predominance toward high-risk CaP. While group separation was observed between CaP risk groups (Figure 4.6, Figures A2.5, 2.6), we could not truly exclude CaP in patients with a negative biopsy due to limitations in biopsy-based CaP detection and known metabolic changes in early tumorigenesis, which may lead to confounding overlap between groups and invalid statistical models. Given the studied population of 151 men was almost double that reported by Serkova and colleagues (n=78), statistical power was considered to be sufficient. To exclude uncertainty among the control group, a sub-analysis of the presence or absence of Gleason pattern 4 on RP histology showed overlap of groups (M10, Figure
A2.4g,h), likely owing to reduced metabolite influence in poorly differentiated tumours. Similarly, limitations of biopsy-based risk stratification given known upstaging at RP in up to 40% of patients may confound the accuracy of risk subgroup analyses. When analyses based only on RP-based diagnosis were expanded to include biopsy-based diagnosis to increase sample size, sample grouping was less obvious despite similar metabolite patterns being observed in the loadings plot (Figure 4.6). Thus, a larger low-/intermediate-risk RP cohort would be expected to accurately “up-classify” (upstage) low-risk samples with metabolite patterns similar to intermediate-/high-risk samples, as shown elsewhere.

**Biologically**, Gene expression and metabolite alterations occur early in tumorigenesis and are more pronounced in lower grade (Gleason≤7) compared with higher grade (Gleason≥8) tumours, supported by our analysis of low- and intermediate-risk patients (Figure 4.6, Figures A2.5, 2.6). In addition to direct metabolic changes, the inverse relationship between lactate and fructose resulting in group separation between low- and intermediate-risk and benign samples may indicate disturbed ejaculate homeostasis of anions (zinc) or enzymes (PSA, PAP) known to improve sperm function, resulting in impaired sperm glycolysis. Indeed, poor discrimination of metabolite profiles from high-risk tumours was demonstrated here and in other studies, likely due to accumulated genetic alterations with disease progression. Thus, patients with lower grade tumours may be amenable to ejaculate-mediated in vitro or MRSI-mediated in vivo assessment or monitoring as a potential substitute for repeat biopsy in active surveillance.

Altered metabolite homeostasis correlates with increased fatty acid synthesis, due to or in association with TMPRSS2-ERG fusion gene translocation associated with aggressive CaP, may account for the overwhelming influence of lipids/lipoproteins in high-risk patients in this study, similar to that reported by others. Higher grade tumours overexpress the oncogene MYC, which is associated with dysregulated lipid metabolism and display altered cholesterol metabolism to increase energy storage. Upregulated lipid subclasses have been described between normal, localized and metastatic prostatic cells, with choline kinase α implicated in *de novo* lipogenesis in aggressive metastatic cells. Systemically, lipid and energy metabolites in serum have been strongly associated with aggressive CaP and may improve CaP detection with 97% accuracy.
HBSS required significant correction using Add-to-Subtract, which did not introduce further influence into the MVSA. Subsequently, the uninhibited changes in choline-based metabolites showed significant influence in the preliminary MVSA. These metabolite peaks were excluded because PAP-catalysed hydrolysis of phosphocholine to choline is a rapid, endogenous reaction to enhance spermatozoal function and protection\textsuperscript{354}. Variations in time from sample production to processing, despite most being done within 2 hours, are likely to cause significant variation among these metabolites independent of underlying CaP due to the unknown degree of reaction completion. Given the postulated role of choline in tumour progression, as indicated by elevated \textit{in vivo} levels, reliable quantification of choline-based metabolites in ejaculate is desired. Thus, a sample collection/storage protocol should be implemented that limits the PAP reaction to 2-3% progression, such as our recommendation that ejaculate samples be collected in a sterile urine jar containing 5 mM tartrate in 20 ml PBS solution cooled to 277 K\textsuperscript{354}. While malignant prostatic metabolite contribution to ejaculate, considering concurrent contributions from multiple organs and resulting proteolysis, may intuitively be minimal or diluted, our findings are like those seen in tissue extracts and \textit{in vivo}\textsuperscript{279,280}, likely to be enhanced by spectral acquisition at 900 MHz.

In conclusion, this validation study was unable to replicate previous performance of ejaculate-based metabolite prediction of CaP in 151 men being investigated for CaP. Multivariate analysis and targeted metabolite profiling were used. Corroborating other studies, grouping was observed when comparing low- and intermediate-risk patients, while lipids/lipoproteins dominated spectra of high grade samples with fewer contributions from other metabolites. Despite being unable to validate previous reports on univariate metabolite analysis for CaP discrimination, dedicated metabolomics protocols ideally in serial collections may maximize information recovery and allow validation of discrimination and reliability of ejaculate metabolomics for routine use. The value of metabolomics analysis of ejaculate for CaP currently appears to be in active surveillance of low- or intermediate-grade tumours suspicious of under-staging, in which \textit{in vivo} correlation with MRSI and monitoring \textit{in vitro} with ejaculate or \textit{in vivo} with MRSI may further clinical practice.
4.4 Preceding analyses for ejaculate metabolomics experiments

4.4.1 – Exogenous glucose adjustment in NMR spectra

The submitted manuscript as Chapter 4.3 of this Thesis was the final product after several confounding variables were encountered. Initial ejaculate metabolomic analysis was performed for 111 samples using a AV500 500MHz Bruker NMR spectrometer with standard spectral processing, data reduction and integral normalisation. Outlying samples were detected due to the presence of glucose which, following exclusion of glucose-containing buckets, the MVSA was mostly influenced by choline and shifting metabolite signals, indicated by neighbouring buckets present at either end of the loadings plot (buckets and dominated the MVSA, as discussed in Chapter 4.3. While correction for shifting peaks was overcome using icoshift, the loss of information regarding metabolites contained in excluded regions (e.g. myo-inositol) prompted other glucose adjustment methods to be trialled.

First, the buffer solution (HBSS) was measured using the same spectrometer and experimental variables compared with the initial measurements. Indeed glucose was a dominant feature of the HBSS spectrum. After the HBSS was appropriately scaled and subtracted from each spectrum using the method described in Chapter 4.3 (Figure 4.7A), baseline distortion artefacts resulted in improper influence of glucose buckets on the MVSA (Figure 4.7B. Furthermore, measurement of additional samples collected in PBS contained exogenous ethanol, also providing influence on OPLS MVSA for CaP presence (Figure 4.7C-D; n=147, k=919, $R^2_Y = 0.073; Q^2_Y = -0.057$).

Secondly, following review of the manuscript by Ye and colleagues, initial Add-to-Subtract using an average glucose spectrum across all samples was attempted but resulted in significant baseline disturbances, even with use of icoshift (data not shown, similar to Figure 4.7a-b).

Finally, the best adjustment to be made for the exogenous glucose in samples from use of HBSS was remeasuring all samples (n = 151) using Add-to-Subtract, resulting in 302 measured NMR spectra, as described in Chapter 4.3. Given the requirement for re-measurement and availability of automated sample changing at our facility, this was opportunistically performed at higher field strength (900 MHz).
Figure 4.7: Attempted adaptations to Add-to-Subtract to improve ejaculate metabolomics analyses. A – 1D NOESY spectral overlay of ejaculate (orange) and HBSS (blue) spectra, demonstrating the high glucose content of HBSS and minor peak offsets between spectra. B – 1D NOESY spectral overlay in the region of 3.17 – 3.27 ppm near choline-containing metabolites (Cho, choline; PCho, phosphocholine; GPCho, glycerophosphocholine) demonstrating baseline distortion artefacts near the glucose (Glu) peaks following Add-to-Subtract due to improper spectral alignment. C – Scores plot showing minimal separation due to CaP (red) and benign (blue) status. D – Loadings plot showing significant influence of glucose (Glu) despite subtraction, as well as exogenous ethanol (Eth) and choline (Cho).

The resulting spectra demonstrated baseline signal dispersion near the water signal, which required manual correction for all 302 spectra using cubic spline correction. Initial glucose subtraction was performed in TopSpin according to the published method, however this resulted in significant baseline disturbances (similar to that seen in Figure 4.7B).

To correct for this, as described in the manuscript, individual spectra were aligned on glucose and then in equal segments prior to subtraction.

4.4.2 – Unbalanced choline regulation in ejaculate samples

As demonstrated initially and in the final ejaculate metabolomics analysis, choline-based metabolites were additional sources of potential CaP-independent variation within the
MVSA (Figure 4.3). Given the known endogenous hydrolysis of phosphocholine to choline mediated by PAP, these findings prompted the investigation and sample optimisation for metabolomics studies, as presented in Chapter 3.2. While the final analysis resulted in exclusion of choline-containing buckets, many adjustment methods were attempted prior to this.

Initial addition of choline-based metabolite (choline, phosphocholine, glycerophosphocholine) buckets was overly influential on the MVSA (Appendix 2 Figure A2.9; Figure 4.7C-D; n=154, k=879, $R^2X = 0.877; Q^2 = 0.367$).

Given that multiple reports implicate choline in CaP in vivo using MRSI, preservation of choline-based metabolite information was desired. A manuscript describing data normalisation in the presence of unbalance regulation provided a potential solution, where the spectra are scaled based on metabolites with least variance (based on Shapiro-Wilk testing). The Add-to-Subtract 0.01 ppm bucket width matrix was not normally distributed (Shapiro-Wilk $p<0.0001$). When the matrix was subject to resamp_mswsd(y), there was no progression away towards normal distribution with varying percentage of features used. The resulting graphs varied from that in the manuscript, with “features used” in the manuscript different to “percentage of features” in the output script (Figure 4.8).

![Figure 4.8](image)

**Figure 4.8:** Feature-mswsd box plots demonstrating data spread with variable feature use. When this method was tried with ejaculate data (left panel), the outcome was uninterpretable compared with published results (right panel).

Following scaling, various normalisation methods were available, including Variance Stabilisation Normalisation and PQN. The leading author was contacted regarding this discrepancy in the published manuscript and alterations are pending.
Chapter 5 – Post-ejaculate urethral washings as a novel biofluid for prostate cancer biomarkers

5.1 Synopsis

The aim of this chapter is to investigate if a novel prostatic biofluid, PEUW contains CaP biomarkers and their diagnostic performance and feasibility for clinical utilisation.

The presence of prostatic RNA biomarkers and their diagnostic performance in the epithelial cell fraction of PEUW was investigated and presented as a peer-reviewed manuscript (section 5.2)

PEUW was then analysed using NMR spectroscopy with three objectives. Initially, to determine if prostatic metabolite biomarkers are present, both in clinical samples and in a dedicated cohort comparing with mid-stream urine. Secondly, to describe the kinetic behaviour of these markers and if they behave similarly to that observed in ejaculate. Finally, metabolite profiles in PEUW collected from a clinical cohort were analysed to detect and characterise CaP. The preliminary findings of these NMR-based metabolomics experiments are presented in section 5.3.
5.2 “Prostate-based biofluids for the detection of prostate cancer: a comparative study of the diagnostic performance of cell-sourced RNA biomarkers.” (published manuscript)


Since publication, the manuscript has been received 1 and 2 citations for Scopus and Google Scholar, respectively (as at 10/02/17).

The contributions of the co-authors to this manuscript were as follows: Experiment design, data processing and statistical analysis drafting was performed by Matthew Roberts, the PhD candidate, under the supervision of Dr Horst Schirra and Prof RA “Frank” Gardiner. Renee Richards, Clement Chow, Marion Buck and Luke Selth assisted with sample collection, experiment execution and data processing. Suhail Doi provided guidance for statistical analysis and data interpretation. Hema Samaratunga, Joanna Perry-Keene, Diane Payton and John Yaxley provided advice on histological and clinical interpretation. Prof Martin Lavin provided critical manuscript revisions. Preparation of the manuscript and associated figures was performed by Matthew Roberts, the PhD candidate.

The manuscript text is presented as originally published by the journal. To maintain a consistent style between chapters in this thesis, abbreviations, placement of figures and tables within the text, as well as numbering of pages, figures and tables has been adjusted. The references have been collated in the section, ‘Bibliography’, with others contained in this thesis.
Prostate-based biofluids for the detection of prostate cancer: a comparative study of the diagnostic performance of cell-sourced RNA biomarkers

Matthew J. Roberts¹, ², ³, ⁴, Renee Richards², Clement W. K. Chow², Suhail A. R. Doi⁵, ⁶, ⁷, Horst Joachim Schirra³, Marion Buck⁸, Hema Samaratunga⁹, Joanna Perry-Keene¹⁰, Diane Payton¹⁰, John Yaxley⁴, Martin F. Lavin², Robert A. Gardiner², ⁴, ¹¹#

¹ The University of Queensland, School of Medicine, Brisbane, Qld 4006, Australia
² The University of Queensland, Centre for Clinical Research, Brisbane, Qld 4006, Australia
³ The University of Queensland, Centre for Advanced Imaging, Brisbane, Qld 4006, Australia
⁴ Department of Urology, Royal Brisbane and Women’s Hospital, Brisbane, Qld 4006
⁵ Research School of Population Health, The Australian National University, Canberra, Australia
⁶ School of Agricultural, Computational and Environmental Sciences, University of Southern Queensland, Toowoomba, Australia
⁷ College of Medicine, Qatar University, Doha, Qatar
⁸ Department of Environmental Health Sciences, University Medical Centre Freiburg, 79106 Freiburg, Germany
⁹ Aquesta Pathology, Brisbane, Qld 4066, Australia
¹⁰ Anatomical Pathology, Pathology Queensland, Brisbane, Qld 4006, Australia
¹¹ Edith Cowan University, Joondalup, Western Australia
5.2.1 Abstract

**Background:** Prostate cancer diagnosis requires improvement with the aid of more accurate biomarkers. PEUW could be a physiological equivalent to urine obtained following rectal prostatic massage, the current basis for the PCA3 test. The aim of this study was to investigate if PEUW contained prostate-based material, evidenced by presence of PSA, and to evaluate the diagnostic performance of PEUW-based biomarkers.

**Methods:** Male patients referred for elevated serum PSA or abnormal digital rectal examination provided ejaculate and PEUW samples. PSA, PCA3 and β2M were quantified in ejaculate and PEUW and compared with absolute and clinically significant (according to D’Amico criteria) CaP presence, as determined by biopsies. Diagnostic performance was determined and compared with serum PSA using ROC analysis.

**Results:** From 83 who provided PEUW samples, paired analysis with ejaculate samples was possible for 38 patients, while analysis in an unpaired, extended cohort was possible for 62 patients. PSA and PCA3 were detected in PEUW, normalized to β2M, and PCA3:PSA was calculated. In predicting absolute CaP status, PCA3:β2M in ejaculate (AUC 0.717) and PEUW (AUC 0.569) were insignificantly better than PCA3:PSA (AUC 0.668 and 0.431, respectively) and comparable with serum PSA (AUC 0.617) with similar trends observed for the extended cohort. When considering clinically significant CaP presence, serum PSA in the comparison (AUC 0.640) and extended cohorts (AUC 0.665) was comparable with PCA3: β2M (AUC 0.667) and PCA3:PSA (AUC 0.605) in ejaculate, with lower estimates for PEUW in the comparison (PCA3: β2M AUC 0.496; PCA3:PSA AUC 0.342) and extended (PCA3: β2M AUC 0.497; PCA3:PSA AUC 0.469) cohorts. The statistical analysis was limited by sample size.

**Conclusion:** PEUW contains prostatic material, but has limited diagnostic accuracy when considering cell-derived DNA analysis. PCA3-based markers in ejaculate are comparable to serum PSA and DRE-urine markers.

**Keywords:** biomarker, ejaculate, mRNA, PCA3, prostate cancer, PSA
5.2.2 Introduction

The detection of CaP is fraught with difficulties that include limitations of currently available biomarkers, access to imaging and tissue sampling. Total serum PSA is currently the single most widely used marker clinically for identifying men at risk of CaP, but it is a non-specific indicator of prostatic pathology that includes CaP among others so that population and opportunistic screening is discouraged due to over-investigation and over-detection resulting in overtreatment\(^5\). Contemporary CaP detection approaches have included, amongst others, advocacy for a biomarker panel, \textit{phi} and the 4-kallikrein protein test\(^{193,359}\).

Although imaging modalities, such as multiparametric MRI and even PSMA-PET, are being integrated into the detection strategy for triaging patients with an elevated PSA and may improve detection of clinically significant CaP\(^{360}\), there are limitations that detract from their widespread use. The false negative rate (approximately 15-20\%) for multiparametric MRI suggests a significant proportion of clinically significant tumours may be missed\(^{360}\), which is also observed with PSMA-PET imaging of tumours that do not express PSMA\(^{361}\). That imaging is establishing a niche in detection strategies is undeniable however it is the cost of such imaging methods that really limits application to mainstream clinical practice. Improved patient selection for imaging with accurate biomarkers is likely to optimize their practical application clinically.

Prostate-specific biofluids are an ongoing source for investigation using new analytical platforms\(^12\). PCA3, a long non-coding RNA, collected in the first void of urine following a vigorous DRE or prostatic massage, has been examined over the past 20 years\(^{362}\), with studies reporting to improve detection in men undergoing repeat biopsy \(^{363}\), but its role clinically remains uncertain. Modifications of PCA3\(^{364}\), as well as a combination with TMPRSS2-ERG fusion gene have been described to improve detection\(^{365}\) but have not been accepted as a useful addition in routine patient testing\(^{366}\). Other approaches utilizing exosomes, proteomics and metabolomics have the potential to improve early diagnosis of localized disease\(^{20,21,367}\). Indeed molecular and metabolomic markers in ejaculate have been reported to improve diagnosis compared to serum PSA\(^{15,16,335}\). Paralleling the post-massage urine concept, collection of urine following ejaculation, or PEUW potentially represents a new source of prostate-specific biomarkers for CaP detection and characterization, providing a physiologically produced alternative to serum or digital prostatic massage effluent for further biomarker discovery and development.
The aim of this study was to investigate the diagnostic potential of selected molecular markers in PEUW in detecting CaP and comparing diagnostic accuracy with the same markers in ejaculate, as well as serum PSA. The hypothesis tested was that the diagnostic performance of molecular markers in post-ejaculate urine would be comparable with those in ejaculate or serum PSA.

5.2.3 Subjects and Methods

5.2.3.1 Patients

A clinical cohort of men (n = 83) being investigated for CaP on the basis of an abnormal DRE and/or elevated serum PSA provided specimens of ejaculate and PEUW into sterile micro-urine jars, containing 20 ml Hanks Balanced Salt Solution (Gibco) and empty urine jars, respectively, between January 2007 and December 2009. As previously reported\textsuperscript{15}, all specimens were processed within 2 h of production after being delivered to the hospital campus without cooling. All specimens were collected prior to or at least one month following TRUSBx or TPBx.

Ethical approval to conduct this study was obtained from the University of Queensland Medical Research Ethics Committee (Project no. 2006000262) and the RBWH Human Research Ethics Committee (HREC/09/QRBW/320, HREC/09/QRBW/305 together with 1995/088B).

5.2.3.2 Clinical data

All patient data was prospectively collected following recruitment and included clinical details such as age, family history and serum PSA. Initial and updated TRUSbx/TPBx/RP histology specimens were reviewed by DP, JP-K and MLTHS and reported according to the 2005 ISUP classification\textsuperscript{350}, including standard biopsy (number of cores taken, number and percentage of positive cores, Gleason score) and RP (gland size, Gleason score, pathological stage, extracapsular status and margin status) parameters.

In order to identify the patients for whom active treatment would be recommended, risk stratification for biopsies in determining clinically significant CaP presence was performed using the D’Amico criteria recommended in the American Urological Association Guidelines\textsuperscript{337}. The clinically significant CaP category included patients defined as intermediate- and high-risk according to the D’Amico criteria, while the absence of clinically significant CaP was defined as negative/low risk patients according to the...
D'Amico criteria or those without CaP\textsuperscript{337}. The most accurate classification of clinically significant CaP, based on histopathology from TRUSbx, TPBx and/or RP, was used given established disparity between TRUSbx and RP histopathology. As previously reported, patients were subsequently placed in one of two clinical groups based on classification scheme used. Because of the imprecise nature of TRUSbx especially, patient follow-up was pursued for up to 7 years to ensure that those designated as negative for prostate cancer really were negative.

5.2.3.3 – Specimen processing and cellular isolation

Ejaculate specimens combined with 20 ml Hanks Balanced Salt Solution were layered over 10 ml isotonic Percoll (GE Healthcare-Pharmacia) and centrifuged at 974×g for 30-60 min at 4°C, with supernatants subsequently collected in 1 ml aliquots, snap-frozen on dry ice and stored at -80°C. The epithelial cell layer at the Percoll interface, present as a discrete band suspended between supernatant above and sperm and non-cellular components below, was then pipetted, washed with 25 ml PBS or Hanks and centrifuged at 1258×g for 10 min at 4°C.

5.2.3.4 – RNA preparation

Total RNA from collected cells was isolated using TRIzol reagent (Invitrogen) then subjected to on-column DNase treatment and clean-up with the RNeasy kit (Qiagen). Low yield samples were amplified using the SenseAmp kit (Gensisphere).

5.2.3.5 – cDNA synthesis and quantitative reverse transcription PCR

Quantitative reverse transcription PCR (qPCR) was undertaken using the QIAGEN Quantitect SYBR green qPCR Mastermix (QIAGEN, Germany) on a Corbett Rotorgene machine 3000/6000 (Corbett Research, Australia). cDNA synthesis was performed with 200-500 ng of total RNA reverse transcribed using Superscript III (Invitrogen) and random hexamer primers (Promega). The synthesized cDNA was diluted 10-fold and 5 µl was used for the assay in the presence of 7.5 µl Quantitect SYBR green mastermix (Qiagen) and 5 pmol gene specific forward and reverse primers. Each reaction was performed in triplicate for both patient samples and calibrator. Reaction conditions were 95°C for 15 min followed by 45 cycles of 20 s at 95°C, 20 s at 58°C and 20 s at 72°C. Data for each cycle was acquired at the 72°C for 20 s step.
The genes that were characterized were β2M, PCA3 and PSA using the following primers (Sigma-Aldrich, Australia) for qPCR: β2M (Fwd: 5’-AGCAGAGAATGGAAAGTCAAA-3’, Rev:5’-TGCTGCTTACATGTCTCG-3’); PCA3 (Fwd: 5’-GGAAGGACCTGATGATACAGAGGTGAG-3’, Rev: 5’-CACAGGGCGAGGCTCATCG-3’; PSA (Fwd: 5’-GCATCAGGAAACAAAAAGCGTG-3’, Rev:5’-CCTGAGGAATCGATTCTTCA-3’)

Standardized processing (including standard curve fitting, dynamic tube, slope correct) was performed for all runs using Rotor-Gene 6000 Series Software Version 1.7 (Corbett Research, Australia). To maintain quality control, specimens with atypical melt curves or quantitation curves below threshold for any single target (with normal melt curves) were excluded. A standard threshold value was manually set for each individual gene across all samples and used to calculate Ct values, which were exported to Excel (Microsoft Corporation, Redmond, CA). Based on CtAv and CtsD values of triplicates observed across all genes (see Figure A3.1), cut off points defined for exclusion were CtAv > 35 and CtsD > 1.5 for PEUW and ejaculate, as previously described15. Analysis of PSA, PCA3 and β2M in ejaculate and PEUW specimens was possible for 38 patients, with these markers detected in PEUW only from 62 patients.

5.2.3.6 – Relative gene expression determination

Relative gene expression was calculated using the method previously described by Pfaffl316, which uses the following equation:

\[
\text{Expression (R)} = \frac{E_{\text{GOI}}^{\Delta \text{Ct}(\text{Calibrator-sample})}}{E_{\text{RG}}^{\Delta \text{Ct}(\text{Calibrator-sample})}}
\]

“ΔCt (Calibrator - sample)” estimates the amplification (Ct values) difference between the calibrator reaction (uniform template quantity to standardize all runs) and the target gene transcription of the sample reaction (unknown)316. The reaction efficiency (E) of the GOI and endogenous RG are considered without the required use of a standard curve in every run, based on the assumption that reaction efficiency between different runs was consistent and normalized by the calibrator used.

5.2.3.7 – RG variation

β2M, a known housekeeping gene241,362, was used as the endogenous RG with subsequent relative gene expression calculated for PSA and PCA3. The commercial use of PCA3, which unlike PSA is highly overexpressed in prostate cancer209, requires PSA to
be used as the RG on the basis of reports that PSA expression is relatively constant and considers only cells of prostatic origin so the expression of PCA3 relative to PSA was also calculated\(^3\).

5.2.3.8 – Data analysis

Relative gene expression results were analysed considering two clinical classifications, absolute CaP status (cancer versus no cancer) and clinically significant (present versus absent) CaP status, as determined using the D’Amico classification\(^3\). Univariate analysis was conducted with the non-parametric Mann-Whitney \(U\) test.

Receiver Operating Characteristic (ROC) analysis was performed for each marker and compared to each other and serum PSA, with binomial exact Confidence Interval and optimal cut points for each marker calculated in Stata Statistical Software 13 (StataCorp. College Station, Texas USA) using the Liu method\(^3\). Significance thresholds were Bonferroni corrected (<0.05/4 = <0.0125) to adjust for multiple comparisons (\(n = 4\)), and all reported \(p\)-values < 0.0125 were considered statistically significant.

5.2.4 Results

5.2.4.1 – Clinical characteristics

From 83 potential patients with adequate clinical data who donated PEUW samples between January 2007 and December 2009, relative gene expression determination for PSA, PCA3 and β2M was performed. Strict exclusion criteria (see Figure A3.2) were applied, resulting in sample exclusion due to unsatisfactory q-PCR analyses (atypical melt or quantitation curves, \(n = 1\)), insufficient cDNA to reach detection threshold (\(n = 5\), CtAv or CtSD outside determined cutoffs (\(n = 14\)) or coupled with an ejaculate sample excluded for a similar reason (\(n = 25\)). Of the remaining 38 patients, the relationships observed for the entire cohort with respect to median (interquartile range) age [62 (57 – 68) years] and serum PSA [6.7 (4.75 – 9.15) ng/ml] were preserved in this group, with a median age 62 (57 – 69) years (\(p = \text{NS}\)) and serum PSA 6.3 (4.9 – 8.9) ng/ml (\(p = \text{NS}\)). Within the included 38 patients were 25 participants having a prostate biopsy positive for cancer with 21 deemed to have clinically significant CaP.

The demographic information of the cohort including median (IQR) age, serum PSA and relative expression ratios of RNA markers for each group, absolute and clinically significant CaP presence, are presented in Table 5.1A and 5.1B, respectively. Men were of
comparable ages in each classification group. Serum PSA was similar for men diagnosed with CaP compared with men in the no CaP group. This relationship was preserved for serum PSA when comparing men with clinically significant CaP against those without clinically significant CaP. Similar trends were observed in the expanded cohort, which comprised 62 patients with 36 participants with a positive prostate biopsy and 30 deemed to have clinically significant CaP (Table 5.1A, 1B).

5.2.4.2 – Biomarker performance

5.2.4.2.1 – Absolute CaP status

When considering absolute CaP status (Table 5.1A) and compared to a chance AUC of 0.500, serum PSA (AUC 0.617; p=0.217) and PSA:β2M (AUC = 0.600; p = 0.353) in PEUW samples provided similar diagnostic performance. PCA3:β2M (AUC = 0.569; p = 0.522) and PCA3:PSA (AUC = 0.431; p = 0.528) in PEUW demonstrated inferior predictive ability. Similar results were not observed in the expanded cohort, with serum PSA (AUC = 0.610, p = 0.128) performing better than PSA:β2M (AUC = 0.506; p = 0.935), PCA3:β2M (AUC = 0.550; p = 0.531) and PCA3:PSA (AUC = 0.529; p = 728).

In ejaculate, best diagnostic performance was observed for PCA3:β2M (AUC = 0.717; p = 0.033), followed by PCA3:PSA (AUC = 0.668; p = 0.078) followed by less impressive performance for serum PSA (AUC 0.617; p=0.217) and poor performance of PSA:β2M (AUC = 0.486; p = 0.895).

5.2.4.2.2 – Clinically significant CaP

When considering clinically significant CaP (Table 5.1B), similar performance was observed for serum PSA (AUC = 0.640; p = 0.124), PSA:β2M (AUC = 0.608; p = 0.269) and PCA3:PSA (AUC = 0.342; p = 0.093). The performance of PCA3:β2M (AUC = 0.496; p = 0.966) was poorer and less predictive of clinically significant CaP. Within the expanded cohort, only serum PSA (AUC = 0.665; p = 0.018) performed as well with poorer performance for the PEUW-based PSA:β2M (AUC = 0.525; p = 0.740), PCA3:β2M (AUC = 0.503; p = 0.967) and PCA3:PSA (AUC = 0.469; p = 0.681).

In ejaculate samples, PCA3:β2M (AUC = 0.667, p = 0.083) was similar to serum PSA, and PCA3:PSA (AUC = 0.605, p = 0.263). The performance of PSA:β2M (AUC = 0.521, p = 0.828) was less impressive. For graphical purposes, comparison ROC curves are
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Comparison cohort</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Extended PEUW Cohort</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>noCaP ($n = 13$); CaP ($n = 25$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>noCaP ($n = 26$); CaP ($n = 36$)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ROC analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ROC analysis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>AUC (95% CI)</td>
<td>Cutpoint</td>
<td>Sn</td>
<td>Sp</td>
<td>AUC (95% CI)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Absolute Cancer status</td>
<td>Demographic information</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>AUC (95% CI)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>noCaP [median (IQR)]</td>
<td>CaP [median (IQR)]</td>
<td></td>
<td></td>
<td></td>
<td>noCaP [median (IQR)]</td>
<td>CaP [median (IQR)]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Age</td>
<td>65 (58.5 – 71)</td>
<td>63 (57 – 68.25)</td>
<td></td>
<td></td>
<td></td>
<td>64.5 (60 – 69)</td>
<td>63.5 (58 – 69)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Serum PSA</td>
<td>6.20 (4.13 – 7.30)</td>
<td>6.50 (5.13 – 10.4)</td>
<td>0.617 (0.445 to 0.770)</td>
<td>9.05</td>
<td>36</td>
<td>100</td>
<td>6.30 (4.70 – 8.2)</td>
<td>7.00 (5.25 – 11.00)</td>
</tr>
<tr>
<td></td>
<td>PEUW PSA:β2M</td>
<td>0.07 (0.003 – 0.36)</td>
<td>0.10 (0.02 – 1.24)</td>
<td>0.600 (0.429 to 0.755)</td>
<td>0.006</td>
<td>84</td>
<td>46</td>
<td>0.32 (0.005 – 1.18)</td>
<td>0.15 (0.14 – 0.97)</td>
</tr>
<tr>
<td></td>
<td>PCA3:β2M</td>
<td>0.23 (0.01 – 0.60)</td>
<td>0.21 (0.06 – 1.24)</td>
<td>0.569 (0.399 to 0.728)</td>
<td>0.652</td>
<td>44</td>
<td>77</td>
<td>0.19 (0.02 – 0.48)</td>
<td>0.18 (0.07 – 0.99)</td>
</tr>
<tr>
<td></td>
<td>PCA3:PSA</td>
<td>2.72 (0.52 – 61.97)</td>
<td>1.45 (0.51 – 9.66)</td>
<td>0.431 (0.272 to 0.601)</td>
<td>0.764</td>
<td>72</td>
<td>38</td>
<td>1.24 (0.14 – 20.80)</td>
<td>1.19 (0.55 – 9.32)</td>
</tr>
<tr>
<td></td>
<td>Ejaculate PSA:β2M</td>
<td>0.07 (0.003 – 0.36)</td>
<td>0.03 (0.003 – 0.19)</td>
<td>0.514 (0.347 to 0.679)</td>
<td>0.003</td>
<td>80</td>
<td>38</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>PCA3:β2M</td>
<td>0.01 (0.004 – 0.10)</td>
<td>0.04 (0.02 – 0.28)*</td>
<td>0.717 (0.548 to 0.851)</td>
<td>0.023</td>
<td>72</td>
<td>77</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>PCA3:PSA</td>
<td>1.55 (0.12 – 2.80)</td>
<td>4.62 (0.77 – 11.84)</td>
<td>0.668 (0.496 to 0.812)</td>
<td>1.77</td>
<td>64</td>
<td>77</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
### Table 5.1: Demographic information for patients based on absolute cancer status (positive vs. negative biopsy; A) and clinically significant CaP classification (B). Median and interquartile range are shown for age and each marker in post-ejaculate urethral washing and ejaculate samples. All comparisons using the Mann-Whitney U-test (two-tailed; using a Bonferroni corrected threshold) were non-significant ($p > 0.0125$). * $p < 0.05$ for Area under curve with 95% confidence interval (AUC 95% CI) comparisons of markers with chance (AUC 0.5) as determined by the DeLong method. All comparisons between AUC estimates were non-significant ($p > 0.0125$). Abbreviations: $\beta 2M$, $\beta 2$-microglobulin; IQR, interquartile range; PEUW, post-ejaculate urethral washings; PSA, prostate specific antigen; PCA3, prostate cancer antigen 3; Sn, Sensitivity; Sp, Specificity.
available in Figures A3.3 and A3.4 including the comparison P-values against the serum PSA AUC.

5.2.5 Discussion

PEUW could potentially be a new source of prostate-specific biomarkers for CaP detection and characterization, providing an alternative to serum and urine for further biomarker discovery and development. Thus, we investigated in this study the utility of prostatic cells in PEUW as a physiological source of CaP biomarkers. We have shown that the diagnostic performance of the mRNA-based marker PCA3, normalised to PSA or β2M, in PEUW is likely to be inferior to these markers in ejaculate, which were comparable to serum PSA. Overall, the performance is similar to PCA3 in post-massage urine in isolation (AUC 0.62)\textsuperscript{369}, as well as in ejaculate in an expanded cohort (AUC 0.625)\textsuperscript{15}. PCA3 performed best in this cohort in detecting absolute CaP status compared with clinically significant CaP, in accordance with previous reports\textsuperscript{365}. This preliminary investigation helps to build on current CaP biomarker research literature.

The use of PEUW as a prostate-based biofluid is advantageous for a number of reasons. First, it contains prostatic effluent following ejaculation, indicated here by the presence of PSA. In addition, there is no requirement for patient discomfort in contrast to prostatic massage. Furthermore, there is the potential for tumour disruption and dissemination of malignant cells\textsuperscript{370}, given known elevations in serum PSA after TRUSBx and DRE\textsuperscript{371}. Reports regarding serum PSA elevation after ejaculation are mixed, with levels reported to return to normal after 48 hours\textsuperscript{372}, thus PEUW following global contraction of the prostate gland with ejaculation can be considered a physiological equivalent of non-physiological post-massage urine. PEUW sampling enables post-coital donation, which may be more acceptable for some men and has been used to investigate infertility. The combination of urine and ejaculatory components in PEUW allows for assessment of markers reflecting local (ejaculate) and systemic (urine) pathology. While this enables use in clinical scenarios where systemic biological alterations are important to monitor, such as active surveillance and metastatic disease, it is also a potential drawback of PEUW, because the local pathology markers in the ejaculate component can be confounded by the systemic contributions from the urine component. These may contribute to the lower diagnostic performance of PEUW compared to ejaculate described here.

The benefits that we have previously outlined for the use of ejaculate as a prostate-based biofluid also apply to PEUW. Specifically, ejaculate contains malignant prostatic epithelial
cells\textsuperscript{14}, with cell-derived molecular markers PCA3 and Hepsin shown to be comparable diagnostically with PCA3 in post-massage urine\textsuperscript{15}. Analysis of microRNAs in cell-derived mRNA in ejaculate has been reported to improve CaP detection, with miR-200b combined with serum PSA (AUC = 0.751) significantly better than serum PSA alone (AUC = 0.555)\textsuperscript{16}. We have previously reported the use of a composite score, created using contributions from serum PSA, ejaculatory miR-125b and miR-200c, to significantly improve CaP detection (AUC = 0.869) compared with serum PSA alone (AUC = 0.672; \( p < 0.05 \))\textsuperscript{15}. The ability to provide an ejaculate sample may also indicate a favourable “performance” status and consequent survival benefit, with a high and significantly better overall and CaP-specific survival benefit observed for these patients at 10, 15 and 20 years\textsuperscript{17}. Incorporation of exosome and metabolome analysis may improve predictive accuracy using these non-invasively obtained biofluids\textsuperscript{20,373}, reducing anxiety and uncertainty for clinicians and patients managed by active surveillance, in addition to assisting with CaP testing. Use of PEUW may be more favourable than ejaculate as PEUW samples can be provided in the comfort of the home environment and post-intercourse, a strategy more likely to be used by men than the sterile surrounds of the clinic setting. Similar to urine cytology, the sample could be stored overnight in the fridge and brought to the clinic the morning after, assuming RNA integrity is maintained.

The aim of this comparative study was to investigate the diagnostic potential of selected molecular markers in PEUWs in detecting CaP on the basis of the D'Amico classification benchmark\textsuperscript{337}, widely used to stratify in the past. However, the goal posts are in the process of being changed as it is being realized increasingly that intermediate risk CaP is not one condition but a spectrum of conditions. Recently, the management strategy of active surveillance has been extended to include some Gleason 3+4 (ISUP 2) tumours regarded as ‘favourable’ by the National Comprehensive Cancer Network for men with a life-expectancy <10 yrs\textsuperscript{374}, with this change supported by the American Society of Clinical Oncology\textsuperscript{375}. By inference, this means not all intermediate-risk tumours can be considered as clinically significant. However, for the purpose of comparison, risk classifications such as that proposed by D'Amico remain relevant until the entity of clinically significant can be defined better.

The limitations of this preliminary, exploratory study include the small sample size resulting in large confidence intervals and low statistical power. Biologically, the potential for low ejaculatory contribution or dilution reducing the prostate-specific RNA yield may impair results. PCA3:PSA levels in PEUW were lower in men with CaP, both in the comparison
and expanded cohorts, resulting in ROC estimates below 0.5, which was the inverse of that observed for ejaculate samples and other published reports. Although the sample drop-out rates were similar for PEUW and ejaculate, with positive PSA signals suggesting the presence of prostatic material in both sample sets, these results suggest malignant cells may dominate in the ejaculate. A potential explanation for this are changes in cell adhesion molecules, which have been recognized for quite some time in CaP\textsuperscript{376}, with most attention focused on E-cadherin. Loss of E-cadherin is particularly evident in more aggressive tumours with cadherin switching also recently described\textsuperscript{377}. Thus, disaggregated cells or cell clusters from aggressive tumours first appear in ejaculate as a result of global contraction of prostatic smooth muscle following accumulation in acini prior to ejaculation. This in turn may cause relatively fewer cancerous cells to be present in the urethra from the latter part of the ejaculate, which would be dislodged with subsequent voiding. As a result, PEUW may contain less cancerous cells but does contain prostate epithelial cells, evidenced here by the presence of PSA. Thus, in patients suspected to harbor non-aggressive CaP, the presence of prostatic cells in PEUW with low expression of PCA3 may help determine which patients have non-aggressive CaP amenable to active surveillance or watchful waiting.

In conclusion, we introduced and investigated PEUW as a physiological source of CaP biomarkers. We found that PEUW contains prostatic cells, as evidenced by PSA signal. However, significantly upregulated PCA3 levels, consistent with those reported for malignant tissue, were only observed in ejaculate specimens. PEUW may yet prove to be a useful source of cell-free secreted markers, as opposed to cell-derived markers. Further biomarker development using these prostate-specific biofluids may result in improved diagnosis and monitoring of CaP, reducing anxiety and doubt for the benefit of clinicians and patients.
5.3 NMR-based metabolomics of post-ejaculate urethral washings to predict and characterise prostate cancer using local and systemic metabolite changes.

The contributions of the co-authors to this chapter section were as follows: Experiment design, experiment execution, data processing and statistical analysis and drafting of this chapter section was performed by Matthew Roberts, the PhD candidate, under the supervision of Dr Horst Schirra and Prof RA “Frank” Gardiner. Renee Richards, Clement Chow and Marion Buck assisted with sample collection. Jake Hattwell assisted with experiment execution, data processing and statistical analysis. John Yaxley provided advice on histological and clinical interpretation and Prof Martin Lavin provided critical manuscript revisions. Preparation of the manuscript and associated figures was performed by Matthew Roberts, the PhD candidate.

The references have been collated in the section, ‘Bibliography’, with others contained in this thesis.
5.3.1 Introduction

Early diagnosis of CaP to facilitate curative treatment requires an accurate biomarker superior to serum PSA, which is limited in detecting csCaP. Genetic markers in excreted prostatic fluids, such as non-coding RNA PCA3, in combination with the TMPRSS2:ERG fusion gene, improve CaP diagnosis but are used uncommonly as adjunctive tests\textsuperscript{10}. Prostatic fluid excreted in ejaculate contains biomarkers predictive of csCaP, with the PCA3 and Hepsin mRNA and miRNAs detected in the epithelial cell component providing a diagnostic improvement compared to serum PSA alone\textsuperscript{15}. Biomarkers contained in the plasma component, or ejaculate, also improve diagnostic accuracy, such as the metabolites citrate, spermine and myo-inositol and proteins\textsuperscript{12,23}.

Difficulties in collecting ejaculate due to erectile dysfunction, performance anxiety or individual reservations have driven investigation of other prostatic fluids\textsuperscript{12}. PEUW, being the first urine sample following ejaculation, contains PSA RNA indicating a prostatic contribution to this biofluid and opportunity for biomarker discovery\textsuperscript{321}. PEUW has the potential to provide information regarding local pathology from the prostatic contribution, as well as systemic disturbances due to CaP, from the renal contribution. While the diagnostic performance of PEUW prostatic RNA did not improve when compared to serum PSA, this may be due to a dominance of normal prostatic cells over malignant cells and help identify patients without csCaP.

Profiling of metabolites, or metabolomics, applied to urine has provided biomarker panels reported to significantly improve diagnostic accuracy\textsuperscript{293}. Furthermore, local and systemic metabolic disturbances have been suggested, with fatty acid synthesis most commonly implicated\textsuperscript{294}. Thus, metabolite profiling of PEUW is expected to demonstrate locally and systemically deranged metabolite levels and potentially link these to altered metabolic pathways.

The aims of this study were to:

1. Investigate the feasibility of PEUW NMR-based metabolomics and if PEUW contains metabolites from ejaculate that are different to normal MSU metabolites (Feasibility cohort).
2. Test whether the PAP-mediated hydrolysis of phosphocholine to choline happens in PEUW similar as in ejaculate (as described in Chapter 3.2) (Kinetics),
3. Determine the diagnostic performance of PEUW metabolite profiles to predict and characterise CaP in a high-risk clinical cohort compared to diagnosis obtained via prostate biopsy and RP histology (Clinical cohort).

5.3.2 Materials and Methods

This study received ethical approval from the RBWH Human Research Ethics Committee (HREC/09/QRBW/320, HREC/09/QRBW/305 together with 1995/088B) and University of Queensland Medical Research Ethics Committee (Project no. 2006000262).

5.3.2.1 – Participants and Samples

5.3.2.1.1 – Feasibility

Firstly, to investigate the feasibility of PEUW NMR-based metabolomics and if PEUW contains metabolites from ejaculate, PEUW samples were first collected into sterile micro urine jars from eight men attending the RBWH for investigation of an elevated PSA or abnormal digital rectal examination (Pilot cohort). Following standard sample processing, as described in Chapter 5.2, supernatants were snap-frozen on dry ice in 1 ml aliquots and stored at –80°C.

Then to compare PEUW profiles with normal mid-stream urine metabolite profiles, male patients \( (n = 4) \) younger than 30 years old and with no significant family history of CaP provided urine, ejaculate and PEUW specimens either on site or at home (Young cohort). All samples were delivered to the laboratory without cooling as soon as logistically possible and processed within 2 h of production. Subjects provided 10 sample sets, with control MSU sample collections interspersed between collection of ejaculate/PEUW samples per the collection regime outlined in Table 5.2.

5.3.2.1.2 – Kinetics

Secondly, to determine the kinetic behaviour of ejaculate-based metabolites within PEUW and compare with those observed in ejaculate only, a PEUW sample was collected into a sterile micro urine jar from a healthy male (aged 26) for immediate processing and NMR spectroscopy.

5.3.2.1.3 – Clinical cohort

Finally, to determine if PEUW provided accurate discrimination for the diagnosis and characterisation of csCaP, PEUW samples were also collected from the population
Table 5.2: Sample collection regime for Young cohort biofluid samples. Mid-stream urine (MSU; A-D) sample collections are interspersed between collections of ejaculate (E) and post-ejaculate urethral washing (PEUW) samples.

<table>
<thead>
<tr>
<th>Sample Number</th>
<th>Label</th>
<th>Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>MSU</td>
</tr>
<tr>
<td>2</td>
<td>1 E/U</td>
<td>E, PEUW</td>
</tr>
<tr>
<td>3</td>
<td>2 E/U</td>
<td>E, PEUW</td>
</tr>
<tr>
<td>4</td>
<td>B</td>
<td>MSU</td>
</tr>
<tr>
<td>5</td>
<td>3 E/U</td>
<td>E, PEUW</td>
</tr>
<tr>
<td>6</td>
<td>4 E/U</td>
<td>E, PEUW</td>
</tr>
<tr>
<td>7</td>
<td>C</td>
<td>MSU</td>
</tr>
<tr>
<td>8</td>
<td>5 E/U</td>
<td>E, PEUW</td>
</tr>
<tr>
<td>9</td>
<td>6 E/U</td>
<td>E, PEUW</td>
</tr>
<tr>
<td>10</td>
<td>D</td>
<td>MSU</td>
</tr>
</tbody>
</table>

Sample preparation

PEUW samples were thawed on wet ice (Feasibility, Young, Clinical cohorts) or taken fresh (Kinetics sample) and distributed as aliquots comprising 80% NMR tube volume (500 μl for 5mm tube, 200 μl for 3mm tube), 10% 0.1 M sodium phosphate with borate and combined with D$_2$O as lock substance that contained DSS as internal chemical shift standard and DFTMP as internal pH indicator leading to final concentrations of DSS and DFTMP of 100 μM to contribute the remaining 10% NMR tube volume. Samples were transferred to sterile 5 mm (Clinical cohort) and 3 mm (Feasibility, Young cohorts, Kinetics sample) NMR tubes (Bruker Biospin, Rheinstetten, Germany).
Ejaculate samples for the Young cohort were collected and prepared as described in Chapter 4.3.

5.3.2.3 – NMR spectroscopy

NMR spectra of PEUW samples were measured on Bruker Avance spectrometers equipped with a 5 mm self-shielded z-gradient triple resonance cryoprobe and sample changer operating at a $^1\text{H}$ frequency of 900.13 MHz (Bruker Biospin; Feasibility, Young, updated Clinical cohorts, Kinetics sample) and 500 MHz (initial Clinical cohort).

For the Feasibility, Young and Clinical cohorts, 1D NOESY spectra were acquired at 298 K with the “noesypr1d” pulse sequence, accumulating 200 (Feasibility, Young) and 128 (Clinical) transients (following 8 dummy scans) at 32k data points with a spectral width of 14 ppm. The transmitter frequency was set to the water resonance, which was suppressed by continuous wave irradiation during the NOESY mixing time of 0.15 s and relaxation delay of 2.3 s. Tuning/matching, shimming, and data acquisition were performed automatically with the ICON-NMR interface for high-throughput automation.

For the Kinetics sample, a pseudo-2D NMR spectrum was acquired, processed and subject to linear curve fitting as described in Chapter 3.2.

5.3.2.4 – Spectral Processing

NMR spectra were processed in TopSpin 3.2 (Bruker Biospin). The free induction decays were baseline corrected by a Gaussian function (0.1 ppm filter width) for post-acquisition water deconvolution, followed by multiplication with an exponential window function (0.1 Hz line broadening), and Fourier transformation to 64k points. Subsequently, the spectra were manually phased, baseline corrected and referenced to DSS at 0.0 ppm. For all further data manipulation, the spectra were truncated to $\delta=10.0–0.25$ ppm, exported into MATLAB 2015b (The Mathworks Inc., Natick, Massachusetts, United States).

5.3.2.5 – Spectral alignment and data reduction

Spectra were exported either in full resolution from TopSpin or as spectral bins, or buckets, of 0.0005 ppm width using AMIX (version 3.6.6, Bruker Biospin, Rheinsetten, Germany) and subsequently imported into MATLAB 2015b (The Mathworks Inc., Natick, Massachusetts, United States). Automatic peak alignment was performed using “icoshift” based on manually defined segments and data reduced to buckets of 0.01 ppm width over the range 10.0–6.5 and 4.5-0.25 ppm using an in-house MATLAB script.
5.3.2.6 – Multivariate statistical analysis

MVSA was performed using methodology described in Chapter 4.3, except the exploratory, preliminary Clinical cohort analysis of 106 samples which used integral normalisation.

5.3.3 Results

5.3.3.1 – Feasibility cohort

1D NOESY spectra of PEUW samples demonstrated the presence of metabolites from both ejaculate (e.g. choline, phosphocholine) and urine (e.g. creatinine), which are not reciprocally observed in the other biofluid. Metabolites contained in both ejaculate and urine, such as citrate, were also observed (Figure 5.1).

![Figure 5.1: 1D NOESY NMR Spectrum of PEUW, demonstrating the presence of metabolites known to be present in ejaculate (*choline), urine (*creatinine, Cre; *Urea; *hippurate, Hip; *trimethylamine N-oxide, TMAO; *sarcosine, Sar*) and in both (*citrate, Cit*).](image-url)
5.3.3.2 – Young cohort

Three subjects provided samples totalling 14 MSU and 22 paired PEUW and ejaculate samples (Table 5.3).

<table>
<thead>
<tr>
<th>Subject</th>
<th>Cycles completed</th>
<th>Samples provided</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>MSU</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>0.6</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>14</td>
</tr>
</tbody>
</table>

Table 5.3: Collection totals for mid-stream urine (MSU), post-ejaculate urethral washing (PEUW) and ejaculate (E) samples.

To date, 1D NOESY spectra have been acquired for MSU and PEUW samples, which predominantly demonstrate the presence of the choline peak at 3.195 ppm in PEUW samples compared with MSU samples (Figure 5.2).

![Figure 5.2: 1D NOESY segment (3.15 – 3.23 ppm) spectral overlay of a urine-based sample set from the Young cohort. A sample collection cycle shows mid-stream urine (MSU) spectra relative to post-ejaculate urethral washings (PEUW) spectra, with the choline peak (dotted line) predominantly observed in PEUW but not MSU spectra.](image)

Spectral alignment, with potential deconvolution and subsequent MVSA will be performed.
5.3.3.3 – Kinetics

To test whether the PAP-mediated hydrolysis of phosphocholine to choline happens in PEUW similar as in ejaculate (as described in Chapter 3.2), we measured pseudo-2D NOESY spectrum of PEUW (Figure 5.3a). Hydrolysis of phosphocholine to choline was observed (Figure 5.3b). The apparent rate constant for this reaction in PEUW was (0.033 - 0.038 hr\(^{-1}\); R\(^2\) = 0.991) in PEUW. This is considerably slower than in ejaculate at room temperature (0.6761 hr\(^{-1}\)) and is between that seen for 5mM (0.044 – 0.055 hr\(^{-1}\); Figure 3.3) and 10mM tartrate in ejaculate (0.011 – 0.024 hr\(^{-1}\); Figure 3.3) at 298 K.

![Figure 5.3](image)

**Figure 5.3:** Pseudo-2D spectral overlay of serial measurements of PEUW at 298K over time in the region of choline/phosphocholine (A). Hydrolysis and subsequent decrease of phosphocholine results in an increase in choline (B), as demonstrated by the reaction curves, at a rate similar to inhibition of the reaction when 5 - 10mM tartrate is added to ejaculate at 298 K (Figure 3.3).

5.3.3.3 – Clinical cohort – Preliminary analysis

The demographics of the entire clinical cohort are as presented in Chapter 4.3.
A preliminary metabolomics analysis was undertaken on 106 samples, which underwent standard preparation, NMR spectral acquisition at 500 MHz with standard processing and data reduction at full resolution. Following integral normalisation, MVSA was performed in SIMCA using the clinical classifiers most reflective of available tumour grade. Buckets containing ethanol were excluded and the bucket table recalculated.

Principal components analysis yielded a model (n=106, k=765, R²X=0.251; Q²=0.075) with 2 principal components (PCs), with outlying samples present due to the presence of hippurate (PC1), creatinine (PC2) and unknown compounds (UC1, UC2; Figure 5.4a-c). As no discrimination was apparent for csCaP and risk groups and following exclusion of samples containing the UC1 (n = 3) and abnormal spectra (n = 2), supervised PLS using csCaP as the dependent variable (n=106, k=765, A=2 [forced], R²Y=0.413; Q²Y=0.113) showed trimethylamine N-oxide (TMAO), glycine and choline to be significant metabolites (Figure 5.4d-e). PLS analysis that incorporated risk groups as the dependent variable (n=106, k=765, A=2 [forced], R²Y=0.159; Q²Y=-0.077) showed choline, creatinine, glycine and TMAO to be associated with high, intermediate, low and benign risk groups, respectively (Figure 5.4f-g).

With no discrimination according to csCaP, risk subgroup analysis was performed as described in Chapter 4.3. Similar findings were observed, with group separation among low and intermediate samples (n=20, k=765, A=1 [auto], R²Y=0.51; Q²Y=0.164) and negative/low risk and benign samples (n=45, k=765, A=1 [auto], R²Y=0.369; Q²Y=0.124). Despite comparison with intermediate risk and benign samples, with creatinine and TMAO observed to be lower in negative/low risk samples, while glycine and an unknown compound were elevated (Figure 5.5). The remainder of subgroup analyses were invalid, while the metabolite patterns observed with supervised analyses were inconsistently observed in unsupervised analyses with the same spectra (data not shown).
Figure 5.4: Preliminary analysis of 106 PEUW spectra from patients being investigated for CaP.

Panels a-c: Principal Components Analysis suggested influential metabolites to include hippurate (Hip), creatinine (Cre) and unknown compounds (UC1, UC2) without discrimination between csCaP groups (blue = negative/low risk; red = high risk). Panels d-e: Supervised, partial least squares analysis for csCaP demonstrated trimethylamine N-oxide (TMAO), glycine (Gly) and choline (Cho) to be significant metabolites in an invalid model. Panels f-g: PLS analysis for CaP risk (Y variable) groups demonstrated choline, creatinine, glycine and TMAO to be associated with high (red), intermediate (gold), low (green) and benign (blue) risk groups, respectively.
Figure 5.5: Preliminary sub-group analysis of PEUW spectra classified as negative, low risk (green) compared with intermediate risk (gold) and benign (blue). Despite comparison with intermediate risk and benign samples, creatinine (Cre) and TMAO were observed to be lower in negative/low risk samples, while glycine (Gly) and an unknown compound (UC2) were elevated.

When citrate, a metabolite consistently implicated in CaP biofluid studies, was not observed to be significant, as seen in the ejaculate metabolomics analysis in Chapter 4.3, inspection of the spectral overlay showed that the citrate signals were distorted due to the formation of borate adducts from the preservative buffer (Figure 5.6). In order to optimize information recovery and reduce artefacts from citrate adducts, all samples were...
re-measured using sodium azide as preservative on a Bruker 900MHz spectrometer (with cryoprobe and SampleJet sample changer) with similar parameters to previous to acquire 1D NOESY spectra.

**Figure 5.6:** 1D NOESY PEUW spectral overlay showing sarcosine and citrate in the region 2.74 – 2.5 ppm. Samples initially measured using borate preservative demonstrate citrate adducts, despite interval alignment with icoshift (a). Samples were re-measured using azide preservative to eliminate citrate adducts (b), which after optimal peak alignment (c), no peak overlap or adducts were observed.

### 5.3.4 Discussion

In this chapter section, it has been shown that PEUW contains both prostatic and systemic metabolites that require further investigation. Ejaculate-sourced metabolites in PEUW were present in the feasibility cohort and shown to behave similarly to that observed in ejaculate in the kinetics experiment. Ejaculate metabolites in PEUW were observed when compared to MSU and will be substantiated in further analyses comparing PEUW and MSU spectra to ejaculate spectra using MVSA methods. This study will serve to establish the reliability of PEUW as a biofluid that has the ability to provide information of local prostatic pathology, provided by the ejaculate component, as well as systemic changes present in conjunction with or as a result of prostatic pathology.
The reduced hydrolysis of phosphocholine to choline observed in PEUW, where ejaculate metabolites are contained within a urine media, is likely to be multifactorial. Indeed, reduced hydrolysis could be accounted for by differences in pH between ejaculate (pH 6.9) and PEUW (pH 5.14) based on studies in prostate extract where a 50% reduction in PAP activity is observed between these pH levels\textsuperscript{314}. Varied inhibition has also been described due to presence of ions, such as calcium and zinc, which are commonly present in urine and thus PEUW\textsuperscript{314}.

When preliminary analysis was conducted in 106 patients being investigated for CaP, group separation was most pronounced when comparing negative/low risk with benign and intermediate risk samples. This relationship was observed for ejaculate samples in Chapter 4.3 and supported by other investigators, where it has been shown that metabolite changes are greatest early in the malignant transformation process\textsuperscript{279,280}. With ejaculate-metabolites minimally affected by PAP-related hydrolysis, the association of choline with high risk samples may be consistent with that observed \textit{in vivo}, while creatinine has been implicated with CaP based on epidemiological data. Glycine has been associated with CaP tumorigenesis and altered methionine and other amino acid metabolism\textsuperscript{283}. The association of TMAO with benign and intermediate risk samples is likely an artefact of misclassification of the benign samples, as discussed at length in Chapter 4.3. TMAO has been reported in other studies to be associated with aggressive CaP in serum\textsuperscript{289}.

This preliminary, exploratory study is limited by incomplete analysis of the Young cohort experiment and Clinical cohort measured with sodium azide at 900 MHz. Additionally, the results are speculative due to the few low risk tumours included and limited statistical certainty, as these were the result of opportunistic sample collection for this proof-of-concept study. Thus, this study was hypothesis-generating and statistical power considerations were reserved for future, larger validation studies on the basis of estimates presented here. Furthermore, potentially low ejaculate metabolite contributions from dilution with urine or reduced ejaculate volume may be addressed by normalisation to a consistently expressed compound unique to ejaculate, which is currently being investigated. As observed for RNA and discussed in Chapter 5.2, low ejaculate-based metabolites may also be a surrogate indicator of reduced malignant cell excretion in PEUW, however use of PEUW supernatants may serve to overcome this cell-based limitation, as observed in preliminary analyses.
5.3.5 Conclusions

Following proof-of-concept for PEUW use and initially promising clinical results, it is expected that the multivariate model based on the metabolite profiles in PEUW will improve the prediction of CaP in this clinical cohort. These results can then be applied to other cohorts prospectively. The metabolites contained within this unique, non-invasive biofluid are likely to represent CaP-induced local and systemic metabolic disturbances.
Chapter 6 – From bench to bedside: clinical implementation of prostatic fluid biomarkers

6.1 Synopsis

The aim of this chapter is to describe how, based on the previously presented methods and results demonstrating the feasibility and potential to predict and characterise CaP, prostatic fluid biomarker performance can be confirmed in a phase 2 “window” clinical trial. Section 6.2 contains a published peer-reviewed published clinical trial protocol that outlines the implementation of a novel methodology developed within this thesis where serum PSA, combined with prostatic fluid biomarkers may be used to characterise and monitor CaP. The use of MRSI is described to demonstrate how in vivo metabolite profiling may assist in monitoring of disease progression or treatment effects.
6.2 “Can atorvastatin with metformin change the natural history of prostate cancer as characterized by molecular, metabolomic, imaging and pathological variables? A randomized controlled trial protocol.”

(published manuscript)

The manuscript entitled, “Can atorvastatin with metformin change the natural history of prostate cancer as characterized by molecular, metabolomic, imaging and pathological variables? A randomized controlled trial protocol.” has been published by *Contemporary Clinical Trials* (2016; 50:16-20)

The contributions of the co-authors to this manuscript were as follows: Trial co-ordination and recruitment, experiment design, data processing and statistical analysis drafting was performed by Matthew Roberts, the PhD candidate, under the supervision of Dr Horst Schirra and Prof RA “Frank” Gardiner. John Yaxley, Geoff Coughlin, Troy Gianduzzo, Rachel Esler and Nigel Dunglison will contribute to trial recruitment. Suzanne Chambers will provide psychosocial advice and oversight. Robyn Medcraft will assist with trial co-ordination. Renee Richards, Clement Chow, Marion Buck and Luke Selth will continue to assist with sample collection, experiment design and execution with data processing. Nicholas Kienzle, Macy Lu and Ian Brereton will assist with imaging co-ordination and interpretation. Hema Samaratunga, Joanna Perry-Keene and Diane Payton will provide advice on histological interpretation. Chikara Oyama will assist with laboratory experimental design. Suhail Doi provided guidance for sample size calculation and will assist with statistical analysis and data interpretation. Prof Martin Lavin provided critical manuscript revisions. Preparation of the manuscript and associated figures was performed by Matthew Roberts, the PhD candidate.

The manuscript text is presented as originally published by the journal. To maintain a consistent style between chapters in this thesis, abbreviations, placement of figures and tables within the text, as well as numbering of pages, figures and tables has been adjusted. The references have been collated in the section, ‘Bibliography’, with others contained in this thesis.
Can atorvastatin with metformin change the natural history of prostate cancer as characterized by molecular, metabolomic, imaging and pathological variables? A randomized controlled trial protocol.

Matthew J. Roberts\textsuperscript{1, 2, 3, 4}, John W Yaxley\textsuperscript{2}, Geoffrey D Coughlin\textsuperscript{2}, Troy R. J. Gianduzzo\textsuperscript{2, 3}, Rachel C. Esler\textsuperscript{2}, Nigel T. Dunglison\textsuperscript{2}, Suzanne K. Chambers\textsuperscript{1}, Robyn J Medcraft\textsuperscript{1, 2}, Clement W. K. Chow\textsuperscript{1}, Horst Joachim Schirra\textsuperscript{4}, Renee Richards\textsuperscript{1}, Nicholas Kienzle\textsuperscript{6}, Macy Lu\textsuperscript{6}, Ian Brereton\textsuperscript{4}, Hema Samaratunga\textsuperscript{7}, Joanna Perry-Keene\textsuperscript{8}, Diane Payton\textsuperscript{8}, Chikara Oyama\textsuperscript{9}, Suhail A. R. Doi\textsuperscript{10, 11, 12}, Martin F. Lavin\textsuperscript{2}, Robert A. Gardiner\textsuperscript{1, 2, 13#}

\textsuperscript{1} The University of Queensland, Centre for Clinical Research, Building 71/918, RBWH, Herston Brisbane, Qld, 4029, Australia
\textsuperscript{2} Department of Urology, Royal Brisbane and Women’s Hospital, Brisbane, Qld, 4029, Australia
\textsuperscript{3} School of Medicine, The University of Queensland, Herston, Qld, 4006, Australia
\textsuperscript{4} The Centre for Advanced Imaging, The University of Queensland, St. Lucia, Brisbane, Qld 4072, Australia
\textsuperscript{5} Griffith Health Institute, Griffith University, Gold Coast, Qld 4222, Australia
\textsuperscript{6} Department of Medical Imaging, Royal Brisbane and Women’s Hospital, Brisbane, Qld, 4029
\textsuperscript{7} Auesta Pathology, PO Box 1878, Toowong DC, Brisbane, Qld, 4066, Australia
\textsuperscript{8} Anatomical Pathology, Pathology Queensland, Herston, Brisbane, Qld, 4029, Australia
\textsuperscript{9} Department of Urology, Hirosaki University Graduate School of Medicine, 5-Zaifu-cho, Hirosaki 036-8562, Japan
\textsuperscript{10} Research School of Population Health, The Australian National University, Canberra, Australia
\textsuperscript{11} School of Agricultural, Computational and Environmental Sciences, University of Southern Queensland, Toowoomba, Australia
\textsuperscript{12} College of Medicine, Qatar University, Doha, Qatar
\textsuperscript{13} Edith Cowan University, Joondalup, Western Australia
6.2.1 Abstract

**Background:** Atorvastatin and metformin are known energy restricting mimetic agents that act synergistically to produce molecular and metabolic changes in advanced prostate cancer. This trial seeks to determine whether these drugs favourably alter selected parameters in men with clinically-localized, aggressive CaP.

**Methods/design:** This prospective phase II randomized, controlled window trial is recruiting men with clinically significant CaP, confirmed by biopsy following multiparametric MRI and intending to undergo RP. Ethical approval was granted by the Royal Brisbane and Women’s Hospital Human and The University of Queensland Medical Research Ethics Committees.

Participants are being randomized into four groups: metformin with placebo; atorvastatin with placebo; metformin with atorvastatin; or placebo alone. Capsules are consumed for 8 weeks, a duration selected as the most appropriate period in which histological and biochemical changes may be observed while allowing prompt treatment with curative intent of clinically significant CaP. At recruitment and prior to RP, participants provide blood, urine and ejaculate. A subset of participants will undergo 7Tesla magnetic resonance spectroscopy to compare metabolites *in-vivo* with those in ejaculate and biopsied tissue.

The primary end point is biochemical progression, defined using biomarkers (serum PSA; PCA3 and citrate in ejaculate and prostatic tissue). Standard pathological assessment will be undertaken alongside quality of life and psychosocial outcome assessments.

**Discussion:** This study is designed to assess the potential synergistic action of metformin and atorvastatin on CaP tumour biology. The results may determine simple methods of tumour modulation to reduce disease progression.

**Keywords:** Prostate cancer; atorvastatin; metformin; clinical trial; biomarkers; metabolomics
6.2.2 Introduction

Aggressive CaP cells increase glucose uptake and glycolysis under normoxic conditions (the Warburg effect\textsuperscript{179}) producing glycolytic intermediates that also feed biosynthesis and CaP proliferation\textsuperscript{379,380}. Metformin reduces glucose oxidation to increase glutamine metabolism and cell death while inhibiting metastatic behaviour. Epidemiological evidence suggests metformin use is associated with reduced risks of many cancers, including CaP\textsuperscript{381,382} with reduced hyperinsulinaemia by metformin in advanced CaP potentially improving ADT response\textsuperscript{383}. When statins are combined with metformin, further reduction in CaP progression and improved clinical outcomes have been reported, indicating a potential additive or synergistic effect to this medication combination\textsuperscript{384-386}. Statins reduce cholesterol and mevalonic acid biosynthesis, with \textit{in vitro} evidence that statins slow testosterone synthesis by inhibiting pre-cursor molecule transport, improving ADT response\textsuperscript{387}. Hypercholesterolaemia is associated with high risk CaP\textsuperscript{388} and androgen-independent CaP metastasis\textsuperscript{389} with statin use associated with lower PSA levels, percentage positive biopsies and fewer cases of advanced and fatal disease\textsuperscript{385}. Lipophilic statins such as atorvastatin also inhibit CaP cell migration to bone marrow stroma\textsuperscript{390}, however benefit in reducing biochemical recurrence, remains uncertain\textsuperscript{391}.

Medication safety profiles of metformin and atorvastatin are favourable, with significant side effects rarely observed. Concern for metformin regarding lactic acidosis is reserved for patients with significant comorbidities (chronic renal failure, congestive cardiac failure)\textsuperscript{392}. Large cohorts consuming statins report rhabdomyolysis in up to 11 per 100,000 person-years\textsuperscript{393}. The use of metformin as a neoadjuvant agent for 4-12 weeks by Joshua and colleagues was well tolerated and demonstrated a 10% reduction in PSA, 6.5% reduction in IGF-1 and 5% reduction in BMI\textsuperscript{394}. Such ideal drug tolerability and favourable clinical effects supports their adjunctive use in localized prostate cancer without need for a phase I controlled trial in this context.

Metformin and atorvastatin may influence malignant metabolic transformation in the prostate, known to favour ATP production and fatty acid synthesis, by shifting citrate, detectable in ejaculate\textsuperscript{12,20,289}. Markers, such as PCA3, improve CaP detection and disease monitoring but may vary with epigenetic and exogenous stimuli\textsuperscript{15,321}.

Initially promising findings by Joshua and colleagues demonstrated significant changes in molecular markers following neoadjuvant metformin therapy prior to RP\textsuperscript{394}. These
medications are also being explored in Metformin Active Surveillance Trial Study (NCT01864096) in delaying pathologic disease progression. Thus, exploring the role of energy restriction mimetic agents represents an exciting development in managing men with CaP. However, before atorvastatin and metformin can be entertained for use in patients with early CaP, their potential demonstrable beneficial effects with respect to tumour parameters need to be evaluated objectively.

The primary aim of this study is to determine whether these drugs by themselves and together, favourably alter selected parameters in a group of clinically-localized, aggressive tumours.

6.2.3 Materials and Methods

6.2.3.1 – Study design

This is a prospective randomized, double-blinded controlled phase II window trial designed to determine the efficacy on biochemical progression of atorvastatin and metformin, in isolation and together, in a population of men with early, clinically significant CaP. In addition, the effect of these drugs on CaP biology will be assessed in a population not previously studied in this respect whilst these men await definitive treatment by RP, in accordance with a phase II window trial design\textsuperscript{395}. Men with an elevated PSA who have a mpMRI examination that demonstrates a PI-RADS 4 or 5 lesion and who, at consultation, express an intention to proceed to RP should biopsy confirm the suspicion of high-risk CaP, will be approached to enter the study. Our current practice includes in depth counselling prior to biopsy in order to ascertain the benefits to the patient in investigating for CaP. This includes outlining the biopsy and treatment process, with treatment options of surgery, radiotherapy, active surveillance or watchful waiting all discussed. Following written informed consent and randomisation by the manufacturing pharmacy (QPharm) to ensure clinician and participant blinding, four study groups are being examined, as outlined in Figure 6.1.

The protocol is designed and reported according to the SPIRIT guidelines\textsuperscript{396}. Participants will provide blood, urine and ejaculate after 48hrs abstinence of sexual activity. Blood and ejaculate samples will be used to determine biomarkers of interest as defined by the primary and secondary endpoints. Further exploratory analyses will be conducted as outlined in order to determine biochemical effects of these medications in this patient cohort. Prior to giving specimens, a subset of participants, selected by a sub-
randomisation process, will opportunistically undergo a further mpMRI with MRS using a 7Tesla machine at the University of Queensland Centre for Advanced Imaging. Here, we will assess the metabolic profile of participants prostates in-vivo, for comparison with those seen in ejaculate in-vitro, and ascertain if superior imaging is provided by this machine.

Participants will then undergo transperineal prostate biopsy targeting lesions of interest (cognitive biopsies) detected by pre-trial mpMRI, in addition to systematic whole gland biopsies using a template as per the local department protocol. Biopsy samples from index lesions and from non-index lesion areas will be taken for research purposes and stored for subsequent molecular and NMR analysis.

Participants will undergo 8 weeks of capsule consumption, as this duration was determined to be most appropriate in which histological and biochemical changes may be observed while allowing prompt treatment with curative intent of clinically significant CaP. Non-invasively obtained participant samples will be collected again and mpMRI with MRS will be repeated (for those previously randomized to have these investigations). The reason for allocating only a limited number of participants for mpMRI and MRS with the 7 Tesla machine is cost. The biomarker kinetic changes following biopsy are poorly described, however we expect these will be minimally affected by biopsy artefacts with 8 weeks of

---

**Figure 6.1:** Flow chart of study design, incorporating timing of specimen collection, imaging and capsule distribution.
treatment and healing. Latifoltojar and colleagues examined changes in mpMRI parameters following biopsy and described a return to baseline apparent diffusion coefficient parameters 1 month post biopsy\textsuperscript{397}. The effects of biopsy on MRS parameters are currently unknown and will be examined in this study.

Ethical approval has been obtained from the RBWH Human Research Ethics Committee (Approval no. HREC/14/QRBW/153 together with HREC/09/QRBW/320, HREC/09/QRBW/305 and 1995/088B) and The University of Queensland Medical Research Ethics Committee (Approval no. 2014000944 together with 2006000262) using the National Ethical Application Form. Specialist clinicians are overseeing all aspects of management through our established team. This trial has been registered in the Australian New Zealand Clinical Trials Registry (Registration number: ACTRN12615000571572).

Consent to participate in the trial is being obtained from treating urologists or the Trial Coordinator (RM). This process also includes provision for storage and future use of clinical data and biological specimens. Database access is password-protected and limited to the Trial Coordinator (RM) who will then provide information as necessary to the other investigators. The database will be stored on a secure internal (UQ) server. At the conclusion of the trial, data access will be limited to the Trial Coordinator (RM) and relevant Principal Investigators (MJR, RAG).

Adverse events are being monitored by the Trial Coordinator (RM), who contacts participants by telephone on two occasions during the trial to ensure satisfaction and determine the presence of any medication side-effects, or other difficulties. In addition, participants are instructed to contact the trial coordinator should they have any concerns or difficulties at other times during the trial. Further reports to the participants treating urologists and general practitioners, as well as relevant hospitals will be included and promptly assessed in detail.

Untoward events are being determined by participant experiences of new symptoms during the study period, as well as routine monitoring using pathology tests, and the results of any other investigations prompted by other health professionals (e.g. general practitioners, emergency department clinicians) during the study period. The Clavien system is being used to document any adverse effects reported, while study questionnaires are also being monitored for participant satisfaction. Adverse outcomes and trial conduct are regularly audited and discussed with the relevant ethics representatives.
The funding source had no involvement in study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the article for publication.

6.2.3.2 – Study population and Recruitment

Participants are being recruited from the RBWH Urology Outpatient clinic and specialist private practices in Brisbane, Queensland. Clinicians identify men as having clinically significant CaP on the basis of prostate imaging with mpMRI and who intend to proceed to RP. Following informed consent, those fulfilling all inclusion criteria and no exclusion criteria are entered into the randomization process. Recruitment is promoted by treating urologists discussing the study with eligible participants.

6.2.3.2.1 – Inclusion criteria

- suspected of having high-risk CaP on the basis of mpMRI (PIRADS 4/5 lesions), with clinically-localized CaP and who intend to proceed to RP following confirmation of their disease as high-risk following biopsies;
- able to provide informed consent (written and verbal) in English;
- considered by their urologist that involvement in the research programme/involvement in the study is considered to be their best interests, for appropriate selection of patients who are likely to benefit from treatment;
- prepared and able to provide a specimen of ejaculate for monitoring before biopsy and at the completion of treatment, just before RP;
- willing to provide urine and serum samples prior to biopsy and again before RP;
- able to lie flat and willing to undergo mpMRI/MRS scanning on two occasions during the trial period with no previous or current history of claustrophobia;
- normal fasting blood glucose, kidney and liver function tests;

6.2.3.2.2 – Exclusion criteria

- previous history of head injury, dementia, psychiatric illness or concurrent cancer;
- regular administration of any lipid-lowering medication or blood-glucose lowering drugs
- prior experience of any adverse effects with lipid or glucose-lowering drugs;
- currently taking drugs known or thought to have an interaction with metformin and atorvastatin;
- previous ingestion of a 5-alpha reductase inhibitor drug;
- known co-morbidities that would contraindicate commencement of metformin or atorvastatin, such as chronic kidney disease, congestive heart failure, liver disease.

6.2.3.3 – Randomisation and allocation concealment

Following enrolment, participants are randomized using a centralized database by an external research pharmaceutical organisation (QPharm Pty Ltd), which is experienced in participation and formulation of drugs for its own and other clinical trials. The randomization is stored and concealed by QPharm, resulting in blinding of clinicians and participants to study medications. Unblinding will be permissible in the event of an adverse event resulting in cessation of medication use, with liaison between the investigators, QPharm and the participant. All medications are identical in appearance to maintain blinding during medication ingestion.

6.2.3.4 – Intervention

Participants are randomized into one of the four study groups:

1) metformin 500 mg twice a day with placebo once a day;
2) atorvastatin 20 mg once a day with placebo twice a day;
3) metformin 500 mg twice a day with atorvastatin 20 mg once a day;
4) placebo three times a day.

All capsules have been formulated by QPharm Pty Ltd to appear identical in order to comply with blinding for participants and investigators.

6.2.3.5 – Outcome measures

6.2.3.5.1 – Primary endpoint

The primary end point is serum PSA “biochemical evolution” (bEvo), defined as an increase in Serum PSA prior to surgery and following 8 weeks of treatment from baseline measurement of 20% (based on values of reported studies\textsuperscript{394,398,399}). These studies were used instead of those from larger observational studies\textsuperscript{400,401} as the observational studies had a low (25 – 30%) proportion of clinically significant CaP. The term “biochemical progression” is intended for this specific definition and different from “biochemical recurrence” commonly used in the post-treatment scenario.
Secondary endpoints that are being assessed opportunistically are biochemical progression for ejaculate-based biomarkers PCA3:PSA and citrate, defined for each biomarkers as follows:

a) PCA3:PSA measured in ejaculate: Increase from baseline measurement by 20% (based on improved all-cause and cancer-specific survival estimates\textsuperscript{402}).

b) Citrate measured in ejaculate and prostatic biopsies at the time of transperineal biopsies and the prostate ex-vivo after RP as well as MRS: Increase from baseline measurement of 30% (based on serum citrate association with aggressive prostate cancer\textsuperscript{289}).

These biochemical markers are being preferentially assessed ahead of tissue histological markers on the basis that biochemical effects precede histological change, known as the ‘field effect’ in cancer biology\textsuperscript{172}, and are thus more likely to be detected \textit{in vitro} using ejaculate and \textit{in vivo} using MRSI.

\textbf{6.2.3.5.3 – Clinical Data Collection}

Clinical data will be obtained by participating urologists and the Trial Coordinator (RM), who contacts participants by telephone on at least two occasions during the trial to ensure participant satisfaction and determine the presence of any side-effects or other difficulties. The Trial Coordinator is experienced in the data collection tools being used. Data will be collected onto a CAISIS database, used in our randomized, controlled study of open and robotic prostatectomy\textsuperscript{403,404} and with which we are familiar.

\textbf{6.2.3.5.4 – Assessment and Follow-up}

Participants are being assessed for inclusion upon enrolment for prostate biopsy. Those whose biopsies do not include any Gleason 4 or 5 tumour do not receive drug and do not take any further part in the study. Participants continuing receive medication, as per randomization protocol, and are placed on the waiting list for surgery which is planned for 8 weeks after their biopsies.

Serum investigations used clinically include total PSA (Abbott Architect\textsuperscript{®} assay) and free-total PSA estimations as well as serum testosterone and metabolic screening (serum biochemical and lipoprotein profiles). Blood tests will all be performed according to
standard assays by Sullivan & Nicolaides Pathology in addition to glycosylated PSA quantified as previously described\textsuperscript{405}.

The intended duration from diagnosis and treatment is 8 weeks, with assessments performed prior to biopsies and again just before RP. mpMRI and MRS will be assessed by IB, NK and ML. Biopsy and RP slides will be reviewed by specialized uropathologists (JP-K, JDP, MLTHS) with reference to the 2014 ISUP classification\textsuperscript{406} for continuity and expert interpretation. Standard parameters for biopsies (total number of cores, number and percentage of positive cores, Gleason/ISUP score) and RP specimens (gland size, margin status, extracapsular status, pathological stage, Gleason score) are being reported. Surgical follow-up is being determined by the treating urologist with follow-up assessments performed at these appointments until 24 months post-operatively.

Technical details regarding biological sampling and storage as well as molecular and metabolomic investigations are outlined in Chapters 4 and 5.

6.2.3.6 – Sample size and statistical analysis

Sample sizes were based on a randomized selection design with the aim of achieving a 80% probability that the best schedule produced the highest observed response rate\textsuperscript{407-409}. We took the expected baseline freedom from biochemical progression rate for the placebo schedule to be 80% based on results reported for the serum PSA biomarker\textsuperscript{399}. We estimate that we need to study 20 participants per schedule, to have a 80% probability of selecting the schedule that has a true freedom from biochemical progression rate that is at least 15% higher assuming this is the minimal practically significant difference. This calculation can be verified through the online calculator from the Centre for Clinical Research and Biostatistics of the Chinese University of Hong Kong (https://www2.ccrb.cuhk.edu.hk/stat/phase2/Randomized.htm). We were unable to suitably estimate the proportion of men who would be enrolled with a negative biopsy result and so are continuing to recruit until the sample size is achieved.

Differences in categorical variables between groups will be tested using chi squared analysis while continuous variables will be compared using two-sample t tests. Univariate analysis will be performed using the binary bEvo variable as the outcome and intervention schedule category as the predictor using a GLM with a Poisson family, log link and robust error variances in order to generate the relative risk. Multivariable analysis will also be performed using a similar GLM to determine the independence of the schedule of
treatment from other relevant explanatory variables. The latter will be included in the multivariable model if they are found to be associated with outcome (P<0.2) on univariate analysis. Link specification will be tested using the squared linear predictor while the goodness of fit of the model will be evaluated by the ability of the linear predictor from the model to classify participants into progressors and non-progressors (i.e., its predictive performance) will be evaluated using the C statistic, a term equivalent to the area under a receiver operating characteristic curve for the dichotomous outcome (bEvo). All analysis will be done using Stata (StataCorp, College Station, TX) and P<0.05 will be the threshold for significance.

6.2.4 Discussion

Since commencing the study in March 2015, 9 participants have been recruited and 2 have completed the treatment and assessment protocols. One man with a mpMRI PI-RADS 4 lesion did not have CaP detected with his biopsies. As a consequence his participation in the trial was terminated. Compared with recruitment for altruistic ejaculate donation by patients without likelihood of personal benefit, recruitment for this study, which includes ejaculate donation as an essential requisite, has been much easier but considerable greater difficulty is being experienced identifying men who have not been consuming a statin regularly.
Part 4 – Conclusions
Chapter 7 – Summary and discussion

7.1 Summary

The early detection CaP is limited by an absence of biomarkers that accurately reflect clinically significant disease. Currently implemented PSA-based CaP detection method have limitations that, despite use of adjunctive biomarkers such as phi, PCA3 and TMPRSS2:ERG, require new biomarkers to accurately diagnose localised CaP. Prostatic fluid in ejaculate or as PEUW contains CaP-specific biomarkers and provides a surrogate marker for patients with CaP who may survive long enough to benefit from curative treatment, based on improved overall survival of ejaculate donors relative to age-matched patients with CaP. Biomarkers in prostatic fluid have been reported to improve diagnosis in small proof-of-concept studies with minimal validation performed to date.

This thesis investigated prostatic fluid derived from ejaculate and PEUW collected from a “high risk”, clinically applicable cohort to improve CaP diagnosis and characterisation using molecular and metabolomic detection methods. Specifically, this Thesis also sought to optimise ejaculate sample processing for metabolomics studies and evaluate the diagnostic performance of mRNA, microRNA and metabolomic markers in ejaculate and PEUW to complement serum PSA in detecting clinically significant CaP.

To optimise ejaculate sample processing for metabolomics studies, the addition of tartrate and cooling of ejaculate samples improved the stability of choline and phosphorylcholine levels reflected by NMR-based metabolite kinetics. Clinical sample collection into a sterile urine jar containing 5 mM (on-site) or 10 mM (off-site) tartrate in 20 ml PBS solution cooled to 277 K and cooled during transport until processing would result in at most a 2-3% change in choline and phosphorylcholine to facilitate sample collection off-site without significant effect on choline-based metabolites.

To evaluate the diagnostic performance of mRNA and microRNA markers in 66 ejaculate samples to complement serum PSA in detecting clinically significant CaP, a Hepsin:PCA3 ratio together with serum PSA best predicted absolute and clinically significant CaP. For 20 patients with matched mRNA and microRNA expression, serum PSA combined with miR-200c and miR-125b improved prediction of absolute CaP (AUC=0.869 vs 0.672; p<0.05), improving specificity (67%) at 90% sensitivity compared with PSA alone (11%).
Following adjustment for variable glucose concentration attributable to the initial buffer used and unbalanced metabolite (choline/phosphocholine) regulation, metabolomic analysis of ejaculate supernatants showed that metabolites best predicted low and intermediate risk CaP with grouping observed between these groups and benign and intermediate risk samples. Lipoproteins dominated spectra of high grade samples with less metabolite contributions. Overall CaP prediction using metabolites described in previous studies was not validated.

To establish that PEUW contains prostatic-specific biomarkers, prostatic cell RNA isolation, amplification and qPCR for β2M, PSA, PCA3 and Hepsin was possible for 38 PEUW samples, with corresponding ejaculate markers. PCA3:PSA in ejaculate (AUC=0.668) and PEU (AUC=0.569) were comparable with serum PSA (AUC=0.617) in predicting absolute CaP status, which may be due to a dominance of normal prostatic cells over malignant cells and help identify patients without csCaP. To corroborate these findings, NMR-based metabolomics analysis of PEUW demonstrated that PEUW contains prostate-specific biomarkers attributable to the presence of ejaculate, both in young healthy volunteers and at-risk patients. Kinetics methodology confirmed these metabolites to behave kinetically similar to those observed in ejaculate. Initial metabolomic analysis of PEUW demonstrated best discrimination between low and intermediate risk groups as well as benign samples, but was complicated by the presence of borate-citrate adducts. Following adjusted sample preparation and data acquisition, spectral alignment and statistical analysis is ongoing.

A published study protocol for a phase II “window” randomised controlled clinical trial that commenced in March 2015 described how these in vitro NMR-based metabolomic findings may translate to in vivo metabolite quantification using MRSI. These combined NMR-based approaches may serve to better triage patients for biopsy or in active surveillance cohorts and monitor altered tumorigenesis by metabolically restricting drugs, such as metformin and atorvastatin.
7.2 Future directions

The promising methods and early performance of prostatic fluid biomarkers presented within this Thesis provide a strong platform from which to expand future research. Suitable research topics include:

1. Completion of the PEUW metabolomics analysis following measurement using sodium azide. With improved sample preservation and measurement at 900 MHz for improved sensitivity and spectral resolution, metabolites and/or relationships not previously described may be used to improve diagnosis and characterisation of CaP.

2. Confirmation of the reproducibility, feasibility and diagnostic performance of biomarkers contained in prostatic cells in ejaculate and PEUW as well as metabolomic markers in ejaculate and PEUW supernatants in independent cohorts.

3. Once diagnostic performance reliability is established, sample collection protocols can be tailored for practical use as a screening test with incorporated optimization methods (such as that presented in Section 3.2). This research would encourage use of these methods by other researchers to advance this novel approach.

4. Evaluation of genetic markers in the cell-free fraction of ejaculate, predominantly contained in exosomes, or prostasomes, which have been shown by other investigators to contain durable RNA species for biomarker quantification in other biofluids. Accurate description of these RNA species may provide more accurate biomarkers and insight into the underlying pathogenesis of CaP.

5. Matched analysis of ejaculate and PEUW to provide a “one stop shop” for CaP characterisation of low and high risk tumours, given both biofluids contain valuable, independent markers for assessing both disease states. A streamlined method for accurate non-invasive CaP diagnosis would be invaluable to clinicians in patients.

6. Matched ejaculate, prostate tissue and prostate MRSI analysis of metabolites contained in tumour foci with adjacent non-cancerous tissue to determine which metabolic changes may be monitored in vitro or in vivo to predict CaP prior to histological transformation. Prediction and/or observation of pre-malignant changes using non-invasive methods would allow for more accurate diagnosis at an early state to facilitate successful curative treatment.

7. Metabolic pathway analysis to determine the aetiology of initial prostatic malignant transformation and identify targets for imaging or therapeutics for early treatment and possible prevention of CaP in young men.


effectiveness and cost-effectiveness of the PROGENSA(R) prostate cancer antigen
3 assay and the Prostate Health Index in the diagnosis of prostate cancer: a
systematic review and economic evaluation. *Health Technol Assess* 2015; **19**(87): i-
xxxi, 1-191.

12. Roberts MJ, Richards RS, Gardiner RA, Selth LA. Seminal fluid: a useful source of


14. Gardiner RA, Samaratunga ML, Gwynne RA, Clague A, Seymour GJ, Lavin MF.
Abnormal prostatic cells in ejaculates from men with prostatic cancer--a preliminary

performance of expression of PCA3, Hepsin and miR biomarkers in ejaculate in
combination with serum PSA for the detection of prostate cancer. *Prostate* 2015;
**75**(5): 539-549.

seminal fluid as a source of prostate cancer-specific microRNA biomarkers. *Endocr

Survival Outcomes for Men Who Provided Ejaculate Specimens for Prostate
200-206.

18. Chew KK, Gibson N, Sanfilippo F, Stuckey B, Bremner A. Cardiovascular mortality
in men with erectile dysfunction: increased risk but not inevitable. *J Sex Med* 2011;
**8**(6): 1761-1771.

19. Fiehn O. Metabolomics--the link between genotypes and phenotypes. *Plant Mol Biol*

to early and noninvasive prostate cancer detection. *Korean J Urol* 2011; **52**(2): 79-
89.

21. Clarke RA, Schirra HJ, Catto JW, Lavin MF, Gardiner RA. Markers for detection of

22. Kline EE, Treat EG, Averna TA, Davis MS, Smith AY, Sillerud LO. Citrate
concentrations in human seminal fluid and expressed prostatic fluid determined via


62. Averna TA, Kline EE, Smith AY, Sillerud LO. A decrease in 1H nuclear magnetic resonance spectroscopically determined citrate in human seminal fluid


201. Lucarelli G, Fanelli M, Larocca AM, Germinario CA, Rutigliano M, Vavallo A et al. Serum sarcosine increases the accuracy of prostate cancer detection in patients with total serum PSA less than 4.0 ng/mL. *Prostate* 2012; **72**(15): 1611-1621.


300. Riches SF, Payne GS, Morgan VA, Dearnaley D, Morgan S, Partridge M et al. Multivariate modelling of prostate cancer combining magnetic resonance derived...


Appendices

Appendix 1 – Chapter 3
<table>
<thead>
<tr>
<th>Subject</th>
<th>Temperature (K)</th>
<th>[Tartrate] (mM)</th>
<th>Metabolite</th>
<th>a</th>
<th>Standard Error</th>
<th>b</th>
<th>Standard error</th>
<th>c</th>
<th>Standard error</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>279</td>
<td>0</td>
<td>P-Cho</td>
<td>38103858.4230</td>
<td>304721.3457</td>
<td>0.1352</td>
<td>0.0059</td>
<td>1317449.2016</td>
<td>164916.5174</td>
<td>0.9996</td>
</tr>
<tr>
<td>1</td>
<td>279</td>
<td>1</td>
<td>P-Cho</td>
<td>53747867.2310</td>
<td>1183268.5773</td>
<td>0.0183</td>
<td>0.0009</td>
<td>-</td>
<td>-</td>
<td>0.9842</td>
</tr>
<tr>
<td>1</td>
<td>279</td>
<td>3</td>
<td>P-Cho</td>
<td>48264917.8107</td>
<td>8906090.8846</td>
<td>0.0078</td>
<td>0.0018</td>
<td>3928866.5777</td>
<td>9024694.5405</td>
<td>0.9978</td>
</tr>
<tr>
<td>1</td>
<td>279</td>
<td>5</td>
<td>P-Cho</td>
<td>53801463.6366</td>
<td>115913.7570</td>
<td>0.0042</td>
<td>6.2466E-005</td>
<td>-</td>
<td>-</td>
<td>0.9983</td>
</tr>
<tr>
<td>1</td>
<td>279</td>
<td>10</td>
<td>P-Cho</td>
<td>52460768.9812</td>
<td>235701.9838</td>
<td>0.0019</td>
<td>0.0001</td>
<td>-</td>
<td>-</td>
<td>0.9676</td>
</tr>
<tr>
<td>1</td>
<td>279</td>
<td>20</td>
<td>P-Cho</td>
<td>55062996.4100</td>
<td>335650.4872</td>
<td>0.0011</td>
<td>0.0002</td>
<td>-</td>
<td>-</td>
<td>0.8442</td>
</tr>
<tr>
<td>1</td>
<td>279</td>
<td>0</td>
<td>Cho</td>
<td>40294692.2026</td>
<td>850908.5798</td>
<td>0.1366</td>
<td>0.0160</td>
<td>33301468.6564</td>
<td>735503.2491</td>
<td>0.9969</td>
</tr>
<tr>
<td>1</td>
<td>279</td>
<td>1</td>
<td>Cho</td>
<td>69451615.9273</td>
<td>21770972.9779</td>
<td>0.0139</td>
<td>0.0067</td>
<td>18501941.1790</td>
<td>2008376.6902</td>
<td>0.9700</td>
</tr>
<tr>
<td>1</td>
<td>279</td>
<td>3</td>
<td>Cho</td>
<td>51577287.9334</td>
<td>8475533.1906</td>
<td>0.0080</td>
<td>0.0017</td>
<td>15402977.9459</td>
<td>253286.7039</td>
<td>0.9981</td>
</tr>
<tr>
<td>1</td>
<td>279</td>
<td>5</td>
<td>Cho</td>
<td>62074627.6958</td>
<td>16849787.7718</td>
<td>0.0041</td>
<td>0.0013</td>
<td>15641323.1738</td>
<td>131248.0263</td>
<td>0.9972</td>
</tr>
<tr>
<td>1</td>
<td>279</td>
<td>10</td>
<td>Cho</td>
<td>157615153.4223</td>
<td>4059239458.7058</td>
<td>7.9614E-005</td>
<td>0.0021</td>
<td>14631253.4209</td>
<td>118322.0284</td>
<td>0.9970</td>
</tr>
<tr>
<td>2</td>
<td>279</td>
<td>20</td>
<td>Cho</td>
<td>1737122804.2224</td>
<td>18200186019.8261</td>
<td>4.2918E-005</td>
<td>0.0045</td>
<td>15204230.9720</td>
<td>154147.3216</td>
<td>0.9857</td>
</tr>
<tr>
<td>2</td>
<td>279</td>
<td>0</td>
<td>P-Cho</td>
<td>2402151.0406</td>
<td>34405.3437</td>
<td>0.2313</td>
<td>0.0570</td>
<td>466311.7839</td>
<td>17375.6995</td>
<td>0.9985</td>
</tr>
<tr>
<td>2</td>
<td>279</td>
<td>1</td>
<td>P-Cho</td>
<td>10351138.8972</td>
<td>60594.8648</td>
<td>0.0413</td>
<td>0.0006</td>
<td>912672.9738</td>
<td>57676.6990</td>
<td>0.9999</td>
</tr>
<tr>
<td>2</td>
<td>279</td>
<td>3</td>
<td>P-Cho</td>
<td>9799443.1905</td>
<td>330835.4009</td>
<td>0.0169</td>
<td>0.0009</td>
<td>2112304.5491</td>
<td>346770.0855</td>
<td>0.9994</td>
</tr>
<tr>
<td>2</td>
<td>279</td>
<td>5</td>
<td>P-Cho</td>
<td>7630811.2090</td>
<td>1059172.4069</td>
<td>0.0103</td>
<td>0.0019</td>
<td>3915456.4594</td>
<td>1083182.2525</td>
<td>0.9975</td>
</tr>
<tr>
<td>2</td>
<td>279</td>
<td>10</td>
<td>P-Cho</td>
<td>12242928.4153</td>
<td>61285.9509</td>
<td>0.0034</td>
<td>0.0001</td>
<td>-</td>
<td>-</td>
<td>0.9855</td>
</tr>
<tr>
<td>2</td>
<td>279</td>
<td>20</td>
<td>P-Cho</td>
<td>11882917.6267</td>
<td>90290.2163</td>
<td>0.0010</td>
<td>0.0002</td>
<td>0.02241</td>
<td>-204.7, 204.7</td>
<td>0.7485</td>
</tr>
<tr>
<td>2</td>
<td>279</td>
<td>0</td>
<td>Cho</td>
<td>3547237.5526</td>
<td>183893.5199</td>
<td>0.0927</td>
<td>0.0145</td>
<td>17503996.6498</td>
<td>150393.4192</td>
<td>0.9824</td>
</tr>
<tr>
<td>2</td>
<td>279</td>
<td>1</td>
<td>Cho</td>
<td>12156586.0058</td>
<td>133578.7990</td>
<td>0.0277</td>
<td>0.0006</td>
<td>10644222.9645</td>
<td>42896.4762</td>
<td>0.9998</td>
</tr>
<tr>
<td>2</td>
<td>279</td>
<td>3</td>
<td>Cho</td>
<td>20717430.0584</td>
<td>2934789.6702</td>
<td>0.0073</td>
<td>0.0013</td>
<td>8208371.8203</td>
<td>68898.6172</td>
<td>0.9999</td>
</tr>
<tr>
<td>2</td>
<td>279</td>
<td>5</td>
<td>Cho</td>
<td>23076839.1354</td>
<td>8205266.2361</td>
<td>0.0039</td>
<td>0.0015</td>
<td>7390693.3729</td>
<td>52808.9642</td>
<td>0.9985</td>
</tr>
<tr>
<td>2</td>
<td>279</td>
<td>10</td>
<td>Cho</td>
<td>8566587.4940</td>
<td>1857210.3462</td>
<td>0.0082</td>
<td>0.0023</td>
<td>7443275.9106</td>
<td>54665.4962</td>
<td>0.9966</td>
</tr>
<tr>
<td>2</td>
<td>279</td>
<td>20</td>
<td>Cho</td>
<td>4231653.7028</td>
<td>944654.6186</td>
<td>0.0128</td>
<td>0.0042</td>
<td>7019873.1039</td>
<td>69115.5093</td>
<td>0.9886</td>
</tr>
<tr>
<td>3</td>
<td>279</td>
<td>0</td>
<td>Pho</td>
<td>647152.9466</td>
<td>3592.2321</td>
<td>0.1344</td>
<td>0.0013</td>
<td>12506.0315</td>
<td>775.2215</td>
<td>0.9985</td>
</tr>
<tr>
<td>3</td>
<td>279</td>
<td>1</td>
<td>Pho</td>
<td>857404.2570</td>
<td>4823.9129</td>
<td>0.0226</td>
<td>0.0003</td>
<td>17987.6467</td>
<td>5737.3078</td>
<td>0.9994</td>
</tr>
<tr>
<td>3</td>
<td>279</td>
<td>3</td>
<td>Pho</td>
<td>799722.6348</td>
<td>59567.6846</td>
<td>0.0097</td>
<td>0.0011</td>
<td>74182.6003</td>
<td>62065.2167</td>
<td>0.9927</td>
</tr>
<tr>
<td>3</td>
<td>279</td>
<td>5</td>
<td>Pho</td>
<td>625938.9961</td>
<td>39313.7092</td>
<td>0.0081</td>
<td>0.0007</td>
<td>218060.6912</td>
<td>40449.7390</td>
<td>0.9968</td>
</tr>
<tr>
<td>3</td>
<td>279</td>
<td>10</td>
<td>Pho</td>
<td>514495.2618</td>
<td>82413.6554</td>
<td>0.0060</td>
<td>0.0012</td>
<td>319664.5811</td>
<td>83737.4398</td>
<td>0.9904</td>
</tr>
<tr>
<td>3</td>
<td>279</td>
<td>10(2)</td>
<td>Pho</td>
<td>409699.9401</td>
<td>23208.0808</td>
<td>0.0105</td>
<td>0.0009</td>
<td>476917.3685</td>
<td>24347.0687</td>
<td>0.9947</td>
</tr>
<tr>
<td>3</td>
<td>279</td>
<td>0</td>
<td>Cho</td>
<td>685343.2148</td>
<td>5964.4212</td>
<td>0.1265</td>
<td>0.0019</td>
<td>369913.1895</td>
<td>6084.7792</td>
<td>0.9963</td>
</tr>
<tr>
<td>3</td>
<td>279</td>
<td>1</td>
<td>Cho</td>
<td>940721.7920</td>
<td>7230.4875</td>
<td>0.0200</td>
<td>0.0003</td>
<td>201526.9689</td>
<td>2253.9355</td>
<td>0.9993</td>
</tr>
<tr>
<td>3</td>
<td>279</td>
<td>3</td>
<td>Cho</td>
<td>1201927.7331</td>
<td>115571.6732</td>
<td>0.0065</td>
<td>0.0008</td>
<td>168570.1818</td>
<td>3117.8211</td>
<td>0.9958</td>
</tr>
<tr>
<td>3</td>
<td>279</td>
<td>5</td>
<td>Cho</td>
<td>1339537.5833</td>
<td>226968.1645</td>
<td>0.0034</td>
<td>0.0006</td>
<td>158887.9521</td>
<td>1601.7431</td>
<td>0.9972</td>
</tr>
<tr>
<td>3</td>
<td>279</td>
<td>10</td>
<td>Cho</td>
<td>150800.4529</td>
<td>1504.3530</td>
<td>0.0023</td>
<td>0.0008</td>
<td>150800.4529</td>
<td>1504.3530</td>
<td>0.9956</td>
</tr>
<tr>
<td>3</td>
<td>279</td>
<td>10(2)</td>
<td>Cho</td>
<td>5547542.1445</td>
<td>8420879.2608</td>
<td>0.0006</td>
<td>0.0009</td>
<td>167662.7321</td>
<td>1655.9938</td>
<td>0.9948</td>
</tr>
<tr>
<td>1</td>
<td>298</td>
<td>0</td>
<td>P-Cho</td>
<td>22209489.7905</td>
<td>181326.6944</td>
<td>0.6761</td>
<td>0.0141</td>
<td>1332439.2205</td>
<td>47634.7365</td>
<td>0.9995</td>
</tr>
<tr>
<td>1</td>
<td>298</td>
<td>3</td>
<td>P-Cho</td>
<td>45101043.6580</td>
<td>183461.0530</td>
<td>0.0764</td>
<td>0.0005</td>
<td>-</td>
<td>-</td>
<td>0.9998</td>
</tr>
<tr>
<td>1</td>
<td>298</td>
<td>5</td>
<td>P-Cho</td>
<td>52394228.0472</td>
<td>265866.5087</td>
<td>0.0511</td>
<td>0.0005</td>
<td>-</td>
<td>-</td>
<td>0.9994</td>
</tr>
<tr>
<td>1</td>
<td>298</td>
<td>10</td>
<td>P-Cho</td>
<td>50545590.4709</td>
<td>268682.6090</td>
<td>0.0250</td>
<td>0.0004</td>
<td>-</td>
<td>-</td>
<td>0.9982</td>
</tr>
<tr>
<td>1</td>
<td>298</td>
<td>20</td>
<td>P-Cho</td>
<td>48454768.5224</td>
<td>274597.5368</td>
<td>0.0113</td>
<td>0.0004</td>
<td>-</td>
<td>-</td>
<td>0.9909</td>
</tr>
<tr>
<td>1</td>
<td>298</td>
<td>0</td>
<td>Cho</td>
<td>22643267.4205</td>
<td>435491.3116</td>
<td>0.5228</td>
<td>0.0240</td>
<td>36225005.6536</td>
<td>426319.4110</td>
<td>0.9975</td>
</tr>
<tr>
<td>1</td>
<td>298</td>
<td>3</td>
<td>Cho</td>
<td>43345377.2891</td>
<td>348251.7688</td>
<td>0.0829</td>
<td>0.0018</td>
<td>14807784.4947</td>
<td>218052.9035</td>
<td>0.9997</td>
</tr>
<tr>
<td>1</td>
<td>298</td>
<td>5</td>
<td>Cho</td>
<td>47537726.3565</td>
<td>356051.3724</td>
<td>0.0551</td>
<td>0.0010</td>
<td>14492313.5208</td>
<td>134309.0114</td>
<td>0.9999</td>
</tr>
<tr>
<td>1</td>
<td>298</td>
<td>10</td>
<td>Cho</td>
<td>49783816.6239</td>
<td>1278681.0883</td>
<td>0.0249</td>
<td>0.0009</td>
<td>12383644.0943</td>
<td>74272.1601</td>
<td>0.9999</td>
</tr>
<tr>
<td>1</td>
<td>298</td>
<td>20</td>
<td>Cho</td>
<td>36091259.8225</td>
<td>3927866.1897</td>
<td>0.0180</td>
<td>0.0025</td>
<td>11103210.8580</td>
<td>102377.3865</td>
<td>0.9995</td>
</tr>
<tr>
<td>2</td>
<td>298</td>
<td>0</td>
<td>P-Cho</td>
<td>1452898.6810</td>
<td>Infinity</td>
<td>53.4403</td>
<td>0.0004</td>
<td>406988.7864</td>
<td>Infinity</td>
<td>0.9414</td>
</tr>
<tr>
<td>2</td>
<td>298</td>
<td>3</td>
<td>P-Cho</td>
<td>9693784.3910</td>
<td>100662.8214</td>
<td>0.1588</td>
<td>0.0041</td>
<td>761092.0954</td>
<td>63548.3961</td>
<td>0.9992</td>
</tr>
<tr>
<td>2</td>
<td>298</td>
<td>10</td>
<td>P-Cho</td>
<td>11614260.4769</td>
<td>123931.4720</td>
<td>0.0399</td>
<td>0.0009</td>
<td>-</td>
<td>-</td>
<td>0.9965</td>
</tr>
<tr>
<td>2</td>
<td>298</td>
<td>20</td>
<td>P-Cho</td>
<td>11393037.0701</td>
<td>87152.9838</td>
<td>0.0192</td>
<td>0.0006</td>
<td>-</td>
<td>-</td>
<td>0.9941</td>
</tr>
<tr>
<td>2</td>
<td>298</td>
<td>0</td>
<td>Cho</td>
<td>2505535.8946</td>
<td>103763.4229</td>
<td>0.3087</td>
<td>0.0343</td>
<td>14911955.3622</td>
<td>98451.2428</td>
<td>0.9899</td>
</tr>
<tr>
<td>2</td>
<td>298</td>
<td>3</td>
<td>Cho</td>
<td>9345535.5600</td>
<td>82844.1713</td>
<td>0.1767</td>
<td>0.0038</td>
<td>8574215.7630</td>
<td>80045.7208</td>
<td>0.9995</td>
</tr>
<tr>
<td>2</td>
<td>298</td>
<td>10</td>
<td>Cho</td>
<td>10459737.4442</td>
<td>260422.9935</td>
<td>0.0563</td>
<td>0.0029</td>
<td>6053031.7447</td>
<td>75816.5546</td>
<td>0.9993</td>
</tr>
<tr>
<td>2</td>
<td>298</td>
<td>20</td>
<td>Cho</td>
<td>10670687.8822</td>
<td>727400.8455</td>
<td>0.0266</td>
<td>0.0026</td>
<td>5698269.4567</td>
<td>44069.8900</td>
<td>0.9994</td>
</tr>
<tr>
<td>3</td>
<td>298</td>
<td>0</td>
<td>Pho</td>
<td>647152.9466</td>
<td>3592.2321</td>
<td>0.1344</td>
<td>0.0013</td>
<td>12506.0315</td>
<td>775.2215</td>
<td>0.9985</td>
</tr>
<tr>
<td>3</td>
<td>298</td>
<td>1</td>
<td>Pho</td>
<td>857404.2570</td>
<td>4823.9129</td>
<td>0.0226</td>
<td>0.0003</td>
<td>17987.6467</td>
<td>5737.3078</td>
<td>0.9994</td>
</tr>
<tr>
<td>3</td>
<td>298</td>
<td>3</td>
<td>Pho</td>
<td>4920809.8736</td>
<td>6499.5874</td>
<td>0.0360</td>
<td>0.0000</td>
<td>7.1377E-005</td>
<td>-</td>
<td>0.9998</td>
</tr>
<tr>
<td>3</td>
<td>298</td>
<td>5</td>
<td>Pho</td>
<td>4983259.1449</td>
<td>6071.7653</td>
<td>0.0269</td>
<td>0.0005</td>
<td>5.3128E-005</td>
<td>-</td>
<td>0.9998</td>
</tr>
<tr>
<td>3</td>
<td>298</td>
<td>10</td>
<td>Pho</td>
<td>514405.2618</td>
<td>82413.6554</td>
<td>0.0060</td>
<td>0.0012</td>
<td>319664.5811</td>
<td>83737.4398</td>
<td>0.9904</td>
</tr>
<tr>
<td>3</td>
<td>298</td>
<td>10(2)</td>
<td>Pho</td>
<td>4969350.5641</td>
<td>6859.0618</td>
<td>0.0148</td>
<td>0.0005</td>
<td>4.4646E-005</td>
<td>-</td>
<td>0.9994</td>
</tr>
<tr>
<td>3</td>
<td>298</td>
<td>0</td>
<td>Cho</td>
<td>3823184.8682</td>
<td>42122.8310</td>
<td>0.3538</td>
<td>0.0062</td>
<td>1421822.1890</td>
<td>42264.4416</td>
<td>0.9945</td>
</tr>
<tr>
<td>Time (s)</td>
<td>Temp (°C)</td>
<td>Phosphorylcholine (mM)</td>
<td>Phosphorylcholine Hydrolysis (mg/l)</td>
<td>Choline (mM)</td>
<td>Phosphorylcholine Hydrolysis (mg/l)</td>
<td>R² Value</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>-----------</td>
<td>------------------------</td>
<td>-----------------------------------</td>
<td>-------------</td>
<td>-------------------------------</td>
<td>----------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>298</td>
<td>Cho</td>
<td>4435188.8130</td>
<td>17698.2542</td>
<td>0.0900</td>
<td>951199.7732</td>
<td>18364.5560</td>
<td>0.9991</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>298</td>
<td>Cho</td>
<td>4839361.1861</td>
<td>15420.9763</td>
<td>0.0335</td>
<td>763969.1487</td>
<td>13116.1839</td>
<td>0.9994</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>298</td>
<td>Cho</td>
<td>4992300.0244</td>
<td>25038.6922</td>
<td>0.0242</td>
<td>778759.5921</td>
<td>11795.1999</td>
<td>0.9994</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>298</td>
<td>Cho</td>
<td>5193269.4087</td>
<td>64792.7117</td>
<td>0.0131</td>
<td>811892.5964</td>
<td>7873.1410</td>
<td>0.9995</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>298</td>
<td>Cho</td>
<td>5376125.0250</td>
<td>67547.8058</td>
<td>0.0124</td>
<td>754768.7830</td>
<td>7198.1964</td>
<td>0.9996</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table A1.1:** Calculated apparent first-order kinetics exponential fit parameters for spectral intensities for phosphorycholyne-choline hydrolysis. The equation used to model the decay of phosphorylcholine was: \( f(t) = a \cdot e^{(-b \cdot t)} + c \). A # symbol after the R² value denotes a model that would not converge as a 3 or 2 parameter model. The equation used to model the production of choline was: \( f(t) = a \cdot (1 - e^{(-b \cdot t)}) + c \).
<table>
<thead>
<tr>
<th>Metabolite</th>
<th>Chemical shift range (ppm)</th>
<th>Number of Protons per signal</th>
<th>Subject 1 (Mean ± StDev)</th>
<th>Subject 2 (Mean ± StDev)</th>
<th>Subject 3 (Mean ± StDev)</th>
<th>Absolute Molar Concentration (mM)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>298 K ( n=6 )</td>
<td>279 K ( n=6 )</td>
<td>279 K ( n=6 )</td>
<td>298 K ( n=5 ) ( n=4 )</td>
</tr>
<tr>
<td>Choline</td>
<td>CHEBI:15354</td>
<td>3.186 – 3.206</td>
<td>9</td>
<td>0.53 ± 0.33</td>
<td>0.25 ± 0.10</td>
<td>0.13 ± 0.01</td>
</tr>
<tr>
<td>Phosphorylcholine</td>
<td>CHEBI:18132</td>
<td>3.206 – 3.216</td>
<td>9</td>
<td>0.82 ± 0.31</td>
<td>0.73 ± 0.04</td>
<td>0.10 ± 0.04</td>
</tr>
<tr>
<td>Glycerophosphorylcholine</td>
<td>CHEBI:16870</td>
<td>3.216 – 3.226</td>
<td>9</td>
<td>0.26 ± 0.04</td>
<td>0.19 ± 0.005</td>
<td>0.10 ± 0.04</td>
</tr>
<tr>
<td>Citrate</td>
<td>CHEBI:30769</td>
<td>2.486 – 2.706</td>
<td>4 (sum)</td>
<td>1.58 ± 0.15</td>
<td>1.08 ± 0.02</td>
<td>2.25 ± 0.18</td>
</tr>
<tr>
<td>Lactate</td>
<td>CHEBI:24996</td>
<td>1.306 – 1.336</td>
<td>3</td>
<td>0.51 ± 0.20</td>
<td>0.25 ± 0.02</td>
<td>0.45 ± 0.31</td>
</tr>
<tr>
<td>Spermine</td>
<td>CHEBI:15746</td>
<td>1.776 – 1.816</td>
<td>4</td>
<td>0.31 ± 0.03</td>
<td>0.20 ± 0.008</td>
<td>0.44 ± 0.03</td>
</tr>
<tr>
<td>Fructose</td>
<td>CHEBI:28645</td>
<td>3.976 – 4.036</td>
<td>1.45</td>
<td>1.63 ± 0.46</td>
<td>1.14 ± 0.04</td>
<td>0.56 ± 0.10</td>
</tr>
<tr>
<td>Tartrate (reference)</td>
<td>CHEBI:15671</td>
<td>4.296 – 4.346</td>
<td>2</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>

**Table A1.2:** Descriptive statistics for calculated concentrations of key metabolites identified in the multivariate statistical analysis (Figure 3.6) using tartrate as a concentration standard. The number of spectra included \( (n) \) for each subject at each temperature are provided. CHEBI – Chemical Entities of Biological Interest.
Figure A1.1: Representative 1D NOESY spectra of human ejaculate (Subject 3). Panel a – Single 1D NOESY spectrum (range -2 – 12 ppm) containing 1 mM tartrate acquired at t=0, with the segment to the left of the excluded water signal (5.1 – 4.5 ppm) up scaled by a factor of eight to improve metabolite visualisation. Panel b – Stacked plot showing 1D NOESY spectra containing 10 mM tartrate, acquired at 2 hour intervals across a 60-hour period. The figure shows the stability of the metabolite signals provided by the addition of tartrate.
Figure A1.2: Time-dependent phosphorylcholine (PCho) hydrolysis to choline (Cho) in ejaculate at 277 K with 1 mM tartrate characterised by NMR spectroscopy. Panel a – Stacked plot showing 1D NOESY spectra in the range 2.3 – 4.4 ppm with major metabolites labelled, acquired at various times (7.2, 11.4 and 23.5 h) after initial acquisition (0 h) demonstrating changes in peak intensities of phosphorylcholine (decrease) and choline (increase) over time. Panel b – Section of a pseudo-2D NOESY spectrum demonstrating changes in choline (Cho – increase), phosphorylcholine (PCho – decrease) and glycerophosphorylcholine (G-P-Cho – no change) metabolite peaks over time.
Figure A1.3: Panel a - Overlay of initial (t=0) 1D NOESY NMR spectra acquired from ejaculate samples at 298 K with varying tartrate concentrations (yellow = 0 mM, blue = 1 mM, red = 3 mM, green = 5 mM, purple = 10 mM, orange = 20 mM) over the region 2.4 – 4.4 ppm, demonstrating changes in initial choline (Cho) and phosphorylcholine (PCho) peak heights. The remaining metabolites, glycerophosphorylcholine (G-P-Cho), fructose, citrate and spermine, demonstrate constant peak heights. Panel b – Relative intensity change in phosphorylcholine peak at 3.21 ppm at different temperatures (298 K = squares; 279 K = triangles) after 24 hours for three subjects (blue, red, gold).
Figure A1.4: Relative expression of RNA biomarkers (PCA3, PSA and Hepsin normalized to β2M) in LNCaP cell culture and phosphate buffered saline (PBS - black). Expression levels are unchanged in the presence of 20 mM tartrate (dark grey), 0.05% sodium azide (antimicrobial sample preservative – light grey) and a combination of tartrate and sodium azide (white). * $p > 0.05$.

Figure A1.5: Apparent activation energies ($E_A$) (right panel) calculated from linear fits of apparent rate constants of phosphorylcholine hydrolysis (measured in s$^{-1}$) for each subject at 277 K, 279 K and 298 K in an Arrhenius plot (left panel). Data for all volunteers yield similar slopes in the Arrhenius plot and thus also similar apparent activation energies for PAP. Black circles: Subject 1, open circles: Subject 2, black triangles: Subject 3.
### Table A2.1: Stepwise logistic regression results incorporating serum PSA, Hepsin, PCA3 and Hepsin:PCA3 (n = 66; A) and additional analysis incorporating miRNAs (n = 20; B). All markers available for entry in the model; \( p \) to enter 0.3; \( p \) to remove 0.35; Univariate composite scores based on MLR modelling were created as follows: Composite1 = (Serum PSA) – (Hepsin:PCA3 / 2), Composite2 = (Serum PSA) + (miR-200c*5.5) – (miR-125b*2).
Figure A2.1 – Comparison plot of $Ct_{Av}$ vs. $Ct_{SD}$ plots for qPCR of PSA, Hepsin and PCA3. This plot was used to determine a suitable cut off at which sample numbers and quantification were optimized, specifically $Ct_{Av} > 35$ and $Ct_{SD} > 1.5$. 
Figure A2.2 – Flowchart demonstrating loss of samples for consideration in final analysis from initial cohort. These exclusions were the result of rigorous consideration based on experimental and statistical factors.
Figure A2.3 – Predictive probability plots illustrating the relationship between the composite scores (Composite 1 – A; Composite 2 – B; x-axis) and the prediction of the binary variable based on the multivariate logistic regression (y-axis). The formulae used to calculate the composite scores, based on the regression coefficients, are indicated in the x-axis label for each figure.
Figure A2.4—Supervised partial least squares multivariate statistical analysis of ejaculate samples with different predictive classifiers against benign samples. Analysis of high risk (red triangles) versus benign (blue squares) samples (M7; Panels a, b), intermediate risk (yellow diamonds) versus benign samples (M8, Panels c, d), negative/low risk (green circles) versus benign samples (M9, Panels e, f), and of samples with presence (filled squares) and absence (empty squares) of primary Gleason pattern 4 or 5 on radical prostatectomy histology (M10, Panels g,h) did not show significant group separation among the invalid models. Scores plots (panels a, c, e, g) show sample groupings, and the loadings plots (b, d, f, h) show responsible metabolite signals. Metabolite abbreviations: Ac = acetate; Cit = citrate; Cre = creatine; Lac = lactate; Gly = glycine; Lip = lipids/lipoproteins; Lys = lysine; PG = propylene glycol; Pyr = pyruvate; Spe = spermine; Uri = uridine; UC1 = unknown compound 1.
Figure A2.5– Supervised partial least squares multivariate statistical analysis of ejaculate samples containing only PBS buffer, with different predictive sample classifiers compared with benign samples. Analysis of csCaP versus benign samples (M12; Panels a,b), low versus intermediate risk (M13; Panels c,d), intermediate risk versus benign (M14; Panels e,f), high risk versus benign (M15; Panels g,h), and all risk groups (M16; Panels i,j) were similar to those seen in the main cohort. Scores plots (panels a, c, e, g, i) show sample groupings, and loadings plots (b, d, f, h, j) show the responsible metabolite peaks. Blue squares = benign; Green circles = negative/low risk; Yellow diamonds = intermediate risk; Red triangles = high risk/cancer present. Metabolite abbreviations: Cit = citrate; Cre = creatine; Fru = fructose; Glut = glutamine; Lac = lactate; Lip = lipids/lipoproteins; Lys = lysine; PG = propylene glycol; Pyr = pyruvate; Spe = spermine; Uri = uridine.
Figure A2.6 – Unsupervised multivariate statistical analysis (principal components analysis) of valid supervised models to test validity of the supervised models. Comparison of intermediate (yellow diamonds) and low (green circles) risk based on RP histology (panels a, b) and on best available histology in samples collected in PBS (panels c, d) are shown. Comparison of low-risk and benign (blue squares) samples are shown in panels e, f. Scores plots (panels a, c, e) show sample groupings, and loadings plots (b, d, f) show the responsible metabolite peaks. Metabolite abbreviations: Cit = citrate; Fru = fructose; Glut = glutamine; Lac = lactate; Leu = leucine; Lip = lipids/lipoproteins; Lys = lysine; Pyr = pyruvate; Spe = spermine; UC2 = unknown compound 2.
Figure A2.7 – Box plots comparing lactate bucket intensity (following quantile normalisation) against risk categories (0 = benign, 1 = low, 2 = intermediate, 3 = high) for all samples (a) and samples collected in PBS only (b).
Figure A2.8 – Supervised multivariate statistical analysis (partial least squares) of samples according to presence or absence of the TMPRSS2:ERG fusion gene (M11). The scores plot (a) shows sample grouping, and the loadings plot (b) the responsible metabolite peaks. Empty boxes = negative; Filled boxes = positive. Metabolite abbreviations: Cit = citrate; Glu = glutamine; Lac = lactate; Lip = lipids/lipoproteins; Lys = lysine.
<table>
<thead>
<tr>
<th>Model</th>
<th>Type</th>
<th>Group 1</th>
<th>Group 2</th>
<th>Scaling</th>
<th>n&lt;sup&gt;a&lt;/sup&gt;</th>
<th>k&lt;sup&gt;b&lt;/sup&gt;</th>
<th>A&lt;sup&gt;c&lt;/sup&gt;</th>
<th>$R^2_X$ &lt;sup&gt;d&lt;/sup&gt;</th>
<th>$R^2_Y$ &lt;sup&gt;d&lt;/sup&gt;</th>
<th>Q&lt;sup&gt;d&lt;/sup&gt; &lt;sub&gt;2&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>PCA</td>
<td>All samples (exclude 2 outliers)</td>
<td>Par</td>
<td>149</td>
<td>907</td>
<td>6</td>
<td>0.757</td>
<td>--</td>
<td>--</td>
<td>0.624</td>
</tr>
<tr>
<td>M2</td>
<td>PCA</td>
<td>M1 after Add-to-Subtract</td>
<td>Par</td>
<td>151</td>
<td>914</td>
<td>16</td>
<td>0.885</td>
<td>--</td>
<td>0.5</td>
<td>--</td>
</tr>
<tr>
<td>M3</td>
<td>PCA</td>
<td>M2 after excluding Cho, PCho, GPCho</td>
<td>Par</td>
<td>151</td>
<td>884</td>
<td>15</td>
<td>0.874</td>
<td>--</td>
<td>--</td>
<td>0.643</td>
</tr>
<tr>
<td>M4</td>
<td>PLS</td>
<td>Benign</td>
<td>CaP</td>
<td>Par</td>
<td>151</td>
<td>884</td>
<td>2</td>
<td>0.413</td>
<td>0.114</td>
<td>-0.128</td>
</tr>
<tr>
<td>M5</td>
<td>PLS</td>
<td>Low (RP)</td>
<td>Intermed (RP)</td>
<td>Par</td>
<td>11</td>
<td>884</td>
<td>2</td>
<td>0.479</td>
<td>0.868</td>
<td>0.449</td>
</tr>
<tr>
<td>M6</td>
<td>PLS</td>
<td>Low (Bx/RP)</td>
<td>Intermed (Bx/RP)</td>
<td>Par</td>
<td>27</td>
<td>884</td>
<td>2</td>
<td>0.449</td>
<td>0.507</td>
<td>-0.193</td>
</tr>
<tr>
<td>M7</td>
<td>PLS</td>
<td>Benign</td>
<td>Intermed</td>
<td>Par</td>
<td>69</td>
<td>884</td>
<td>2</td>
<td>0.333</td>
<td>0.209</td>
<td>-0.21</td>
</tr>
<tr>
<td>M8</td>
<td>PLS</td>
<td>Benign</td>
<td>High</td>
<td>Par</td>
<td>123</td>
<td>884</td>
<td>2</td>
<td>0.418</td>
<td>0.157</td>
<td>-0.076</td>
</tr>
<tr>
<td>M9</td>
<td>PLS</td>
<td>Benign</td>
<td>Low</td>
<td>Par</td>
<td>63</td>
<td>884</td>
<td>2</td>
<td>0.398</td>
<td>0.339</td>
<td>-0.011</td>
</tr>
<tr>
<td>M10</td>
<td>PLS</td>
<td>Primary Gleason ≥4</td>
<td>Primary Gleason &lt;4</td>
<td>Par</td>
<td>61</td>
<td>884</td>
<td>2</td>
<td>0.275</td>
<td>0.263</td>
<td>-0.21</td>
</tr>
<tr>
<td>M11</td>
<td>PLS</td>
<td>PBS benign</td>
<td>PBS csCaP</td>
<td>Par</td>
<td>38</td>
<td>884</td>
<td>2</td>
<td>0.368</td>
<td>0.278</td>
<td>-0.21</td>
</tr>
<tr>
<td>M12</td>
<td>PLS</td>
<td>PBS Low (RP)</td>
<td>PBS Intermed (RP)</td>
<td>Par</td>
<td>6</td>
<td>884</td>
<td>2</td>
<td>0.555</td>
<td>0.907</td>
<td>0.165</td>
</tr>
<tr>
<td>M13</td>
<td>PLS</td>
<td>PBS Benign</td>
<td>PBS Low</td>
<td>Par</td>
<td>12</td>
<td>884</td>
<td>2</td>
<td>0.491</td>
<td>0.867</td>
<td>0.292</td>
</tr>
<tr>
<td>M14</td>
<td>PLS</td>
<td>PBS Benign</td>
<td>PBS High</td>
<td>Par</td>
<td>32</td>
<td>884</td>
<td>2</td>
<td>0.435</td>
<td>0.305</td>
<td>-0.21</td>
</tr>
<tr>
<td>M15</td>
<td>PCA</td>
<td>Low (RP)</td>
<td>Intermed (RP)</td>
<td>Par</td>
<td>11</td>
<td>884</td>
<td>2</td>
<td>0.584</td>
<td>--</td>
<td>0.296</td>
</tr>
<tr>
<td>M16</td>
<td>PCA</td>
<td>PBS Low (RP)</td>
<td>PBS Intermed (RP)</td>
<td>Par</td>
<td>6</td>
<td>884</td>
<td>2</td>
<td>0.69</td>
<td>--</td>
<td>0.143</td>
</tr>
<tr>
<td>M17</td>
<td>PCA</td>
<td>PBS Benign</td>
<td>PBS Low</td>
<td>Par</td>
<td>12</td>
<td>884</td>
<td>3</td>
<td>0.724</td>
<td>--</td>
<td>0.416</td>
</tr>
<tr>
<td>M18</td>
<td>PLS</td>
<td>TMPRSS2:ERG positive</td>
<td>TMPRSS2:ERG negative</td>
<td>Par</td>
<td>73</td>
<td>884</td>
<td>2</td>
<td>0.44</td>
<td>0.175</td>
<td>-0.21</td>
</tr>
</tbody>
</table>

Table A2.2: Figures of Merits of all MVSA models. Abbreviations: Bx = biopsy-based histology (RP histology used where available); CaP = prostate cancer; Cho = choline; csCaP = clinically significant prostate cancer; GPCho = glycerophosphocholine; Intermed = intermediate risk CaP; Par = Pareto scaling; PBS = phosphate buffered saline; PCA = principal components analysis; PCho = phosphocholine; PLS = partial least squares; RP = radical prostatectomy only histology; *Number of samples  b Number of $X$ variables  c Number of principal/latent components  

$d$ $R^2_X$ and $R^2_Y$ are the fraction of the sum of squares for the selected component representing the variance of $X$ and $Y$ variables, and $Q^2$ is the predictive ability parameter of the model, which is estimated by cross-validation. Negative $Q^2$ values in the table denote weak statistical models.
<table>
<thead>
<tr>
<th>Metabolite</th>
<th>Log base 10 of</th>
<th>Mean (± Std. Error) (mM)</th>
<th>Logistic Regression</th>
<th>ROC analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>P-value</td>
<td>Coefficient</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Std. Error</td>
<td>Std. Error</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>AUC</td>
<td>Std. Error</td>
</tr>
<tr>
<td>Prostate cancer status</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>alanine</td>
<td>0.1734 (0.0208)</td>
<td>0.7998</td>
<td>0.0291</td>
<td>2.0552</td>
</tr>
<tr>
<td>choline</td>
<td>1.3326 (0.1392)</td>
<td>0.2011</td>
<td>0.9263</td>
<td>0.555</td>
</tr>
<tr>
<td>citrate</td>
<td>2.9243 (0.2643)</td>
<td>0.8616</td>
<td>0.9631</td>
<td>0.542</td>
</tr>
<tr>
<td>creatine</td>
<td>0.1156 (0.0114)</td>
<td>0.12433</td>
<td>0.9144</td>
<td>0.629</td>
</tr>
<tr>
<td>fructose</td>
<td>3.2694 (0.2118)</td>
<td>0.2439</td>
<td>0.8559</td>
<td>0.541</td>
</tr>
<tr>
<td>glucose</td>
<td>0.5802 (0.0604)</td>
<td>0.1629</td>
<td>0.1629</td>
<td>1.4923</td>
</tr>
<tr>
<td>glutamine</td>
<td>0.2259 (0.0236)</td>
<td>0.5812</td>
<td>0.0008</td>
<td>1.8686</td>
</tr>
<tr>
<td>glycerophosphocholine</td>
<td>0.8520 (0.0645)</td>
<td>0.5879</td>
<td>0.0291</td>
<td>2.0552</td>
</tr>
<tr>
<td>lactate</td>
<td>0.4067 (0.0416)</td>
<td>0.1042</td>
<td>0.1595</td>
<td>0.592</td>
</tr>
<tr>
<td>leucine</td>
<td>0.3251 (0.0238)</td>
<td>0.1042</td>
<td>0.1595</td>
<td>0.592</td>
</tr>
<tr>
<td>myo-inositol</td>
<td>0.1810 (0.0459)</td>
<td>0.1042</td>
<td>0.1595</td>
<td>0.592</td>
</tr>
<tr>
<td>phosphocholine</td>
<td>0.3501 (0.0373)</td>
<td>0.1042</td>
<td>0.1595</td>
<td>0.592</td>
</tr>
<tr>
<td>serum PSA</td>
<td>3.2694 (0.2118)</td>
<td>0.1042</td>
<td>0.1595</td>
<td>0.592</td>
</tr>
<tr>
<td>pyruvate</td>
<td>0.3709 (0.0373)</td>
<td>0.1042</td>
<td>0.1595</td>
<td>0.592</td>
</tr>
<tr>
<td>uridine</td>
<td>0.1793 (0.0167)</td>
<td>0.1042</td>
<td>0.1595</td>
<td>0.592</td>
</tr>
<tr>
<td>valine</td>
<td>0.3206 (0.0392)</td>
<td>0.1042</td>
<td>0.1595</td>
<td>0.592</td>
</tr>
<tr>
<td>Prostate cancer risk</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>alanine</td>
<td>0.1734 (0.0208)</td>
<td>0.1629</td>
<td>0.2293</td>
<td>1.64212</td>
</tr>
<tr>
<td>choline</td>
<td>1.3326 (0.1392)</td>
<td>0.1595</td>
<td>0.1536</td>
<td>0.592</td>
</tr>
<tr>
<td>citrate</td>
<td>2.9243 (0.2643)</td>
<td>0.8808</td>
<td>0.1703</td>
<td>0.584</td>
</tr>
<tr>
<td>creatine</td>
<td>0.1156 (0.0114)</td>
<td>0.1536</td>
<td>0.584</td>
<td>0.584</td>
</tr>
<tr>
<td>fructose</td>
<td>3.2694 (0.2118)</td>
<td>0.3672</td>
<td>0.1147</td>
<td>0.584</td>
</tr>
<tr>
<td>glucose</td>
<td>0.5802 (0.0604)</td>
<td>0.1147</td>
<td>0.584</td>
<td>0.584</td>
</tr>
<tr>
<td>glutamine</td>
<td>0.2259 (0.0236)</td>
<td>0.1147</td>
<td>0.584</td>
<td>0.584</td>
</tr>
<tr>
<td>glycerophosphocholine</td>
<td>0.8520 (0.0645)</td>
<td>0.1147</td>
<td>0.584</td>
<td>0.584</td>
</tr>
<tr>
<td>lactate</td>
<td>0.4067 (0.0416)</td>
<td>0.1147</td>
<td>0.584</td>
<td>0.584</td>
</tr>
<tr>
<td>leucine</td>
<td>0.3251 (0.0238)</td>
<td>0.1147</td>
<td>0.584</td>
<td>0.584</td>
</tr>
<tr>
<td>myo-inositol</td>
<td>0.1810 (0.0459)</td>
<td>0.1147</td>
<td>0.584</td>
<td>0.584</td>
</tr>
<tr>
<td>phosphocholine</td>
<td>0.3501 (0.0373)</td>
<td>0.1147</td>
<td>0.584</td>
<td>0.584</td>
</tr>
<tr>
<td>serum PSA</td>
<td>3.2694 (0.2118)</td>
<td>0.1147</td>
<td>0.584</td>
<td>0.584</td>
</tr>
<tr>
<td>pyruvate</td>
<td>0.3709 (0.0373)</td>
<td>0.1147</td>
<td>0.584</td>
<td>0.584</td>
</tr>
<tr>
<td>uridine</td>
<td>0.1793 (0.0167)</td>
<td>0.1147</td>
<td>0.584</td>
<td>0.584</td>
</tr>
<tr>
<td>valine</td>
<td>0.3206 (0.0392)</td>
<td>0.1147</td>
<td>0.584</td>
<td>0.584</td>
</tr>
</tbody>
</table>

Table A2.3: Logistic regression weightings following targeted metabolite quantification using Chenomx, similar to that reported by\(^\text{18}\). Among 151 patients, CaP status (positive 98, negative 53) and D’Amico risk (high = 82, low = 69) were used as dependent variables (Enter, \(p_{\text{enter}} 0.05, p_{\text{exclude}} >0.1\)). *units = ng/ml
<table>
<thead>
<tr>
<th>Metabolites(^a)</th>
<th>Moieties: (\delta , ^1\text{H} \text{ in ppm (multiplicity)})</th>
</tr>
</thead>
<tbody>
<tr>
<td>acetate</td>
<td>CH(_3): 1.91 (s)</td>
</tr>
<tr>
<td>alanine</td>
<td>(\beta\text{CH}_3): 1.47(d), (\alpha\text{CH}): 3.78(q)</td>
</tr>
<tr>
<td>asparagine</td>
<td>(\beta^y\text{CH}_2): 2.84(m), (\beta^x\text{CH}_2): 2.94(m), (\alpha\text{CH}): 3.99(m)</td>
</tr>
<tr>
<td>choline</td>
<td>N-CH(_3): 3.19(s), 2-CH(_2): 3.51(dd) 1-CH(_2): 4.06(ddd)</td>
</tr>
<tr>
<td>citrate</td>
<td>CH(^y\text{H}^x): 2.53(d), 2.66(d)</td>
</tr>
<tr>
<td>creatine</td>
<td>(\gamma\text{CH}_3): 3.03(s), (\alpha\text{CH}_2): 3.92(s)</td>
</tr>
<tr>
<td>formate</td>
<td>CH: 8.4 (s)</td>
</tr>
<tr>
<td>fructose</td>
<td>3.59(m), 3.70(m), 3.82(m), 3.89(dd), 3.99(m), 4.02(dd), 4.12(m)</td>
</tr>
<tr>
<td>(\alpha)-glucose</td>
<td>5-CH: 3.41(t), 3-CH: 3.53(dd), 4-CH: 3.71(t), CH(^y\text{H}^x): 3.77(m), 3.84(m), 6-CH: 3.83(ddd), 2-CH: 5.23(d)</td>
</tr>
<tr>
<td>(\beta)-glucose</td>
<td>3-CH: 3.24(dd), 5-CH: 3.40(t), 6-CH: 3.46(dd), 4-CH: 3.49(t), CH(^y\text{H}^x): 3.73(m), 3.90(dd), 2-CH: 4.64(d)</td>
</tr>
<tr>
<td>glutamate</td>
<td>2.04(m), (\beta\text{CH}_2): 2.12 (m), (\beta\text{CH}_2): 2.34(m), (\alpha\text{CH}): 3.75(m)</td>
</tr>
<tr>
<td>glutamine</td>
<td>(\beta\text{CH}_2): 2.13(m), (\gamma\text{CH}_2): 2.45(m), (\alpha\text{CH}): 3.77(t)</td>
</tr>
<tr>
<td>glycerophosphocholine</td>
<td>N-(CH(_3))(_3): 3.22(s), 2-CH(_2): 3.61(m), glycerol-CH/CH2: 3.90(m), 1-CH(_2): 4.30(m)</td>
</tr>
<tr>
<td>glycine</td>
<td>CH(_2): 3.55 (s)</td>
</tr>
<tr>
<td>isobutyrate</td>
<td>(\beta\text{CH}_3): 1.06(d), (\alpha\text{CH}): 2.38(m)</td>
</tr>
<tr>
<td>isoleucine</td>
<td>(\delta\text{CH}_3): 0.93(t), (\gamma\text{2CH}_3): 1.00(d), (\gamma\text{CH}_2): 1.25 (m), (\gamma\text{1CH}_2): 1.46(m), (\beta\text{CH}): 1.97(m)(^c), (\alpha\text{CH}): 3.65(d)</td>
</tr>
<tr>
<td>lactate</td>
<td>(\beta\text{CH}_3): 1.32(d)(^b), (\alpha\text{CH}): 4.11(q)(^b)</td>
</tr>
<tr>
<td>leucine</td>
<td>(\delta\text{CH}_3): 0.95(d), (\delta\text{CH}_3): 0.96(d), (\beta\text{CH}_2): 1.71(m), (\gamma\text{CH}): 1.71(m), (\alpha\text{CH}): 3.73(m)</td>
</tr>
<tr>
<td>lipids/lipoproteins</td>
<td>0.84 (b), 0.9(b), 1.45(b), 1.66 (b), 2.02(b), 2.25(b), 2.35(b), 3.01(b), 3.10(b)</td>
</tr>
<tr>
<td>lysine</td>
<td>(\gamma\text{CH}_2): 1.43(m), (\delta\text{CH}_2): 1.72(m), (\beta\text{CH}_2): 1.90(m), (\varepsilon\text{CH}_2): 3.02(m), (\alpha\text{CH}): 3.75(m)</td>
</tr>
</tbody>
</table>
Table A2.4: Assignments of metabolites identified in 1D NOESY ejaculate spectra. The resonance assignment of metabolites was based on Chenomx NMR Suite 8.2 as well as reported by Lynch et al. (1994). Key to multiplicities: s, singlet; d, doublet; dd, doublet of doublets; t, triplet; q, quartet; m, multiplet; b, broad; TMAO, trimethylamine-N-oxide; UC, unknown compound.

*a The assignments of most metabolites are equivalent to Metabolomics Standard Initiative level 2 (putatively annotated compounds, i.e. assignments based upon physicochemical properties and/or spectral similarity with public/commercial spectral libraries but no internal standard used) except for UC1 & UC2 which are Metabolomics Standard Initiative level 4 (unidentified metabolites).
Figure A2.9 – Principal components multivariate statistical analysis of samples following sum of choline-containing buckets (Cho_all; choline, phosphocholine, glycerophosphocholine) and citrate (cit_all), coloured according to CaP presence (red = CaP; blue = benign). Lipids/lipoproteins (2.02, 0.90, 0.84) were also influential as seen in Chapter 4.3.
Appendix 3 – Chapter 5
Figure A3.1 – Comparison plot of average $Ct$ ($CtAv$) vs. $Ct$ standard deviation ($CtSD$) plots for qPCR of PSA, PCA3 and $\beta$2M in PEUW. This plot was used to determine a suitable cutoff at which sample numbers and quantification were optimized, specifically $CtAv > 35$ and $CtSD > 1.5$ (indicated by the red dashed line).
Figure A3.2 – Flowchart demonstrating sample loss during workflow towards final analysis from initial cohort. Experimental and statistical factors were rigorously considered, resulting in these exclusions.

- 1 = qPCR analysis (atypical melt or quantitation curves)
- 5 = insufficient RNA to reach threshold (all PCA3)
- 14 = Ct values (CtAv or CtSD) outside defined cutoffs
- 25 = corresponding ejaculate sample ineligible

83 PEUW samples measured for B2M, PSA and PCA3

38 samples remaining for analysis across all variables
<table>
<thead>
<tr>
<th>Biomarker</th>
<th>Area</th>
<th>Std. Err.</th>
<th>chi2</th>
<th>df</th>
<th>Pr&gt;chi2</th>
</tr>
</thead>
<tbody>
<tr>
<td>serumpsa (standard)</td>
<td>0.6169</td>
<td>0.0946</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>peu_psa_b2m</td>
<td>0.6000</td>
<td>0.1077</td>
<td>0.0139</td>
<td>1</td>
<td>0.9063</td>
</tr>
<tr>
<td>peu_pca3_b2m</td>
<td>0.5692</td>
<td>0.1082</td>
<td>0.0974</td>
<td>1</td>
<td>0.7550</td>
</tr>
<tr>
<td>peu_pca3_psa</td>
<td>0.4308</td>
<td>0.1098</td>
<td>1.4340</td>
<td>1</td>
<td>0.2311</td>
</tr>
<tr>
<td>ej_psa_b2m</td>
<td>0.4862</td>
<td>0.1044</td>
<td>0.8108</td>
<td>1</td>
<td>0.3679</td>
</tr>
<tr>
<td>ej_pca3_b2m</td>
<td>0.7169</td>
<td>0.1021</td>
<td>0.4912</td>
<td>1</td>
<td>0.4834</td>
</tr>
<tr>
<td>ej_pca3_psa</td>
<td>0.6677</td>
<td>0.0952</td>
<td>0.1825</td>
<td>1</td>
<td>0.6692</td>
</tr>
</tbody>
</table>

**Figure A3.3** – ROC comparison figures for each analysis, comprising PEUW (paired cohort) and Ejaculate (paired cohort) for absolute CaP. The *p*-values in the table (“Pr>chi2”) are for the difference between each biomarker area vs the reference (serum PSA). Std. Err. = standard error, chi2 = $\chi^2$, df = degrees of freedom.
### Table A3.4

<table>
<thead>
<tr>
<th>Biomarker</th>
<th>ROC Area</th>
<th>Std. Err.</th>
<th>chi2</th>
<th>df</th>
<th>Pr&gt;chi2</th>
</tr>
</thead>
<tbody>
<tr>
<td>serumpsa (standard)</td>
<td>0.6401</td>
<td>0.0909</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>peu_psa_b2m</td>
<td>0.6078</td>
<td>0.0976</td>
<td>0.0539</td>
<td>1</td>
<td>0.8165</td>
</tr>
<tr>
<td>peu_pca3_b2m</td>
<td>0.4958</td>
<td>0.0997</td>
<td>0.9638</td>
<td>1</td>
<td>0.3262</td>
</tr>
<tr>
<td>peu_pca3_psa</td>
<td>0.3417</td>
<td>0.0943</td>
<td>4.6728</td>
<td>1</td>
<td>0.0306</td>
</tr>
<tr>
<td>ej_psa_b2m</td>
<td>0.5210</td>
<td>0.0963</td>
<td>0.6973</td>
<td>1</td>
<td>0.4037</td>
</tr>
<tr>
<td>ej_pca3_b2m</td>
<td>0.6667</td>
<td>0.0963</td>
<td>0.0336</td>
<td>1</td>
<td>0.8546</td>
</tr>
<tr>
<td>ej_pca3_psa</td>
<td>0.6050</td>
<td>0.0938</td>
<td>0.0742</td>
<td>1</td>
<td>0.7853</td>
</tr>
</tbody>
</table>

**Figure A3.4** – ROC comparison figures for each analysis, comprising PEUW (paired cohort) and Ejaculate (paired cohort) for clinically significant CaP. The *p*-values in the table ("Pr>chi2") are for the difference between each biomarker area vs the reference (serum PSA). Std. Err. = standard error, chi2 = \( \chi^2 \), df = degrees of freedom.
Appendix 4 – Publication links


Chapter 4.2 - http://espace.library.uq.edu.au/view/UQ:354600

Chapter 4.3 - http://www.sciencedirect.com/science/article/pii/S2287888217300387
