AD_____

Award Number: W81XWH-10-1-0323

TITLE: Metabolomics and Prostate Cancer

PRINCIPAL INVESTIGATOR: Fredrik Wiklund

CONTRACTING ORGANIZATION:
Karolinska Institutet, SE-171 77 Stockholm, Sweden

REPORT DATE: May 2013

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
                         Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
May 2013	Final	01-May-2010 to 30-April-2013

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Metabolomics and Prostate Cancer	W81XWH-10-1-0323
	5b. GRANT NUMBER
	W81XWH-10-1-0323
	5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)	5d. PROJECT NUMBER
Fredrik Wiklund	
	5e. TASK NUMBER
E-Mail: Fredrik.wiklund@ki.se	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
Karolinska Institutet SE-171 77 Stockholm Sweden	

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012	
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

We performed untargeted serum metabolomic profiling of two large population-based prostate cancer populations. Molecular features have been derived and explored for association with prostate cancer status (6,138 features, 475 subjects) and prostate cancer survival (5,209 features, 534 subjects). Assessment of metabolite features revealed two features as study-wide significantly associated with prostate cancer status; however, due to low observed abundance we were not able to identify the molecular identity of these features. Among features indicated in pairwise analysis molecular identification revealed Caprolactam, L-Phosphatidic acid, and Peptide (Tyr-Lys-Thr) as possibly associated with prostate cancer aetiology. Genome-wide assessment of the four top associated metabolite features implicated four genes *PDE7B*, *NRG3*, *ILI3RA1* and *UGT3A1* of which the association between metabolite feature 174.1_53 and gene *ILI3RA1* was genome-wide significant (P = 1.4×10^{-8}). This finding is interesting since although *IL13RA1* itself has not been associated with prostate cancer, the alpha 2 chain of the same receptor (*IL13RA2*) has been reported to be differentially expressed in a metastatic prostate cancer cell line, and suggested as a target for prostate cancer treatment. In summary, several metabolite features associated with prostate cancer were discovered that warrant further investigation to advance our understanding of the underlying biological processes.

15. SUBJECT TERMS
Prostate cancer, metabolomic profiling, biomarkers

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON USAMRMC
a. REPORT	b. ABSTRACT	c. THIS PAGE			
U	U	U	UU		19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

INTRODUCTION

We hypothesize that the profiling of the human serum metabolome can unveil underlying biological processes that are associated with the initiation, aggressiveness, and prognosis of prostate cancer.

BODY

Specific Aim 1: To compare pre-treatment metabolite levels between population controls and prostate cancer cases

Timetable of research accomplishments of Specific Aim 1 as outlined in the Statement of Work:

Task 1   *Perform metabolic profiling in pre-treatment serum samples from controls, localized cases and aggressive cases:*

1.a   Deliver serum samples from population controls, indolent cases and aggressive cases from Sweden to Colorado State University. (Months 1-3).

1.b   Sample preparation at Colorado State University. (Months 1-6).

1.c   Metabolic profiling of serum samples at Colorado State University. (Months 7-12).

1.d   Statistical analysis of metabolite levels. (Months 13-15).

1.e   Manuscript preparation/submission. (Months 16-18).

Progress report
All serum samples have been delivered from Sweden to Colorado State University (1.a) where they have been prepared for metabolomic profiling (1.b). Metabolomic profiling has been completed for all samples (1.c) and statistical analysis of generated data has been completed (1.d). Manuscript has been prepared and submitted for publication (1.e).

Metabolite profiling
The raw metabolomics profiling data, which has three dimensions, mass-to-charge ratio, retention time, and signal intensity, was analyzed to identify peaks and assess magnitudes using XCMS (version 1.23.7)[1] in R version 2.12.1[2] (R Development Core Team, 2008). Identification of peaks in each chromatogram was performed by the "matchedFilter" method in XCMS with default parameters except setting full width at half maximum to 8 seconds, the signal to noise ratio threshold to 3, and allowing 100 peaks at maximum for each extracted ion chromatogram.

Table 1. Descriptive statistics by prostate cancer status[1].

	Controls (N = 188)	Cases (less aggressive) (N = 188)	Cases (more aggressive) (N = 99)
Tumor stage (T)			
1-2		100% (0)	36% (36)
3-4		0% (0)	64% (63)
Nodal stage (N)			
N0		6% (12)	15% (15)
N1		0% (0)	5% (5)
NX[2]		94% (176)	80% (79)
Distal stage (M)			
M0		24% (45)	46% (46)
M1		0% (0)	5% (5)
MX		76% (143)	48% (48)
Gleason score			
2-6		100% (188)	28% (28)
7		0% (0)	31% (31)
8-10		0% (0)	26% (26)
NA		0% (0)	14% (14)
PSA[3] (ng/ml)	0.9 (0.6-1.4)	6.6 (4.7-8.2)	19.6 (10.4-38.1)
Age[4] (years)	63.7 (60.1-70.7)	65.8 (61.4-70.5)	73.7 (66.5-77.1)
Body mass index (kg/m^2)	26.3 (24.2-27.8)	25.7 (24.1-28.1)	26.0 (23.9-28.7)
Sample storage time (days)			
2161-2421	39% (74)	38% (72)	26% (25)
2448-2716	46% (86)	29% (55)	28% (27)
2721-2990	15% (28)	21% (39)	24% (24)
3016-3276	0% (0)	12% (22)	22% (22)

[1] Continuous variables are reported as median (interquartile range), numbers in brackets are frequencies.
[2] NX and MX = not assessed.
[3] PSA = prostate specific antigen
[4] Age represents age at inclusion (controls) or age at diagnosis (cases)

The peaks that are likely to represent the same molecules were grouped across samples with an 8 second band width and 1% threshold in order to neglect the group in which the peak was identified from less than 1% of the samples. The retention time within a peak group was adjusted by the method "loess" with "gaussian" fitting. The time-wise corrected peaks were re-grouped with the same parameters as above in XCMS. Any samples for which the peaks were missing were filled as if a peak existed at the same retention time. The magnitude of a peak was calculated by integrating intensities. Output from this software is in the form of an aligned data matrix consisting of a large number of features (each feature represents one mass at a given retention time) suitable for further processing.

Population characteristics

Pertinent characteristics of the study population are displayed in Table 1. An increasing trend of age was observed across disease status with lowest age among controls and highest age among more aggressive cases. Body mass index was equally distributed across groups while PSA levels were strongly correlated with disease status. A trend of longer sample storage time was observed among cases as compared to control subjects. According to the study design, Gleason score and TNM stage were strongly shifted against more severe disease among the more aggressive cases compared to the less aggressive cases.

Association between individual profiles and disease

A total of 6,138 unique molecular features from metabolomics profiling were retained for testing for association with prostate cancer status. Association between each normalized feature and prostate cancer status was assessed through linear regression models with each feature's abundance as the outcome and disease status as categorical predictor variable (with levels: control subject, less aggressive disease, more aggressive disease). To adjust for potential confounding factors, all analyses were further adjusted for age at inclusion/diagnosis and sample storage time, represented by a categorical variable dividing storage time into four equally spaced time periods. A quantile-quantile plot of observed versus expected $-\log 10$ p-values with associated 95% confidence intervals is given in Figure 1, indicating a slight excess of significant tests.
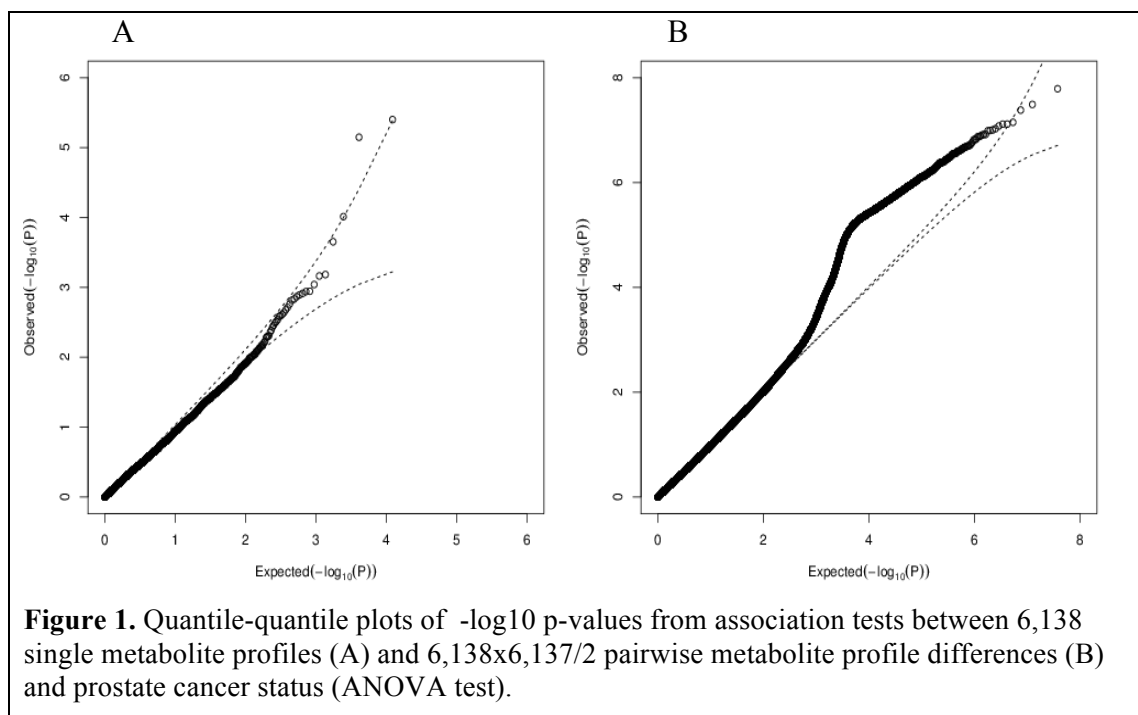
Figure 1. Quantile-quantile plots of $-\log 10$ p-values from association tests between 6,138 single metabolite profiles (A) and 6,138x6,137/2 pairwise metabolite profile differences (B) and prostate cancer status (ANOVA test).

In Table 2, details of the top four significant associations ($P < 1.0 \times 10^{-3}$) are given. Applying a Bonferroni correction (significance threshold $=0.05/6138 = 8.1 \times 10^{-6}$), two features remained study-wide significant (595.4_153, $P=4.0 \times 10^{-6}$; 422.2_315, $p=7.1 \times 10^{-6}$).

Table 2. Metabolite features associated with prostate cancer at $P < 1.0$ x 10^{-3}.

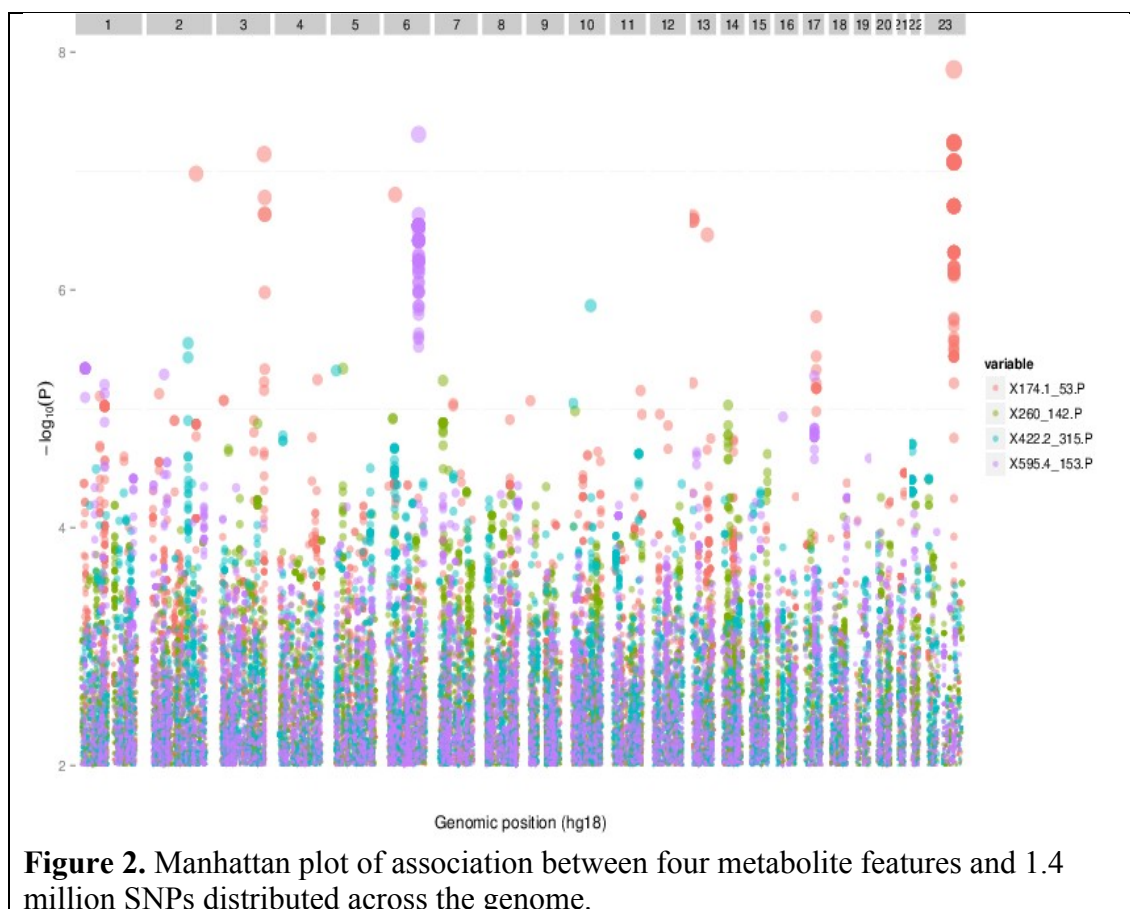| Molecular feature (m/z retention time)[1]	Prostate cancer association P-value[2]	Metabolite identification	Identification confidence
595.4_153	4.0 x 10^{-6}	Unknown	4
422.2_315	7.1 x 10^{-6}	Unknown	4
174.1_53	6.6 x 10^{-4}	Unknown	4
260_142	9.1 x 10^{-4}	Unknown	4

[1] m/z = mass to charge ratio.
[2] P-values from ANOVA test (2 df), adjusted for age and sample storage time.

Metabolite genome-wide association analysis

To further explore identified metabolite features metabolite genome-wide association analysis was performed. Identification of association between metabolites and prostate cancer related genetic variation would further implicate importance of the identified metabolite feature. In addition, associated enzymatic sequence coding may be helpful in feature identification. The four metabolites most strongly associated to prostate cancer (Table 2) were explored for quantative trait association with 1.4 million single nucleotide polymorphisms (SNPs) distributed across the genome. In Figure 2 a Manhattan plot of all association results is displayed. The position of the SNP with the lowest P-value for each feature in Table 2 is reported in Table 3, along with the marker's location in relation to the nearest annotated gene.

For each genome-wide set of metabolite-SNP tests, the Bonferroni- corrected study significance threshold is $0.05/1442840 = 3.5$ x 10^{-8}. For one of the four metabolites, study-wide significance was observed; abundance of metabolite feature 174.1_53 was associated with the SNP rs2247035 at a significance level of 1.4 x 10^{-8}. This SNP is located in an intron of the gene interleukin 13 receptor, alpha 1 (*IL13RA1*) on chromosome Xq24. Although *IL13RA1* itself has not to our knowledge been associated with PC before, the alpha 2 chain of the same receptor (*IL13RA2*) has been reported to be differentially expressed in a metastatic prostate cancer cell line, and suggested as a target for prostate cancer treatment[3].

The second strongest association ($P = 4.9$ x 10^{-8}) was observed between the metabolite feature 595.4_153 and variation in the gene phosphodiesterase 7B (*PDE7B)* on chromosome 6q23, whose protein product hydrolyzes the second messenger cAMP, a key regulator of many important physiological processes.

Figure 2. Manhattan plot of association between four metabolite features and 1.4 million SNPs distributed across the genome.

Finally, metabolite features 422.2_135 and 260_142 showed strongest association with genetic variation in genes neuregulin 3 (*NRG3*, P = 1.4 x 10^{-6}, chromosome10q23) and UDP glycosyltransferase 3 family, polypeptide A1 (*UGT3A1*, P = 4.6 x 10^{-6}, chromosome 5p13), respectively. *NRG3*, encoding a direct ligand for the ERBB4 tyrosine kinase receptor, act as a growth factor and have been suggested in the aetiology of several cancers, including prostate and breast[4]. *UGT3A1* acts on steroids, particularly estrogen analogs[5], and hypermethylation of this gene in breast cancer tissue has been associated with tumor relapse and worse survival[6].

Table 3. Metabolite genome-wide association results.

Molecular feature (m/z retention time)[1]	Metabolite GWAS lowest P value	Genomic position (Chr:bp, hg18)	Nearest gene (SNP location)
595.4_153	4.9 x 10-8	Chr6:136374051	*PDE7B* (intron)
422.2_135	1.4 x 10-6	Chr10:83842772	*NRG3* (intron)
174.1_53	1.4 x 10-8	ChrX:117756115	*ILI3RA1* (intron)
260_142	4.6 x 10-6	Chr5:35999264	*UGT3A1* (intron)

[1] m/z = mass to charge ratio.

Association between pairs of profiles and disease

Next we explored pairwise logarithmically transformed metabolite differences (corresponding to ratios on the original scale) for association with prostate cancer. Metabolite ratios have been suggested to show more robust associations than single metabolic features, since they may correlate with enzyme function or flow through metabolic pathways. Top results from the analysis of 6138*6137/2 pairwise differences are presented in Figure 1 and Table 4. No metabolite pair was study-wide significantly associated to PC after Bonferroni correction for the number of tests performed (significance threshold = 2.7×10^{-9}). Seven metabolite feature pairs were associated at a significance threshold of 1.0E-7. Five of these pairs involved the metabolic feature 595.4_153, which was the most strongly associated feature in univariate analyses (Table 2). Further metabolite features that were implicated in these pairwise assessments were 114.1_118, 411.3_285, 443.3_275, 451.2_266 and 597.4_306. Each of the two remaining pairs significant at the 1.0E-7 threshold included feature 422.2_315, the second most strongly univariate associated metabolite, in combination with features 226.2_212 and 581.3_446.

Table 4. Pairwise metabolite features associated with prostate cancer.

Molecular feature pair (m/z retention time)[1]	Association P value	Metabolite identification	Identification confidence
595.4_153 – 114.1_118	3.2×10^{-8}	Unknown – Caprolactam	4 – 1
595.4_153 – 443.3_275	4.2×10^{-8}	Unknown – Unknown	4 – 4
597.4_306 - 595.4_153	7.1×10^{-8}	L-Phosphatidic acid – Unknown	2 – 4
595.4_153 - 451.2_266	7.7×10^{-8}	Unknown – Unknown	4 – 4
595.4_153 - 411.3_285	8.4×10^{-8}	Unknown – Peptide (Tyr-Lys-Thr)	4 – 3
581.3_440 - 422.2_315	9.4×10^{-8}	Unknown – Unknown	4 – 4
422.2_315 - 226.2_212	1.0×10^{-7}	Unknown – Unknown	4 - 4

[1] m/z = mass to charge ratio.

Feature identification

Identification of the top ranked features from univariate and pairwise analysis was performed according to the following workflow: 1) Accurate mass measurements are searched against a variety of metabolite databases including the Human Metabolome Database, Metlin, and LipidMaps. 2) A combination of the accurate mass measurement and the isotopic distribution of the mass spectrometry peaks are imported into the elemental composition calculator (Waters MassLynx software) to generate a "best fit" molecular formula. 3) The best-fit molecular formula is used to filter the database search results to yield a putative identification. 4) When possible, fragmentation information for the metabolite feature are extracted from the mass spectrometry analysis and compared with fragmentation of the putative metabolite found in the literature, and/or mass spectral database, and/or from a commercially available pure standard. We report metabolite identification confidence based on metabolomics standards initiative recommendations[7]. Specifically, level 1 refers to confident molecular identification based on orthogonal analytical parameters (accurate mass, retention time, and MS/MS fragmentation) relative to an authentic compound. Level 2 refers to a putative identification based on

physicochemical properties and/or spectral similarity with literature or spectral libraries. Level 3 refers to the putative identification of a compound class based on physicochemical properties or spectral similarity. Level 4 refers to an unknown compound

For the four most strongly associated features in univariate analysis (Table 2) we were unsuccessful in providing the molecular identity of any of the feature. From the pairwise analysis we were able to identify three of the seven additional features implicated including caprolactam, L-Phosphatidic acid and the peptide Tyr-Lys-Thr. Each of these molecules was implicated in combination with 595.4_153, the most strongly associated metabolite feature. We were unable to retrieve molecular identities for any of the two features implicated in combination with the 422.2_315 feature (Table 4). Of note, caprolactam is a non-endogenous compound used in the manufacturing of nylon and produced around the world in very large quantities. Phosphatidic acids are fatty acid derivatives of glycerophosphates, and are established intracellular signaling lipids.

Specific Aim 2: To compare post-treatment metabolic levels between prostate cancer patients with lethal and non-lethal disease outcome

Timetable of research accomplishments of Specific Aim 2 as outlined in the Statement of Work:

Task 2 *Perform metabolic profiling in post-treatment serum samples from cases with lethal and non-lethal disease outcome:*

2.a Deliver 608 serum samples from Sweden to Colorado State University. (Months 19-21).

2.b Sample preparation at Colorado State University. (Months 19-24).

2.c Metabolic profiling of serum samples at Colorado State University. (Months 25-30).

2.d Statistical analysis of metabolite levels. (Months 31-33).

2.e Manuscript preparation/submission. (Months 34-36).

Progress report

All serum samples have been delivered from Sweden to Colorado State University (2.a) where they have been prepared for metabolomic profiling (2.b). Profiling of metabolite levels has been completed (2.c). All statistical analysis of metabolite features have been completed (2.d). A manuscript is under preparation (2.e).

Metabolite profiling

The raw metabolomics profiling data was processed as described under specific Aim 1 using XCMS (version 1.23.7)[1] in R version 2.12.1 (R Development Core Team, 2008).

Population characteristics

Pertinent characteristics of the study population are displayed in Table 5. Body mass index was equally distributed between non-lethal and lethal patients. According to the matched study design, prognostic risk group and primary treatment were equally distributed between groups. Majority of patients were in the highest prognostic risk group (metastatic disease) and the most common treatment option was GnRH in combination with antiandrogene.

Table 5. Clinical characteristics of prostate cancer patients.

Characteristic	Alive (N = 267)	Deceased (N = 267)
Follow-up (tears), mean (range)	6.4 (5.2-8.2)	2.8 (0.1-7.1)
Age at diagnosis (years), mean (SD)	69.2 (7.0)	69.1 (7.0)
BMI (kg/m^2), mean (SD)	26.2 (3.2)	26.4 (3.5)
Prognostic risk group, no (%)		
Intermediate	4 (1.5)	4 (1.5)
High	19 (7.1)	19 (7.1)
Metastatic	244 (91.4)	244 (91.4)
Primary treatment		
Hormones	1 (0.4)	1 (0.4)
Surgical castration	28 (10.5)	28 (10.5)
Antiandrogene	46 (17.2)	46 (17.2)
GnRH	74 (27.7)	74 (27.7)
GnRH and antiandrogene	118 (44.2)	118 (44.2)

<u>Association between individual profiles and prostate cancer survival</u>

A total of 5,209 unique molecular features from metabolomics profiling were retained for testing for association with prostate cancer survival. Association between each normalized feature and disease survival was assessed through stratified Cox regression proportional hazard models. All patients were followed from date of diagnosis until date of death from prostate cancer or censoring (at death from other causes other than prostate cancer or at end of follow-up).

Table 6. Metabolite features associated with prostate cancer survival at $P < 1.0 \times 10^{-3}$.

Molecular feature (m/z retention time)[1]	Hazard ratio (95% CI)	P-value[2]
148.5_415	0.92 (0.88-0.96)	1.3×10^{-4}
272.7_415	0.91 (0.86-0.96)	3.9×10^{-4}
244.7_415	0.35 (0.19-0.63)	4.1×10^{-4}
508.3_309	1.71 (1.26-2.32)	6.0×10^{-4}
743.6_671	0.93 (0.89-0.97)	7.7×10^{-4}
639.4_383	4.07 (1.80-9.24)	7.7×10^{-4}
742.6_671	0.98 (0.96-0.99)	9.1×10^{-4}
631.6_740	1.33 (1.12-1.58)	9.7×10^{-4}

[1] m/z = mass to charge ratio.
[2] P-values from stratified Cox regression analysis.

In Table 6, the top eight significant associations ($P < 1.0 \times 10^{-3}$) are given. Applying a Bonferroni correction (significance threshold $=0.05/5209 = 9.6 \times 10^{-6}$), no feature was study-wide significant. The lack of association between metabolite features and prostate cancer survival was apparent from the quantile-quantile plot of observed versus expected –log10 p-values with associated 95% confidence intervals (Figure 3), indicating no excess of significant associations.
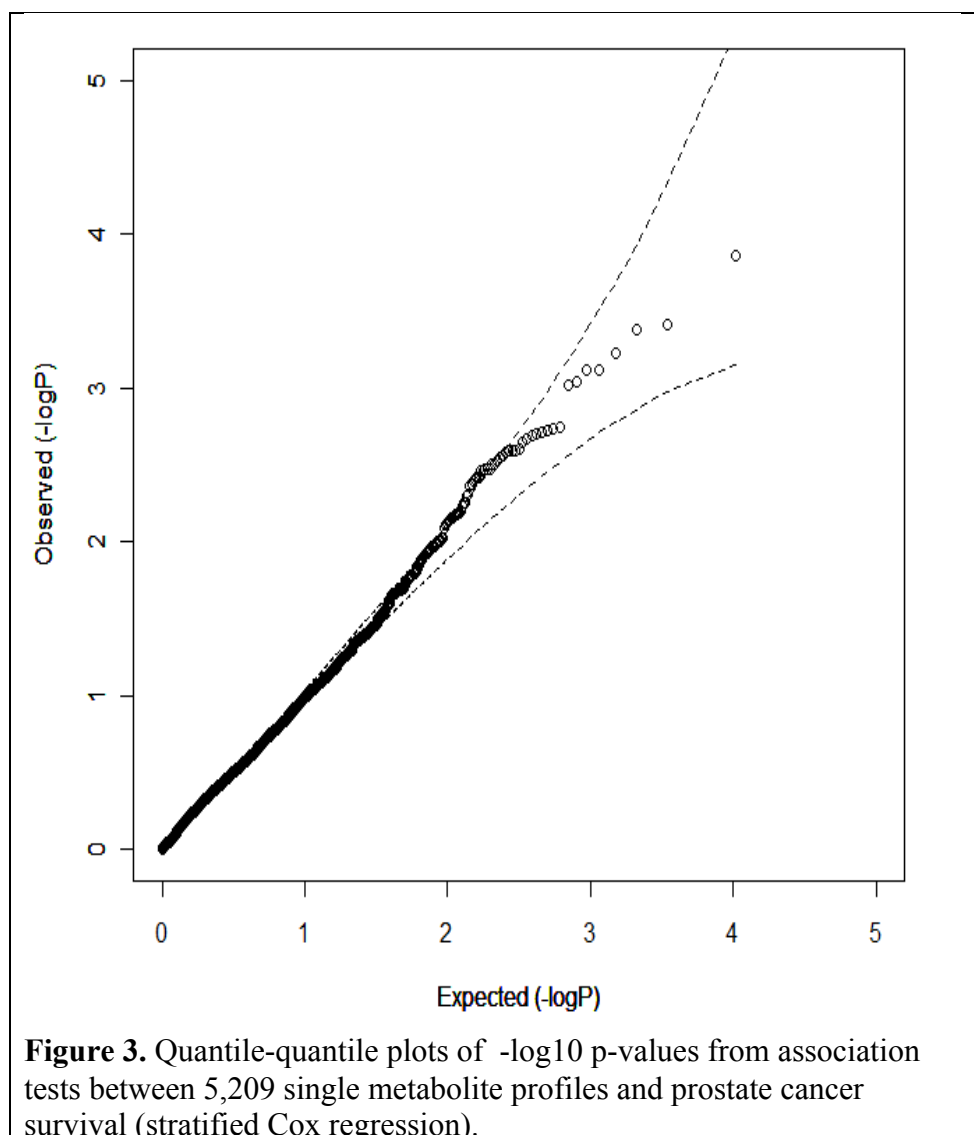
Figure 3. Quantile-quantile plots of -log10 p-values from association tests between 5,209 single metabolite profiles and prostate cancer survival (stratified Cox regression).

Table 7. Pairwise metabolite features associated with prostate cancer survival at P $< 5.0 \times 10^{-6}$.

Molecular feature pair (m/z retention time)[1]	Hazard ratio (95% CI)	P-value[2]
413.2_411 – 394.3_504	1.98 (1.70-2.27)	2.2×10^{-6}
508.3_309 – 394.3_504	1.70 (1.48-1.92)	2.3×10^{-6}
372.1_55 – 332.3_449	1.31 (1.20-1.42)	2.5×10^{-6}
344.3_211 – 229_137	1.84 (1.58-2.09)	2.9×10^{-6}
350.1_56 – 521.2_632	1.20 (1.12-1.27)	2.9×10^{-6}
350.1_56 – 332.3_449	1.20 (1.13-1.28)	2.9×10^{-6}
513.4_349 – 394.3_504	2.09 (1.78-2.40)	3.0×10^{-6}
350.1_56 – 538.2_632	1.19 (1.12-1.27)	3.2×10^{-6}
350.1_56 – 610.2_681	1.20 (1.12-1.27)	3.4×10^{-6}
508.3_309 – 319.2_201	1.75 (1.51-1.99)	3.5×10^{-6}
513.4_349 – 319.2_201	2.33 (1.97-2.69)	3.6×10^{-6}
344.3_211 – 417.8_707	1.80 (1.56-2.05)	3.6×10^{-6}
413.2_411 – 229_137	1.90 (1.63-2.18)	3.8×10^{-6}
743.6_671 – 508.3_309	0.61 (0.40-0.82)	3.9×10^{-6}
508.3_309 – 417.8_707	1.62 (1.42-1.83)	4.0×10^{-6}
508.3_309 – 491.4_471	1.50 (1.32-1.67)	4.0×10^{-6}
413.2_411 – 503.1_632	1.67 (1.45-1.88)	4.1×10^{-6}
344.3_211 – 129.1_153	1.87 (1.60-2.13)	4.1×10^{-6}
404.8_679 – 383.2_240	0.63 (0.44-0.83)	4.2×10^{-6}
350.1_56 – 543.1_631	1.21 (1.13-1.28)	4.3×10^{-6}
508.3_309 – 404.8_679	1.40 (1.26-1.55)	4.4×10^{-6}
394.3_504 – 344.3_211	0.56 (0.31-0.81)	4.7×10^{-6}
742.6_671 – 508.3_309	0.62 (0.42-0.83)	4.8×10^{-6}
485.3_309 – 394.3_504	1.70 (1.47-1.93)	4.9×10^{-6}
742.6_671 – 585.4_510	0.53 (0.25-0.80)	4.9×10^{-6}

[1] m/z = mass to charge ratio.
[2] P-values from stratified Cox regression analysis.

Metabolite genome-wide association analysis

As in specific Aim 1 we performed metabolite genome-wide association analysis for the metabolites reported in Table 6. This effort revealed no study-wide significant associations between any genetic variants for any of the eight metabolite features (data not-shown).

Association between pairs of profiles and prostate cancer survival

Next we explored pairwise logarithmically transformed metabolite differences (corresponding to ratios on the original scale) for association with prostate cancer survival. Top results from the analysis of 5209*5208/2 pairwise differences are presented in Table 7. No metabolite pair was study-wide significantly associated to PC after Bonferroni correction for the number of tests performed (significance threshold = 2.7×10^{-9}).

Of the eight top significant metabolite features identified in univariate analysis (Table 6) three were implicated in the pairwise analysis. Feature 508.3_309 was observed associated with prostate cancer survival in combination with features 394.3_504, 319.2_201, 417.8_707,

491.4_471 and 404.8_679. In addition, feature 508.3_309 was also implicated in combination with features 743.6_671 and 742.6_671, both of which were observed in the univariate analysis.

Feature identification
Identification of the top ranked features from univariate and pairwise analysis is ongoing according to same protocol as described under specific aim 1.

KEY RESEARCH ACCOMPLISHMENTS

- A total of 6,132 metabolite features have been derived for the first population of population controls, localized prostate cancer cases, and aggressive prostate cancer cases.
- A total of 5,209 metabolite features have been derived for the second population contrasting lethal and non-lethal outcome of prostate cancer.
- Statistical assessment of association between metabolite features and prostate cancer status and prostate cancer survival has been performed.
- Molecules Caprolactam, L-Phosphatidic acid, and Peptide (Tyr-Lys-Thr) have been identified as weakly associated with prostate cancer status.
- Metabolite genome-wide association has been performed for prostate cancer related metabolite features.
- Four genes have been observed as associated with prostate cancer related metabolite features: *PDE7B*, *NRG3*, *ILI3RA1*, and *UGT3A1*.
- Several metabolite features weakly associated with prostate cancer survival have been identified and molecular identification of these is ongoing.

REPORTABLE OUTCOMES

Manuscript

R. Karlsson, M. Hong, J. Prenni, C. Broeckling, H Grönberg, J. Prince, F. Wiklund. Untargeted serum metabolomic profiling of prostate cancer. Submitted.

Abstract

R. Szulkin, R. Karlsson, A. Heuberger, M. Hong, C. Broeckling, J. Prenni, J. Prince, F. Wiklund. Serum metabolomics and prostate cancer survival. Abstract #1189T. Presented at the 62nd Annual Meeting of The American Society of Human Genetics, November 7, 2012 in San Francisco, California, US.

CONCLUSION

In this project we have successfully performed untargeted serum metabolomic profiling of two large population-based prostate cancer populations. Molecular features have been derived and explored for association with prostate cancer status (6,138 features, 475 subjects) and prostate cancer survival (5,209 features, 534 subjects). Assessment of metabolite features revealed two features as study-wide significantly associated with prostate cancer status; however, we were not able to identify the molecular identity of these features, probably due to their low observed abundance. In pairwise metabolite feature assessment only weak association (not study-wide significantly associated) with prostate cancer status was observed. Among features indicated in the pairwise analysis molecular identification revealed Caprolactam, L-Phosphatidic acid, and Peptide (Tyr-Lys-Thr) as possibly associated with prostate cancer aetiology. Finally we

performed genome-wide assessment of the four top associated metabolite features. Four genes were implicated from this effort including *PDE7B*, *NRG3*, *ILI3RA1* and *UGT3A1* of which the association between metabolite feature 174.1_53 and gene *ILI3RA1* was genome-wide significant (P = 1.4 x 10^{-8}). This finding is interesting since although *IL13RA1* itself has not been associated with prostate cancer, the alpha 2 chain of the same receptor (*IL13RA2*) has been reported to be differentially expressed in a metastatic prostate cancer cell line, and suggested as a target for prostate cancer treatment[3]. Compared to the findings related to the first study population (prostate cancer status) no study-wide significant associations between any metabolite feature and prostate cancer survival was observed in the second study population. Overall these results are negative regarding our effort to identify novel biomarker of clinical use for early prostate cancer detection and treatment monitoring. It remains to be shown if our results will improve our understanding of the underlying biological processes that are associated with initiation and prognosis of prostate cancer.

We have performed metabolite profiling using blood samples collected throughout Sweden. Participating subjects were asked to visit nearest health clinic to donate a blood sample. Drawn blood was sent by mail overnight to the biobank at Umeå University for preparation and storage (-80 °C). It is possible that molecules relevant for prostate cancer initiation and progression may have degraded during this process and thereby been impossible to detect in our study. Therefore we may have missed molecules involved in important biological processes related to prostate cancer aetiology due to sample handling. However, regarding identifying new clinically relevant biomarkers we argue that our design was appropriate since quickly degrading molecules will be of limited clinical use.

Application of metabolomics to identify novel disease biomarkers has attracted increasing interest[8]. Early-stage diagnosis of incident cancer may considerably improve clinical outcome through early treatment. Metabolomic profiling has been reported for numerous types of malignancies including colorectal cancer[9], lung cancer[10], primary liver cancer[11], ovarian cancer[12], and breast cancer[13]. The most common used biomarker for prostate cancer detection to date is the prostate specific antigen (PSA). Although PSA has adequate sensitivity the lack of specificity would results in considerable overdiagnosis and overtreatment in a population-based screening program[14]. In 2009, applying metabolomic screening on both plasma and urine samples from prostate cancer patients and controls, Sreekumar and coworkers reported a potential role of sarcosine in prostate cancer prognosis[15]. However, their finding has been difficult to replicate in independent populations[16,17]. Miyagi and coworkers[18] recently reported that plasma free amino acids show great potential to discriminate between healthy controls and prostate cancer patients. Estrogen and androgen metabolites has been proposed as potential biomarkers for prostate cancer[19,20], while Thysell and coworkers reported high levels of cholesterol in prostate cancer bone metastases in a metabolomic study of prostate cancer tissue and plasma samples[21].

In conclusion, our project has failed in identifying novel prostate cancer biomarkers of clinical use. Future work may benefit from stricter sample handling that would increase number of molecules possible to study. Although no clinically relevant biomarkers were identified we did observe several metabolite features that associated with prostate cancer status. We were also able to locate several genes associated with the abundance of these metabolites. These results are novel and may advance our understanding of the biological processes related to the aetiology of prostate cancer and our research group intends to continue explore the these findings in continued research.

REFERENCES

1.	Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Analytical chemistry.* Feb 1 2006;78(3):779-787.

2.	*R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing; 2008.

3.	He H, Xu J, Nelson PS, Marshall FF, Chung LW, Zhau HE, He D, Wang R. Differential expression of the alpha2 chain of the interleukin-13 receptor in metastatic human prostate cancer ARCaPM cells. *The Prostate.* Jun 15 2010;70(9):993-1001.

4.	Montero JC, Rodriguez-Barrueco R, Ocana A, Diaz-Rodriguez E, Esparis-Ogando A, Pandiella A. Neuregulins and cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research.* Jun 1 2008;14(11):3237-3241.

5.	Meech R, Mackenzie PI. UGT3A: novel UDP-glycosyltransferases of the UGT superfamily. *Drug metabolism reviews.* Feb 2010;42(1):45-54.

6.	Hill VK, Ricketts C, Bieche I, Vacher S, Gentle D, Lewis C, Maher ER, Latif F. Genome-wide DNA methylation profiling of CpG islands in breast cancer identifies novel genes associated with tumorigenicity. *Cancer research.* Apr 15 2011;71(8):2988-2999.

7.	Sansone SA, Fan T, Goodacre R, Griffin JL, Hardy NW, Kaddurah-Daouk R, Kristal BS, Lindon J, Mendes P, Morrison N, Nikolau B, Robertson D, Sumner LW, Taylor C, van der Werf M, van Ommen B, Fiehn O. The metabolomics standards initiative. *Nature biotechnology.* Aug 2007;25(8):846-848.

8.	Yin P, Xu G. Metabolomics for tumor marker discovery and identification based on chromatography-mass spectrometry. *Expert review of molecular diagnostics.* May 2013;13(4):339-348.

9.	Leichtle AB, Nuoffer JM, Ceglarek U, Kase J, Conrad T, Witzigmann H, Thiery J, Fiedler GM. Serum amino acid profiles and their alterations in colorectal cancer. *Metabolomics : Official journal of the Metabolomic Society.* Aug 2012;8(4):643-653.

10.	Maeda J, Higashiyama M, Imaizumi A, Nakayama T, Yamamoto H, Daimon T, Yamakado M, Imamura F, Kodama K. Possibility of multivariate function composed of plasma amino acid profiles as a novel screening index for non-small cell lung cancer: a case control study. *BMC cancer.* 2010;10:690.

11.	Tan Y, Yin P, Tang L, Xing W, Huang Q, Cao D, Zhao X, Wang W, Lu X, Xu Z, Wang H, Xu G. Metabolomics study of stepwise hepatocarcinogenesis from the model rats to patients: potential biomarkers effective for small hepatocellular carcinoma diagnosis. *Molecular & cellular proteomics : MCP.* Feb 2012;11(2):M111 010694.

12.   Chen J, Zhang X, Cao R, Lu X, Zhao S, Fekete A, Huang Q, Schmitt-Kopplin P, Wang Y, Xu Z, Wan X, Wu X, Zhao N, Xu C, Xu G. Serum 27-nor-5beta-cholestane-3,7,12,24,25 pentol glucuronide discovered by metabolomics as potential diagnostic biomarker for epithelium ovarian cancer. *Journal of proteome research.* May 6 2011;10(5):2625-2632.

13.   Chen Y, Zhang R, Song Y, He J, Sun J, Bai J, An Z, Dong L, Zhan Q, Abliz Z. RRLC-MS/MS-based metabonomics combined with in-depth analysis of metabolic correlation network: finding potential biomarkers for breast cancer. *The Analyst.* Oct 2009;134(10):2003-2011.

14.   Ilic D, Neuberger MM, Djulbegovic M, Dahm P. Screening for prostate cancer. *Cochrane Database Syst Rev.* 2013;1:CD004720.

15.   Sreekumar A, Poisson LM, Rajendiran TM, Khan AP, Cao Q, Yu J, Laxman B, Mehra R, Lonigro RJ, Li Y, Nyati MK, Ahsan A, Kalyana-Sundaram S, Han B, Cao X, Byun J, Omenn GS, Ghosh D, Pennathur S, Alexander DC, Berger A, Shuster JR, Wei JT, Varambally S, Beecher C, Chinnaiyan AM. Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. *Nature.* Feb 12 2009;457(7231):910-914.

16.   Jentzmik F, Stephan C, Lein M, Miller K, Kamlage B, Bethan B, Kristiansen G, Jung K. Sarcosine in prostate cancer tissue is not a differential metabolite for prostate cancer aggressiveness and biochemical progression. *The Journal of urology.* Feb 2011;185(2):706-711.

17.   Jentzmik F, Stephan C, Miller K, Schrader M, Erbersdobler A, Kristiansen G, Lein M, Jung K. Sarcosine in urine after digital rectal examination fails as a marker in prostate cancer detection and identification of aggressive tumours. *European urology.* Jul 2010;58(1):12-18; discussion 20-11.

18.   Miyagi Y, Higashiyama M, Gochi A, Akaike M, Ishikawa T, Miura T, Saruki N, Bando E, Kimura H, Imamura F, Moriyama M, Ikeda I, Chiba A, Oshita F, Imaizumi A, Yamamoto H, Miyano H, Horimoto K, Tochikubo O, Mitsushima T, Yamakado M, Okamoto N. Plasma free amino acid profiling of five types of cancer patients and its application for early detection. *PLoS One.* 2011;6(9):e24143.

19.   Kosti O, Xu X, Veenstra TD, Hsing AW, Chu LW, Goldman L, Bebu I, Collins S, Dritschilo A, Lynch JH, Goldman R. Urinary estrogen metabolites and prostate cancer risk: a pilot study. *The Prostate.* Apr 2011;71(5):507-516.

20.   Yang L, Gaikwad NW, Meza J, Cavalieri EL, Muti P, Trock B, Rogan EG. Novel biomarkers for risk of prostate cancer: results from a case-control study. *The Prostate.* Jan 1 2009;69(1):41-48.

21.   Thysell E, Surowiec I, Hornberg E, Crnalic S, Widmark A, Johansson AI, Stattin P, Bergh A, Moritz T, Antti H, Wikstrom P. Metabolomic characterization of human prostate cancer bone metastases reveals increased levels of cholesterol. *PLoS One.* 2010;5(12):e14175.

Untargeted serum metabolomic profiling of prostate cancer

Robert Karlsson[1,*], Mun-Gwan Hong[1], Jessica Prenni[2,3], Corey Broeckling[2], Henrik Grönberg[1], Jonathan A. Prince[1], Fredrik Wiklund[1]

1) Department of Medical Epidemiology and Biostatistics, Karolinska institutet, 171 77 Stockholm, Sweden

2) Proteomics and Metabolomics Facility, Colorado State University, Fort Collins, Colorado, USA

3) Department of Biochemistry and Molecular Biology, Colorado State University, Fort Collins, Colorado, USA

* Corresponding author: Robert Karlsson, Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Box 281, SE-171 77 Stockholm, Sweden.

Tel: +46 8 524 822 21. Fax: +46 8 31 49 75. E-mail: robert.karlsson@ki.se

Abstract

Background: Prostate cancer (PC) is a common disease affecting older men. The current clinical test for PC measures serum prostate specific antigen (PSA). Due to insufficient sensitivity and specificity, overdiagnosis and overtreatment of harmless or nonexistent tumors, as well as missed aggressive tumors, are common occurrences. New biomarkers for PC, independent of PSA, would thus be highly useful. In this study we examined the serum metabolome, the set of all small molecules, for such biomarkers using an untargeted ultra-high performance liquid chromatography-mass spectrometry (UPLC-MS) approach.

Materials and Methods: Serum samples, taken before treatment, from 287 PC cases (of which 99 had advanced disease) and 188 population controls were analyzed by UPLC-MS. Detected metabolite features and pairwise feature differences were tested for association with PC status using linear regression and the ANOVA F-test, adjusting for sample storage time and patient age. The most PC-associated features were further tested for association to single nucleotide polymorphisms (SNPs) genome-wide.

Results: 6138 metabolite features were quantified and tested for association with PC status. Two associations were statistically significant after Bonferroni correction for 6138 tests (mass/charge ratio [m/z] 595.4: P=4.0E-06; m/z 422.2: P=7.1E-06). No pairwise feature difference associations were significant after Bonferroni correction for 6138*6137/2 tests. The four strongest PC-associated features (P-values < 1E-3) all had their strongest SNP associations located in introns of annotated genes (*PDE7B*, *NRG3*, *IL13RA1*, *UGT3A1*).

Conclusion: No metabolite features useful as PC biomarkers were found in this study, and the features associated with PC status could not be assigned a molecular identity. Studies analyzing an even broader spectrum of molecules than those detectable by UPLC-MS may be more successful. The PC-metabolite-associated genes discovered may indicate processes involved in PC aetiology.

Introduction

Prostate cancer (PC) is the most common non-cutaneous malignancy and the second leading cause of cancer death among men in developed countries. It has been estimated that in the year 2007, almost 800,000 men will be diagnosed with prostate cancer worldwide and 250,000 will die of the disease (Crawford 2009).

For several years, serum prostate-specific antigen (PSA) testing and digital rectal exams (DRE) have been the standard measures for diagnosis of prostate cancer. However, since a high proportion of men with abnormal findings from PSA and DRE are not proven to have prostate cancer, unnecessary intervention is common. In addition, once prostate cancer is diagnosed, choice of treatment remains a major challenge.

The risk of overtreatment is substantial considering the excellent prognosis of a high proportion of men with untreated localized disease (Johansson et al. 2004) and the morbidity associated with curative treatment. Management by active surveillance with selective delayed intervention based on early PSA changes has been proposed as a strategy to reduce overtreatment of patients with indolent disease (Klotz 2005). However, although both baseline PSA measurements and rate of PSA change are important prognostic factors, they perform poorly in distinguishing those who will develop a lethal prostate cancer from those at low risk of disease progression (Fall et al. 2007). To this end, improved tools to distinguish lethal from indolent disease to guide clinicians in treatment decisions is critical.

Cancer development and progression is characterized by multiple, complex molecular events. To decipher the molecular networks involved in tumor initiation and neoplastic progression, gene and protein expression have been extensively profiled in human tumors; however, few efforts have been performed to explore global metabolite alterations in this context.

Metabolomics is a field of research that attempts to provide a comprehensive picture of the physiological state of an organism by providing precise measurements of a large number of small molecules (Becker et al. 2012; Xie, Waters, and Schirra 2012; Patti 2011). One of two methodologies is most frequently applied in metabolomics – either nuclear magnetic resonance spectroscopy (NMR) or mass spectrometry (MS) based techniques. Commonly, chromatography is coupled to MS, a system which offers the benefits of partial separation of a complex mixture via chromatography, an additional mass dimension of separation, and molecular weight and fragmentation information.

To date several studies have applied metabolomic profiling to identify novel biomarkers in cancer research (Spratlin, Serkova, and Gail Eckhardt 2009). Although these studies have been restricted in number of metabolites being profiled and number of samples from disease-affected and unaffected individuals being screened, several interesting biomarkers have been identified including sarcosine in prostate cancer assessment (Sreekumar et al. 2009). The utility of sarcosine as a prostate cancer biomarker has however not been definitely proven (Issaq and Veenstra 2011). While these studies demonstrate the potential of metabolomics in identification of cancer diagnostic biomarkers, expanding the coverage of the metabolome, increasing the sample size, and using clinically relevant endpoints are actions likely to improve our ability to identify novel prostate cancer biomarkers.

The aim of the present study was to explore global serum metabolite profiles in a large population-based prostate cancer study. Serum metabolite levels were contrasted between unaffected population controls, prostate cancer cases with indolent disease, and prostate cancer cases with aggressive disease.

Materials and methods

Study sample

The patients and controls for this study were selected from a biobank that was established as part of the Cancer of the Prostate in Sweden (CAPS) study of genetic and dietary risk factors for prostate cancer. Details of the sample collection procedure has been previously published (Lindmark et al. 2004). In brief, CAPS is a population-based case-control sample of Swedish men diagnosed with prostate cancer between 2001 and 2003, and population controls who were frequency matched to the expected age distribution and geographic region of cases. Cases were identified from the Swedish cancer register, and controls from the Swedish population register. Clinical characteristics of cases were obtained from the national prostate cancer register (Adolfsson et al. 2007). All study participants provided written informed consent, and the study was approved by the local institutional review board. The full study biobank constitutes blood samples, separated into serum, plasma, and DNA, from 2875 PC cases and 1746 population controls.

For the present study, we analyzed serum from188 control samples, 99 samples from patients with aggressive PC, and 188 samples from patients with less aggressive PC. Aggressive disease was defined as fulfilling one or more of the following criteria: T stage $\geq$ 3, positive lymph node status, positive metastasis status, Gleason score $\geq$ 8, or blood PSA level $\geq$ 50 ng/ml at diagnosis, while patients not fulfilling any of the criteria for aggressive disease were classified as having less aggressive disease. For all prostate cancer patients included in the present study, serum was extracted from blood samples drawn before any treatment for their disease had been initiated.

Ultra-Performance Liquid Chromatography-Mass Spectrometry profiling

Serum metabolomic profiles were acquired by ultra-performance liquid chromatography (UPLC) coupled with mass spectrometry (MS) at the Proteomics and Metabolomics Facility, Colorado State University, USA. Frozen serum samples, delivered from the Medical Biobank at Umeå University were thawed and 200 µL transferred to an eppendorf tube. Proteins were precipitated by adding 800µL of ice cold methanol, and the tube was spun at 5000g for 15 minutes to separate protein from supernatant. 400µL of the supernatant was transferred to an autosampler vial for UPLC-MS analysis. One µL injections were performed on a Waters Acquity UPLC system (Waters Corp., Milford, MA, USA). Separation was performed through a Waters Acquity UPLC C8 column (1.8 µM, 1.0 x 100 mm), using a gradient from solvent A (95% water, 5% methanol, 0.1% formic acid) to solvent B (95% methanol, 5% water, 0.1% formic acid). Injections were made in 100% A, which was held for 0.1 min. A succession of linear gradients was used, from 0% B to 40% B in 0.9 minutes, then to 70% B in 2 minutes, and finally to 100% B in 8 minutes. The mobile phase was held at 100% B for 6 minutes, returned to starting conditions over 0.1 minute, and allowed to re-equilibrate for 5.9 minutes for a total run time of 23 minutes. Flow rate was maintained at 140 µL/min for the duration of the run. The column was held at 50°C and samples were held at 5°C. Column eluate was infused into a Waters Q-Tof Micro MS fitted with an electrospray source. Data was collected in positive ion mode, scanning from 50-1000 at a rate of 2 scans per second with 0.1 second interscan delay. Calibration was performed prior to sample analysis via infusion of sodium formate solution, with mass accuracy within 5 ppm. The capillary voltage was held at 3000 V, the sample cone at 30 V, the source temperature at 130°C, and the desolvation temperature at 300°C with a nitrogen desolvation gas flow rate of 400 L/hr. The collision cell was held at collision energy of 7 eV.

Postprocessing of metabolite features

The software package XCMS (Smith et al. 2006) was used to align and extract measured ion intensities from the UPLC-MS chromatograms. Integrated peak intensities were assessed through the "matchedFilter" method in XCMS, which fits a second derivative Gaussian filter function to each peak to suppress noise. Peak detection parameters were set to 8 seconds full peak width at half maximum intensity, a minimum signal to noise ratio of 3, and a maximum of 100 peaks for each slice of the m/z domain considered.

Peak grouping across samples was performed with an 8 second bandwidth and a minimum fraction of 1% of all samples needed to display a peak. Retention time correction of the grouped peaks was performed using a loess smoothing function. After retention time correction, peak grouping was repeated as above, and signal intensities for each peak and sample were calculated by integrating the intensity curve. The same range was integrated in all samples whether a peak had been detected or not.

The sample/feature matrix was normalized by scaling the feature intensities of each analyzed sample by a factor, so that all samples were given the same mean intensity (the mean of sample means before normalization). After intensity normalization, and removal of outliers (where the variance within replicate groups significantly exceeded the variance for the whole sample), final feature intensities were set to the means across three (if no outliers were removed) replicate samples for each biological sample, transformed to the 10-logarithm of the measured intensities.

Statistical analyses

Association between each normalized LC-MS feature and PC status was assessed through linear regression, in models with each measured feature's abundance as the outcome, and PC status as a

categorical predictor variable (with levels control, less aggressive, more aggressive). The analyses were further adjusted for age at inclusion, and sample storage time categorized in four equally long time bands, which were potential confounding factors. The ANOVA F-test was used to test the overall statistical significance of the PC status factor variable as a predictor for the abundance of each LC-MS feature. Analyses were performed using R version 2.15.1 (R Development Core Team 2012).

It has been suggested that variation in pairwise ratios between metabolites can reflect variation in enzymatic activity and other biological processes (Altmaier et al. 2008). If such variation is also disease-related, analyzing metabolite ratios could provide aetiological insights not possible from single feature analyses. To assess such effects in our sample, we performed the same tests as for the single metabolites for all n(n-1)/2 pairwise differences between LC-MS features. Since data were log-transformed before analysis, this corresponds to investigating ratios on the original scale.

Genotyping and metabolite genome-wide association analysis

For the LC-MS sample at hand, genome-wide genotypes were available from previous studies. An association between a metabolite feature of unknown molecular identity and genetic variation near sequence coding for an enzyme acting on specific molecules may be helpful in identifying the feature at hand. The metabolites most strongly associated with changes in disease were therefore investigated for quantitative trait association to single nucleotide polymorphisms (SNPs) genome-wide.

Genotypes were generated on the Affymetrix (Santa Clara, CA, USA) GeneChip Human Mapping 500K and 5.0 platforms, by collaborators at the Wake Forest University, USA, following the manufacturer's recommendations. The average call rate for genotypes was 99.1%, and the concordance between replicated samples was greater than 99%. SNPs with no call for more than 5% of samples, or deviating from Hardy-Weinberg equilibrium (exact test $P < 10^{-6}$) were excluded from further analysis. After quality control, additional genotypes were imputed using IMPUTE (Marchini et al. 2007) software and the CEU panel of reference haplotypes from the international HapMap project (The International HapMap Consortium 2007). After imputation, genotypes were called from imputed posterior probabilities. Most likely genotypes with a posterior probability greater than or equal to 0.95 were called as that genotype, while those with lower probabilities were set to missing. After imputation, quality control was rerun as described above for the imputed genotypes, leaving 1,442,839 SNPs available for analysis.

Quality control and genome-wide SNP-metabolite quantitative association analysis was performed using PLINK (Purcell et al. 2007).

Feature identification

Molecular identification of peaks from a non-targeted LC-MS metabolite profiling experiment is not straightforward (Wishart 2011; Theodoridis, Gika, and Wilson 2011). Here, the following workflow was utilized for feature annotation. First, accurate mass measurements were searched against a variety of metabolite databases including the Human Metabolome Database (http://www.hmdb.ca/), Metlin (http://metlin.scripps.edu/), and LipidMaps (http://www.lipidmaps.org/). Second, a combination of the accurate mass measurement and the isotopic distribution of the mass spectrometry peaks were imported into the elemental composition calculator (Waters MassLynx software) to generate a "best fit" molecular formula. Next, the best fit molecular formula was used to filter the database search results to yield a putative metabolite identification. Last, whenever possible, fragmentation information for the metabolite feature was compared with fragmentation of the putative metabolite found in the literature, and/or mass spectral

database, and/or from a commercially available pure standard.

Fragmentation spectra were collected from representative pooled serum samples using a Waters Acquity UPLC coupled with a Waters Xevo G2 TOF MS. Chromatographic conditions were identical to those described above. Mass spectrometry data was collected in positive ion mode, scanning from 50 to 1200 m/z at a rate of 5 scans per second with a 0.014 second inter-scan delay. Calibration was performed prior to sample analysis via infusion of sodium formate solution (0.01 M) in 80 % acetonitrile and 20% water, yielding a mass accuracy within 2 ppm RMS. The capillary voltage was held at 0.8 kV, source temperature at 130, and the desolvation temperature at 450 with a nitrogen gas flow rate of 1200 liters per hour. Data was collected in MS^E mode in which the collision cell voltage is switched between a low voltage state (4 V) and high voltage state (ramped from 12 to 28 V over 200 ms) on alternate acquisitions to generate both molecular mass measurement and fragmentation data. A method recently described by Broeckling et al., was utilized for the reconstruction of MS^E spectra for each significant molecular feature (Broeckling et al. 2012).

We report metabolite identification confidence based on metabolomics standards initiative recommendations (Sumner et al. 2007). Specifically, level 1 refers to confident molecular identification based on orthogonal analytical parameters (accurate mass, retention time, and MS/MS fragmentation) relative to an authentic compound. Level 2 refers to a putative identification based on physicochemical properties and/or spectral similarity with literature or spectral libraries. Level 3 refers to the putative identification of a compound class based on physicochemical properties or spectral similarity. Level 4 refers to an unknown compound.

Results

In total, 6138 metabolite peaks were called from the raw LC-MS data using XCMS. Peaks were assigned identifiers on the format "{mass/charge ratio}_{median retention time}", which will be used henceforth when referring to specific features. Initially we performed association analysis between PC status and each LC-MS feature in linear regression models adjusted for age and sample storage time. A quantile-quantile plot of observed versus expected $\log_{10}(P)$-values is given in Figure 1a, indicating a slight excess of significant tests. In Table 2, details of the top four significant associations (P<1.0E-3) are given. Applying a Bonferroni correction (significance threshold = 0.05/6138 ≈ 8.1E-06), two LC-MS features remained study-wide significant (595.4_153, P=4.0E-6; 422.2_315, p=7.1E-6).

The four metabolites most strongly associated to PC were explored for association with 1.4 million SNPs distributed across the genome. An overview of the results is given in a manhattan plot in Figure 2 (-$\log_{10}(P)$-value vs genomic position). The position of the SNP with the lowest P-value for each feature in Table 2 is reported in Table 3, along with the marker's location in relation to the nearest annotated gene. For each genome-wide set of metabolite-SNP tests, the Bonferroni-corrected study significance threshold is 0.05/1442840 ≈ 3.5E-08.

For one of the four metabolites, study-wide significance was observed; abundance of metabolite feature 174.1_53 was associated with the SNP rs2247035 at a significance level of 1.4E-08. This SNP is located in an intron of the gene interleukin 13 receptor, alpha 1 (*IL13RA1*) on chromosome Xq24. Furthermore, for each of the other three metabolites, the strongest association with genetic markers was observed within an intron of an annotated gene; phosphodiesterase 7B (*PDE7B*) on chromosome 6q23, neuregulin 3 (*NRG3*) on chromosome 10q23, and UDP glycosyltransferase 3 family, polypeptide A1 (*UGT3A1*) on chromosome 5p13.

Next we explored pairwise log(metabolite) differences (corresponding to ratios on the original scale) for association to PC. Metabolite ratios have been suggested to show more robust

associations than single metabolic features, since they may correlate with enzyme function or flow through metabolic pathways. Top results from the analysis of 6138*6137/2 pairwise differences are presented in Figure 1b and Table 4. No metabolite pair was study-wide significantly associated to PC after Bonferroni correction for the number of tests performed (significance threshold ≈ 2.7E-09). Seven metabolite feature pairs were associated at a significance threshold of 1.0E-7. Five of these pairs involved the metabolic feature 595.4_153, which was the most strongly associated feature in univariate analyses. Further metabolite features that were implicated in these pairwise assessments were 114.1_118, 411.3_285, 443.3_275, 451.2_266 and 597.4_306. Each of the two remaining pairs significant at the 1.0E-7 threshold included feature 422.2_315, the second most strongly univariate associated metabolite, in combination with features 226.2_212 and 581.3_446.

The molecular identities of the most strongly associated features from univariate and pairwise assessments were determined according to the workflow described in Materials and Methods. For the four most strongly associated features in univariate analysis (Table 2) we were unsuccessful in providing the molecular identity of any of the feature. From the pairwise analysis we were able to identify three of the seven additional features implicated including caprolactam, L-Phosphatidic acid and the peptide Tyr-Lys-Thr. Each of these molecules were implicated in combination with 595.4_153, the most strongly associated metabolite feature. We were unable to retrieve molecular identities for any of the two features implicated in combination with the 422.2_315 feature (Table 4).

Discussion

In this study we applied a global untargeted UPLC-MS strategy to identify novel biomarkers for PC detection. Utilizing serum samples from prostate cancer patients collected at time of diagnosis, before initiation of any treatment, and from unaffected population controls, a total of 6138 metabolic features were explored for association with disease status. Potential biomarker utility of explored features was assessed by contrasting normalized abundance levels across controls, patients with indolent disease, and patients with more aggressive disease. Overall the results from this study were negative. In univariate analysis only two of the 6138 metabolic features explored were observed as study-wide significantly associated with disease status after correction for multiple testing. Moreover, we were not able to derive the molecular identity of the two most strongly associated features, probably due to their low observed abundance. Since a robust association is a necessary (but not sufficient) criterion for a new biomarker, the immediate usefulness of these findings as biomarkers is low.

The four metabolite features showing the strongest association to prostate cancer were explored for association with SNPs genome-wide. Interestingly, for each investigated metabolite the strongest SNP association was observed within an annotated gene.

The metabolite feature 595.4_153 was associated with variation in the gene *PDE7B*, whose protein product hydrolyzes the second messenger cAMP, a key regulator of many important physiological processes. Variation in the gene *NRG3*, encoding a direct ligand for the ERBB4 tyrosine kinase receptor, was most strongly associated with metabolite 422.2_315. Neuregulins act as growth factors, and have been suggested in the aetiology of several cancers, including prostate and breast (Montero et al. 2008). Levels of the feature 174.1_53 were associated with variation in the interleukin 13 receptor, alpha 1 chain (*IL13RA1*). Though *IL13RA1* itself has not to our knowledge been associated with PC before, the alpha 2 chain of the same receptor (*IL13RA2*) has been reported to be differentially expressed in a metastatic prostate cancer cell line, and suggested as a target for prostate cancer treatment (He et al. 2010). Finally, the metabolite feature 260_142 was associated with variants in the gene *UGT3A1*, whose protein product conjugates substrates with N-acetylglucosamine to increase water solubility and enhance excretion. UGT3A1 acts on steroids,

particularly estrogen analogs (Meech and Mackenzie 2010, 3), and hypermethylation of this gene in breast cancer tissue has been associated with tumor relapse and worse survival (Hill et al. 2011). Despite our inability to derive explicit molecular identity of the top four PC-associated metabolite features, these genetic mapping results implicates these features in biological processes possibly related to PC aetiology.

We also examined pairwise differences between all LC-MS features (corresponding to ratios on the original scale of measurement) for association to PC. If differential activity of enzymes were associated with PC status, then the ratio of substrate to product would better reflect that association then either abundance by itself. However, no such difference was study-wide statistically significant after correction for multiple testing. Of the seven strongest associated feature-pairs, five involved the feature 595.4_153, which showed the strongest association in univariate analyses. We were able to determine the molecular identity of three of the five metabolite features implicated in combination with 595.4_153 including caprolactam, L-Phosphatidic acid and the peptide Tyr-Lys-Thr. Of note, caprolactam is a non-endogenous compound used in the manufacturing of nylon and produced around the world in very large quantities. Phosphatidic acids are fatty acid derivatives of glycerophosphates, and are established intracellular signaling lipids.

Major strengths of this study include the large sample size, and the unbiased assessment of serum metabolite features obtained using UPLC-MS. Furthermore, the availability of genome-wide SNP data for the same samples allowed us to further characterize the PC-associated features, and speculate on their role in biological pathways related to PC.

Limitations include the difficulties in mapping LC-MS features to molecular identities with sufficient certainty. Furthermore, the sampling strategy was in retrospect found to be suboptimal. Since age was seen to have a strong effect on the levels of many metabolite features (data not shown), as well as being strongly associated to PC status, there was a potential for confounding in the statistical analysis. Sample storage time showed the same problematic properties, because only cases were collected for the first six months of the study. These limitations were partly overcome by adjusting for these potential confounding factors in regression analysis, but the power to detect differential metabolites would have been greater had the sample been more balanced in terms of age and storage time between groups.

UPLC-MS is known to only capture part of the human serum metabolome (Psychogios et al. 2011). Potential prostate cancer biomarkers in the spectrum of molecules outside the UPLC-MS-detectable could thus not be assessed in this study. Studies combining several untargeted detection methods such as UPLC-MS, gas chromatography-mass spectrometry (GC-MS), and nuclear magnetic resonance (NMR) spectroscopy, would increase the possibilities of finding new disease-associated molecules.

In summary, we have examined the human UPLC-MS-detectable serum metabolome for prostate cancer biomarkers in a moderately sized Swedish case-control sample. No features of immediate biomarker utility were found, and most features showing association to PC status could not be tied to a molecule identity with certainty. Patient age and serum sample handling were identified as important covariates to consider when designing and analyzing untargeted metabolomics data.

The success of genome-wide association studies (GWAS) in uncovering new variant-disease associations has shown that an untargeted (or at least very broadly targeted) approach can add important new knowledge to disease aetiology. Since the metabolome is "downstream" of the genome in the path to disease, it makes intuitive sense that disease status and severity should be reflected by metabolomic changes. However, the metabolome is nowhere near as constant as the genome over time, and the field of metabolomics for disease assessment is still in its infancy. If the successes of the GWAS era are to be replicated in metabolomics, increased rigor in sample collection and handling strategies, refinement of biochemical and statistical methods, and increased

sample sizes will be required.

References

Adolfsson, Jan, Hans Garmo, Eberhard Varenhorst, Göran Ahlgren, Christer Ahlstrand, Ove Andrén, Anna Bill-Axelson, et al. 2007. "Clinical Characteristics and Primary Treatment of Prostate Cancer in Sweden Between 1996 and 2005." *Scandinavian Journal of Urology and Nephrology* 41 (6): 456–477. doi:10.1080/00365590701673625.

Altmaier, Elisabeth, Steven L. Ramsay, Armin Graber, Hans-Werner Mewes, Klaus M. Weinberger, and Karsten Suhre. 2008. "Bioinformatics Analysis of Targeted Metabolomics—Uncovering Old and New Tales of Diabetic Mice Under Medication." *Endocrinology* 149 (7) (July 1): 3478–3489. doi:10.1210/en.2007-1747.

Becker, Susen, Linda Kortz, Christin Helmschrodt, Joachim Thiery, and Uta Ceglarek. 2012. "LC–MS-based Metabolomics in the Clinical Laboratory." *Journal of Chromatography B* 883–884 (February 1): 68–75. doi:10.1016/j.jchromb.2011.10.018.

Broeckling, Corey, Adam Heuberger, Jonathan Prince, E. Ingelsson, and Jessica Prenni. 2012. "Assigning Precursor–product Ion Relationships in Indiscriminant MS/MS Data from Non-targeted Metabolite Profiling Studies." *Metabolomics*: 1–11. doi:10.1007/s11306-012-0426-4.

Crawford, E. David. 2009. "Understanding the Epidemiology, Natural History, and Key Pathways Involved in Prostate Cancer." *Urology* 73 (5, Supplement) (May): S4–S10. doi:10.1016/j.urology.2009.03.001.

Fall, Katja, Hans Garmo, Ove Andrén, Anna Bill-Axelson, Jan Adolfsson, Hans-Olov Adami, Jan-Erik Johansson, and Lars Holmberg. 2007. "Prostate-specific Antigen Levels as a Predictor of Lethal Prostate Cancer." *Journal of the National Cancer Institute* 99 (7) (April 4): 526–532. doi:10.1093/jnci/djk110.

He, Hui, Jianchun Xu, Peter S Nelson, Fray F Marshall, Leland W K Chung, Haiyen E Zhau, Dalin He, and Ruoxiang Wang. 2010. "Differential Expression of the Alpha2 Chain of the Interleukin-13 Receptor in Metastatic Human Prostate Cancer ARCaPM Cells." *The Prostate* 70 (9) (June 15): 993–1001. doi:10.1002/pros.21133.

Hill, Victoria K., Christopher Ricketts, Ivan Bieche, Sophie Vacher, Dean Gentle, Cheryl Lewis, Eamonn R. Maher, and Farida Latif. 2011. "Genome-Wide DNA Methylation Profiling of CpG Islands in Breast Cancer Identifies Novel Genes Associated with Tumorigenicity." *Cancer Research* 71 (8) (April 15): 2988–2999. doi:10.1158/0008-5472.CAN-10-4026.

Issaq, Haleem J., and Timothy D. Veenstra. 2011. "Is Sarcosine a Biomarker for Prostate Cancer?" *Journal of Separation Science* 34 (24): 3619–3621. doi:10.1002/jssc.201100572.

Johansson, Jan-Erik, Ove Andrén, Swen-Olof Andersson, Paul W Dickman, Lars Holmberg, Anders Magnuson, and Hans-Olov Adami. 2004. "Natural History of Early, Localized Prostate Cancer." *JAMA: The Journal of the American Medical Association* 291 (22) (June 9): 2713–2719. doi:10.1001/jama.291.22.2713.

Klotz, Laurence. 2005. "Active Surveillance with Selective Delayed Intervention Is the Way to Manage 'Good-risk' Prostate Cancer." *Nature Clinical Practice Urology* 2 (3): 136–142. doi:10.1038/ncpuro0124.

Lindmark, Fredrik, S. Lilly Zheng, Fredrik Wiklund, Jeannette Bensen, Katarina Augustsson Balter, Baoli Chang, Maria Hedelin, et al. 2004. "H6D Polymorphism in Macrophage-Inhibitory Cytokine-1 Gene Associated With Prostate Cancer." *J. Natl. Cancer Inst.* 96 (16) (August 18): 1248–1254. doi:10.1093/jnci/djh227.

Marchini, Jonathan, Bryan Howie, Simon Myers, Gil McVean, and Peter Donnelly. 2007. "A New Multipoint Method for Genome-wide Association Studies by Imputation of Genotypes." *Nat Genet* 39 (7) (July): 906–913. doi:10.1038/ng2088.

Meech, Robyn, and Peter I. Mackenzie. 2010. "UGT3A: Novel UDP-glycosyltransferases of the UGT Superfamily." *Drug Metabolism Reviews* 42 (1) (February): 45–54. doi:10.3109/03602530903205823.

Montero, Juan Carlos, Ruth Rodríguez-Barrueco, Alberto Ocaña, Elena Díaz-Rodríguez, Azucena Esparís-Ogando, and Atanasio Pandiella. 2008. "Neuregulins and Cancer." *Clinical Cancer Research* 14 (11) (June 1): 3237–3241. doi:10.1158/1078-0432.CCR-07-5133.

Patti, Gary J. 2011. "Separation Strategies for Untargeted Metabolomics." *Journal of Separation Science* 34 (24): 3460–3469. doi:10.1002/jssc.201100532.

Psychogios, Nikolaos, David D. Hau, Jun Peng, An Chi Guo, Rupasri Mandal, Souhaila Bouatra, Igor Sinelnikov, et al. 2011. "The Human Serum Metabolome." *PLoS ONE* 6 (2) (February 16): e16957. doi:10.1371/journal.pone.0016957.

Purcell, Shaun, Benjamin Neale, Kathe Todd-Brown, Lori Thomas, Manuel A R Ferreira, David Bender, Julian Maller, et al. 2007. "PLINK: a Tool Set for Whole-genome Association and Population-based Linkage Analyses." *American Journal of Human Genetics* 81 (3) (September): 559–75. doi:10.1086/519795.

R Development Core Team. 2012. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org.

Smith, Colin A., Elizabeth J. Want, Grace O'Maille, Ruben Abagyan, and Gary Siuzdak. 2006. "XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification." *Analytical Chemistry* 78 (3) (February 1): 779–787. doi:10.1021/ac051437y.

Spratlin, Jennifer L., Natalie J. Serkova, and S. Gail Eckhardt. 2009. "Clinical Applications of Metabolomics in Oncology: A Review." *Clinical Cancer Research : an Official Journal of the American Association for Cancer Research* 15 (2) (January 15): 431–440. doi:10.1158/1078-0432.CCR-08-1059.

Sreekumar, Arun, Laila M. Poisson, Thekkelnaycke M. Rajendiran, Amjad P. Khan, Qi Cao, Jindan Yu, Bharathi Laxman, et al. 2009. "Metabolomic Profiles Delineate Potential Role for Sarcosine in Prostate Cancer Progression." *Nature* 457 (7231) (February 12): 910–914. doi:10.1038/nature07762.

Sumner, Lloyd, Alexander Amberg, Dave Barrett, Michael Beale, Richard Beger, Clare Daykin, Teresa Fan, et al. 2007. "Proposed Minimum Reporting Standards for Chemical Analysis." *Metabolomics* 3 (3): 211–221. doi:10.1007/s11306-007-0082-2.

The International HapMap Consortium. 2007. "A Second Generation Human Haplotype Map of over 3.1 Million SNPs." *Nature* 449 (7164) (October 18): 851–861. doi:10.1038/nature06258.

Theodoridis, Georgios, Helen G Gika, and Ian D Wilson. 2011. "Mass Spectrometry-based Holistic Analytical Approaches for Metabolite Profiling in Systems Biology Studies." *Mass Spectrometry Reviews* 30 (5) (September 1): 884–906. doi:10.1002/mas.20306.

Wishart, David S. 2011. "Advances in Metabolite Identification." *Bioanalysis* 3 (15) (August): 1769–1782. doi:10.4155/bio.11.155.

Xie, Baogang, Michael J. Waters, and Horst Joachim Schirra. 2012. "Investigating Potential Mechanisms of Obesity by Metabolomics." *Journal of Biomedicine and Biotechnology* 2012: 1–10. doi:10.1155/2012/805683.

Figures

Figure 1. a) Quantile-quantile plot of 6138 single metabolite to PC association tests. b) Quantile-quantile plot of 6138*6137/2 pairwise metabolite differences to PC association tests.
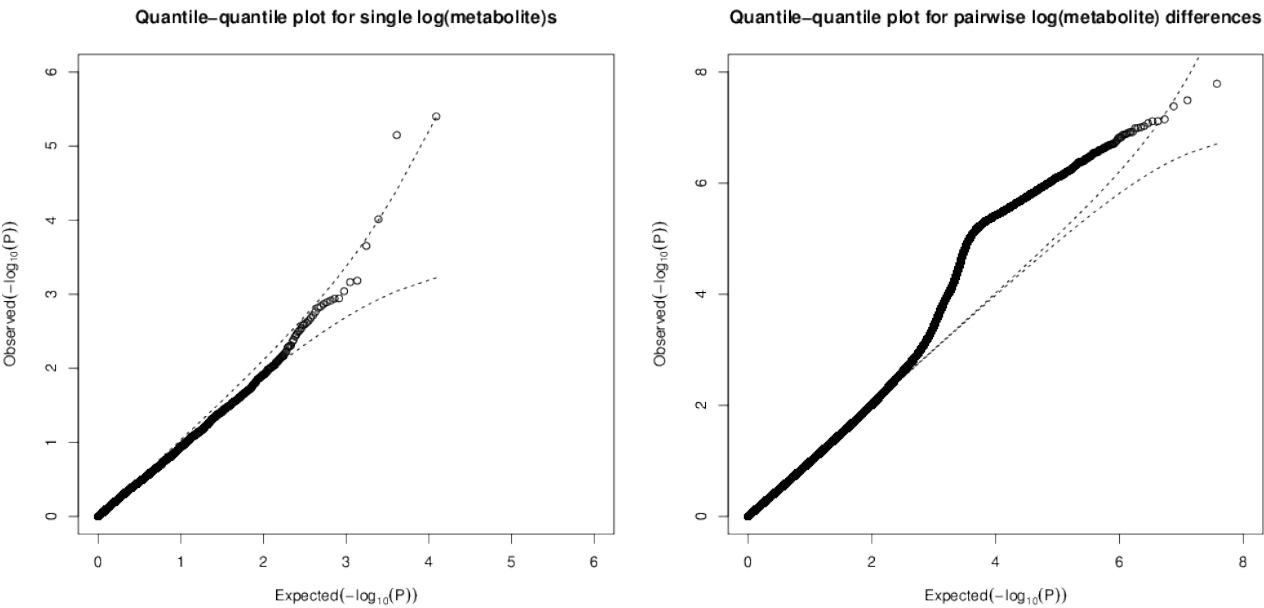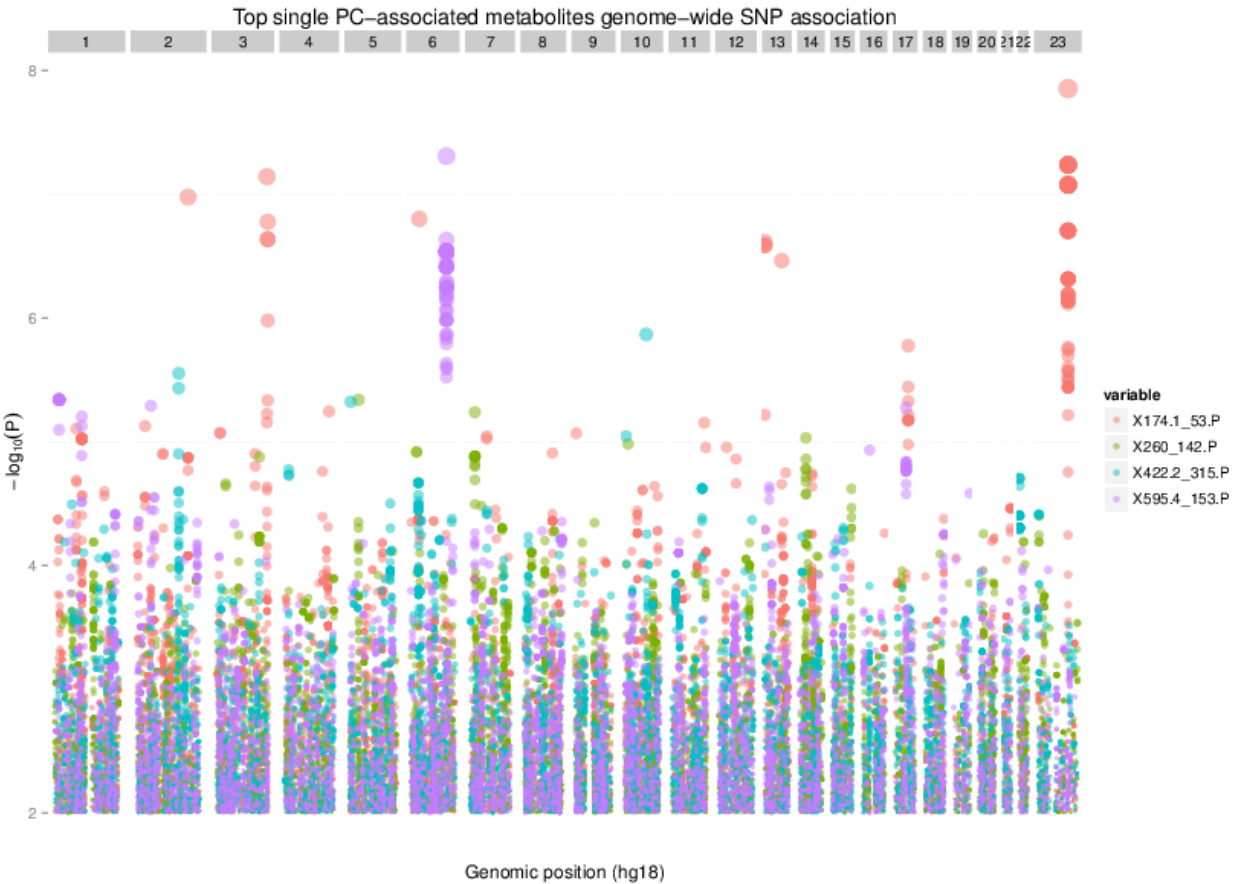

Figure 2. Manhattan plot of top PC-associated metabolites vs genome-wide SNPs

Tables

Table 1: Descriptive Statistics by Prostate cancer status

	Controls	Cases (less aggressive)	Cases (more aggressive)
	N = 188	N = 188	N = 99
Tumor stage (T):			
1 – 2		100% (188)	36% (36)
3 – 4		0%(0)	64% (63)
Nodal stage (N):			
0		6% (12)	15% (15)
1		0% (0)	5% (5)
X		94% (176)	80% (79)
Metastasis stage (M):			
0		24% (45)	46% (46)
1		0% (0)	5% (5)
X		76% (143)	48% (48)
Gleason score:			
2 – 6		100% (188)	28% (28)
7		0% (0)	31% (31)
8 – 10		0% (0)	26% (26)
NA		0% (0)	14% (14)
PSA (ng/ml)	0.9 (0.6 – 1.4)	6.6 (4.7 – 8.2)	19.6 (10.4 – 38.1)
Age at diagnosis/inclusion (years)	63.7 (60.1 – 70.7)	65.8 (61.4 – 70.5)	73.7 (66.5 – 77.1)
Body Mass Index (kg/m^2)	26.3 (24.2 – 27.8)	25.7 (24.1 – 28.1)	26.0 (23.9 – 28.7)
Sample storage time (days)			
2161 – 2421	39% (74)	38% (72)	26% (25)
2448 – 2716	46% (86)	29% (55)	28% (27)
2721 – 2990	15% (28)	21% (39)	24% (24)
3016 – 3276	0% (0)	12% (22)	22% (22)

Continuous variables are reported as "median (interquartile range)".
Numbers after percents are frequencies. X, not assessed. NA, not available. PSA, prostate specific antigen.

Table 2. Top PC-associated single metabolites

Molecular Feature (m/z_retention time)	PC overall association P-value	Metabolite Identification	Identification Confidence
595.4_153	4.0×10^{-06}	Unknown	4
422.2_315	7.1×10^{-06}	Unknown	4
174.1_53	6.6×10^{-04}	Unknown	4
260_142	9.1×10^{-04}	Unknown	4

M/Z, mass to charge ratio. P-values from ANOVA F-test (2 d.f.), adjusted for age at inclusion and sample storage time. Metabolites marked gray were significantly associated to PC status after Bonferroni correction for 6138 tests.

Table 3. GWAS results for top PC-associated single metabolites

Molecular Feature (m/z_retention time)	Metabolite GWAS lowest P	Position (hg18)	Nearest gene
595.4_153	4.9×10^{-08}	chr6:136374051	*PDE7B* (intron)
422.2_315	1.4×10^{-06}	chr10:83843772	*NRG3* (intron)
174.1_53	1.4×10^{-08}	chrX:117756115	*IL13RA1* (intron)
260_142	4.6×10^{-06}	chr5:35999264	*UGT3A1* (intron)

Metabolites marked gray were significantly associated to SNPs at the reported loci after Bonferroni correction for 1442840 tests.

Table 4. Top PC-associated pairwise metabolite differences

Molecular Feature Pair	PC overall association P-value	Metabolite Identification	Identification Confidence
595.4_153 - 114.1_118	3.2×10^{-08}	Unknown - Caprolactam	4 - 1
595.4_153 - 443.3_275	4.2×10^{-08}	Unknown - Unknown	4 - 4
597.4_306 - 595.4_153	7.1×10^{-08}	L-Phosphatidic acid - Unknown	2 - 4
595.4_153 - 451.2_266	7.7×10^{-08}	Unknown - Unknown	4 - 4
595.4_153 - 411.3_285	8.4×10^{-08}	Unknown - peptide (Tyr-Lys-Thr)	4 - 3
581.3_440 - 422.2_315	9.4×10^{-08}	Unknown - Unknown	4 - 4
422.2_315 - 226.2_212	1.0×10^{-07}	Unknown - Unknown	4 - 4

Serum metabolomics and prostate cancer survival. *R. Szulkin[1], R. Karlsson[1], A. Heuberger[2], M. Hong[1], C. Broeckling[2], J. Prenni[1], J. Prince[1], F. Wiklund[1]*
1) Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden; 2) Department of Biochemistry and Molecular Biology, Colorado State University, Fort Collins, USA.

Introduction: Established prognostic factors perform poorly in predicting disease relapse among patients treated for prostate cancer. Identification of novel biomarkers improving the prognostic information is of great importance to guide individual therapy. Materials and Methods: Post-treatment serum samples from a nested case-case design comprising 269 prostate cancer patients with lethal outcome and 269 patients with non-lethal outcome were used. All patients were diagnosed between year 2001 and 2003 in Sweden and followed up for survival until December 2010 through record linkage with the national Cause of Death Registry. Untargeted ultra performance liquid chromatography (UPLC) coupled with mass spectrometry (MS) was employed to screen for novel prostate cancer biomarkers. Normalized and log-transformed metabolite concentrations were explored for association with prostate cancer-specific survival in time-to-event analysis using death from prostate cancer as endpoint. Results: Untargeted metabolomic profiling of prostate cancer serum samples revealed a total of 5209 LC/MS profiles. Univariate analysis of individual normalized feature levels indicated 23 peaks to be study-wide significant associated with prostate cancer-specific survival. Of note, at the 1×10^{-8} significance level we observed 11 associated peaks as compared to 1×10^{-4} expected peaks under the null hypothesis of no association. Further assessment exploring pair-wise ratios between metabolomic peaks revealed additional features significantly associated with prostate cancer prognosis. Conclusion: Untargeted metabolomic profiling of prostate cancer serum samples have identified a considerable number of molecular features strongly associated with disease prognosis. Further analysis is underway to identify these profiles molecular identity and to explore molecular pathways involved.

You may contact the first author (during and after the meeting) at robert.szulkin@ki.se