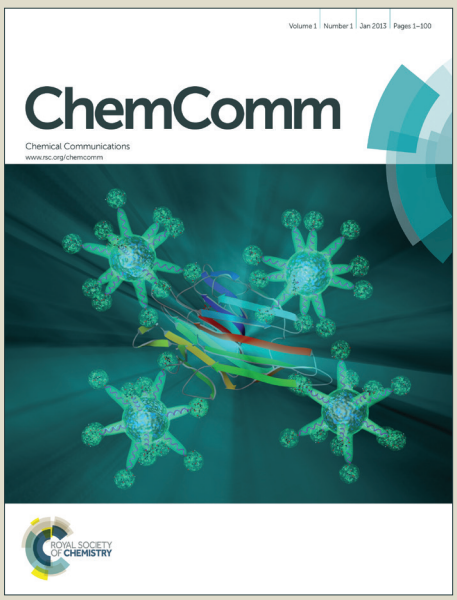
ChemComm

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/chemcomm

Journal Name

RSCPublishing

COMMUNICATION

Parafilm-Assisted Microdissection: A Sampling Method for Mass Spectrometry-Based Identification of Differentially Expressed Prostate Cancer Protein Biomarkers

Cite this: DOI: 10.1039/xoxxoooooox

Received ooth January 2012,
Accepted ooth January 2012

DOI: 10.1039/xoxxoooooox

www.rsc.org/

J. Quanico,[a,b] J. Franck[a] , J.P. Gimeno [a,c], R. Sabbagh[b], M. Salzet[a], R. Day[b] and I. Fournier[a]

Mass spectrometry-based methods for prostate cancer biomarker discovery are hampered by their low-throughput capabilities because of extensive sample preparation. We present the parafilm-assisted microdissection technique coupled with label-free quantification and bioinformatics analysis as a means to evaluate directly protein expression changes on benign and tumor regions.

The early diagnosis of prostate cancer (PC) remains a challenge because of the lack of reliable biomarkers that can effectively describe this disease independent of sample heterogeneity. Current methods for PC detection comprising of digital rectal exam (DRE), prostate-specific antigen (PSA) measurement, and transrectal ultrasound (TRUS) are being used for routine screening prior to undergoing more invasive follow-up procedures such as a biopsy. However, their use has been controversial because overdiagnosis of insignificant tumors led to unnecessary treatment and known harms associated with the tests[1]. Moreover, although elevated PSA levels correlate with PC incidence, PSA alone or in tandem with other screening methods cannot be used to provide reliable prognostic indicators on which decisions to undergo treatment despite of its deleterious effects rely on[2]. Clearly, there is a need for the establishment of biomarkers that can discern PC in its aggressive stages, in order to discriminate patients requiring immediate therapy from those whose likelihood of benefit from treatment is low.

To this end, mass spectrometry (MS) can serve as a method of choice for protein biomarker discovery. Specifically, MS-based shotgun proteomics has been demonstrated to allow the nearly complete characterization of large proteomes by taking advantage of robust developments in rapid data acquisition, high resolution, increased sensitivity and rapid duty cycle. MS-based shotgun proteomics is a powerful tool to detect differences in protein expression levels at low-attomole detection limits. Its major limitation for screening purposes though is that it is low-throughput, due to extensive sample clean-up, long chromatographic gradients during protein separation, and rigorous interpretation of large datasets generated. More importantly, information that can be gleaned from MS-based biomarker discovery-driven methods largely depends on the method of sampling involved. Conventional methods involving fractionation of gross tissues or organs often leads to loss of localization information that may be relevant especially if comparison between normal and diseased regions in the same tissue is desired. Also, less abundant proteins from the diseased region can be diluted among those that were obtained from the normal one. Direct MS-based sampling methods such as liquid extraction surface analysis (LESA[3]) and desorption electrospray ionization (DESI[4]) to name a few, as well as matrix-assisted laser desorption/ionization mass spectrometry (MALDI MS[5]) in both profiling and imaging modes, have been developed to allow direct mapping of protein distributions while minimizing sample preparation but these methods are restricted at best to tissue profiling studies because of their limited sensitivity and/or protein identification capabilities considering that they were not aimed at complete proteome characterization.

To be able to analyze more samples directly at a much faster rate using the MS-based shotgun proteomics approach, we have recently developed the parafilm-assisted microdissection technique (PAM) in order to sample regions of interest (ROIs) directly from tissue sections[6]. ROIs were initially identified by subjecting a tissue section to MALDI MSI analysis. An adjacent section is then mounted on a parafilm-covered glass slide and the ROIs identified previously were excised and subjected to enzymatic digestion and liquid chromatography tandem MS (LC MS). The PAM approach provided a means to obtain millimeter-sized portions that can be easily subjected to LC MS without the use of additional rigorous sample purification steps such as lipid or abundant protein removal, sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS PAGE) or protein precipitation. Obtaining samples of restricted size in such a manner, compared to purifying them from entire organ homogenates, minimizes the dilution of less abundant proteins, and also allows their localization on the tissue where they have been obtained. PAM can be extended to the entire tissue section to provide a map of the distribution of proteins[6]. This time we exploit the advantage of using the PAM approach in profiling mode to directly analyze tumor regions in prostate biopsies as defined by MSI analysis, with the objective of finding candidate protein biomarkers that can be further validated on a larger cohort of samples.

MALDI MS imaging was first used to define ROIs on both the benign and tumor regions of prostate sections. Although the spatial resolution used was only 100 µm, it was sufficient to provide images with discernible localized tumors that coincide very well with those identified by HES staining (Figure 1a and b). But more than just providing optical images, the data generated are actually molecular images. Hierarchical clustering (HC) analysis of the spectra of these molecules (Figure 1c and d) results in simplified images where ROIs are defined based on statistical inference gained from the similarity of the cellular contents of neighbouring cells. This greatly simplifies the selection criterion used to identify ROIs, in contrast to looking at cell morphologies which can at times be subjective or misleading and heavily rely on the experience of the analyst. In the case of prostate tissue, the latter task easily becomes complicated because of the innate heterogeneity of the tissue. We have previously shown that lipid MS images are sufficient alternatives to protein MS images for HC as both show similar component localization but with lipid imaging being done much faster and at better spatial resolution[6]. Lipids, in contrast to proteins, do not need to be incorporated into the matrix crystals to be desorbed in MALDI; small matrix crystals thus do not compromise the efficiency of the ionization/desorption process.

After determining the ROIs, they were excised and subjected to protein extraction, overnight trypsin digestion and shotgun LC-MS. Initial results on four prostate tissue samples subjected to the PAM technique and using an Orbitrap Elite instrument yielded 1221 protein identifications, 374 of which have Fisher's Test p-values ≤ 0.05 when the tumor dataset was compared against the benign set. A total of 122 proteins were identified to have differential expression after filtering using fold change and detectability across samples criteria, 79 of which were classifed as upregulated and 43 as downregulated. In this experiment, single PAM pieces were taken from the benign and tumor regions. In succeeding experiments using the Q-exactive, two PAM pieces from each region were taken from 8 prostate tissue samples and one from a sample of limited size. This yielded 1251 protein IDs, 485 of which fit the Fisher's test criterion. Applying the filtering parameters yielded 208 differentially expressed proteins of which, 135 were upregulated and 73 downregulated. Thus, a total of 273 differentially expressed proteins are reported in this work (Tables 1 and S1).

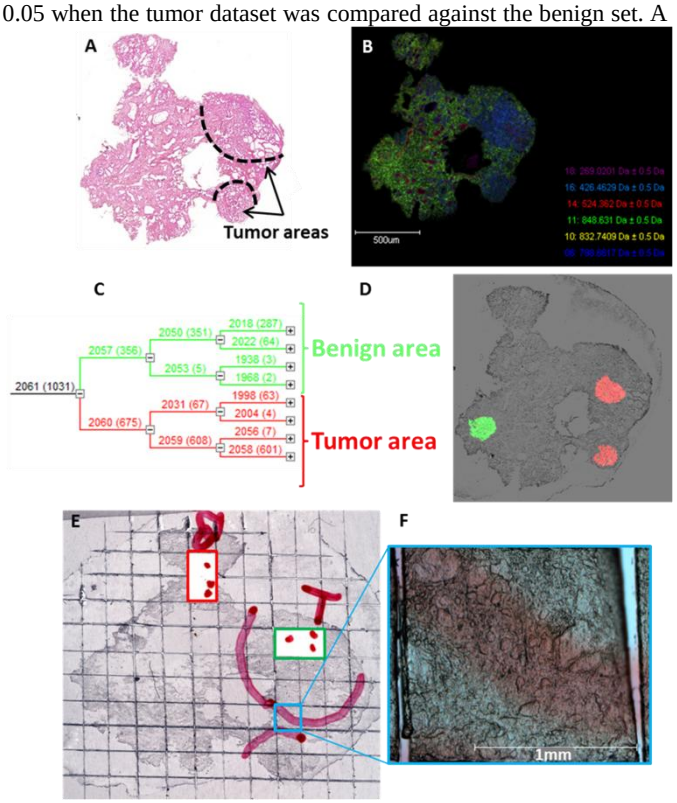


Figure 1. Hierarchical clustering of prostate tumor MS images to define ROIs. A) Tumor regions in HES-stained section. B) Composite image of selected peaks taken from the MS image of an adjacent section. C) Dendrogram of the clustering analysis. D) Clustered spectra plotted back to the optical image. E) Sampled regions on the adjacent section. F) Detail of a single PAM piece.

Examination of the upregulated proteins with 8-16 fold change identified previously reported proteins implicated in PC processes. This includes growth differentiation factor 15 (GDF15), hypoxia upregulated protein 1 (HYOU1), periostin (POSTN) and poly(ADP-ribose) polymerase 1 (PARP1). Proteins implicated in other cancers were also identified, such as hydroxyacyl-coenzyme A (CoA) dehydrogenase/3-ketoacyl-CoA thiolase/enoyl-CoA hydratase (trifunctional protein), alpha subunit (HADHA)[7] and tripartite motif containing 28 (TRIM28)[8]. Highly downregulated proteins are also noted in the list, including PC risk-modulating factor glutathione S-transferase mu 3 (GSTM3)[9], and α-microseminoprotein (MSMB)[10].

The differentially expressed proteins were examined further by establishing the predicted protein-protein interactions (PPI) using STRING, plotting the PPIs using Cytoscape, analyzing the network parameters using NetworkAnalyzer, and performing MCL clustering and Gene Ontology (GO) analysis using clusterMaker and ClueGO, respectively. These were performed in order to determine which pathways were significantly perturbed in the tumor regions. The advantage of this approach is that the results are less prone to individual genetic heterogeneity that greatly decreases the discriminative power of individual markers[11]. Also, proteins and their genes rarely act alone, thus, for their cellular functions to be understood, they have to be examined in the context of the pathways where their roles interplay with those of other proteins[12]. The other

Table 1. Over- and underexpressed proteins with 8-16 fold change observed when the spectral counts of PAM pieces from tumor regions are compared with those from benign regions. Negative values denote underexpression. (For complete list, see Table S1.)

Protein Name	Log2 Fold Change
GOLM1	4
PRKDC	4
POSTN	4
RPL23	4
GDF15	4
HYOU1	4
CTNNA1	4
PSME2	3
PARP1	3
HADHA	3
DAK	3
TRIM28	3
GSTM3	-3
OLFML1	-3
A2M	-3
NID2	-3
MAP1B	-5
MSMB	-5

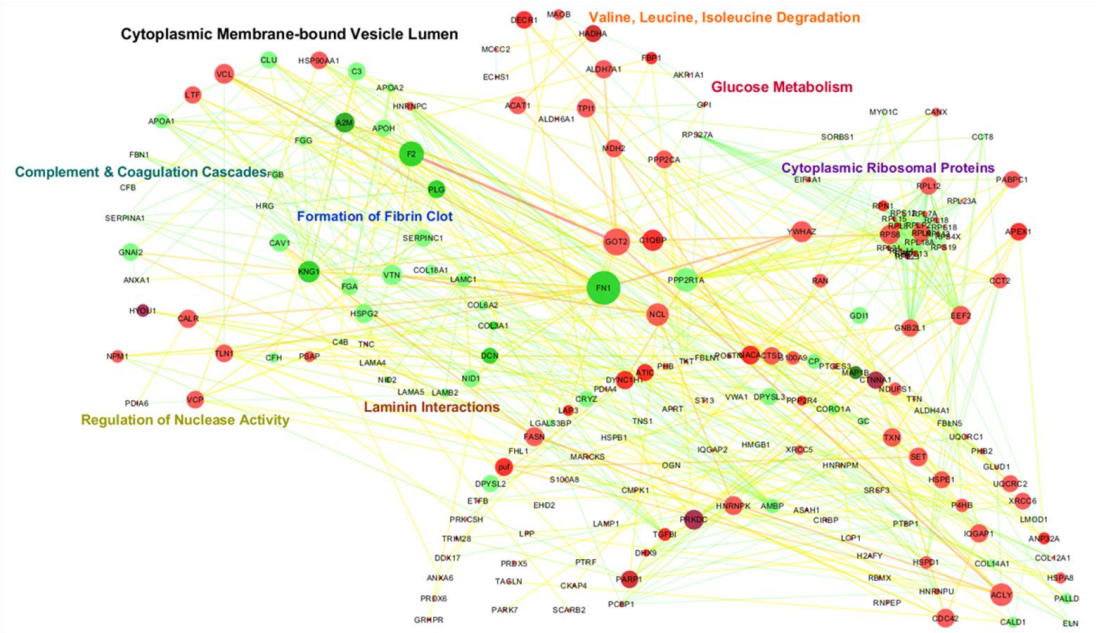


Figure 2. Regulatory network plotted using Betweenness Centrality (node size) and log2Fold Change (color) using the differentially expressed protein dataset. Color scale: green = downregulation, red = upregulation. Color intensity denotes extent of down- or upregulation in the tumor region, with the most intense colors (red or green) corresponding to 16-fold change in spectral count value.

advantage is that, with network analysis, we can perform MCL clustering to identify core modules and locate hubs. Hubs are large nodes in the network that have many connections with the smaller nodes; their centrality denotes that they have more functional relevance compared to other proteins/genes[13]. Hubs are essential for keeping the integrity of networks because they are involved in the relay of information, thus, proteins occupying these positions become more promising biomarker candidates for more precise disease classification[14]. To determine which pathways were enriched in the benign and tumor regions of the prostate samples, the differentially expressed protein list was analyzed using ClueGO. ClueGO can integrate GO terms, KEGG/BioCarta pathways, Reactome and WikiPathways terms and can create a functionally grouped GO/pathway network using kappa statistics for term linkage[15], in contrast to other GO analysis software that generate hierarchical trees based on assessment of overrepresented GO terms.

Results of the network and GO analysis are summarized in Figure 2 and Table S2. After constructing the regulatory network using the PPIs and mapping using network parameters such as degree and betweenness centrality, the proteins were grouped according to their GO term association. The node colors were then changed to log2Fold Change to show the distinctly perturbed pathways based not only on protein member overrepresentation, but also expression. As expected, distinct pathways were perturbed. Majority of the upregulated proteins in the tumor region are involved in cytoplasmic ribosomal protein synthesis and function, including 22 ribosomal proteins. The overexpression of cytoplasmic ribosomal proteins is concomitant with the increased metabolic activity necessary to drive the proliferation of neoplastic cells[16]. Ribosomal proteins have also been proposed to actively mediate certain aspects of tumorigenesis due to the extra-ribosomal cellular functions that they possess independent of protein biosynthesis[17]. Likewise, the enrichment of metabolism-related terms in the tumor region, such as valine, leucine and isoleucine degradation and glucose metabolism, is correlated with increased cell proliferation. Such processes can be utilized by the tumor cells to drive energy production and provide metabolites needed for cellular processes and protection against oxidative stress[18]. Meanwhile, ClueGO analysis of the downregulated terms showed enrichment of proteins involved in complement and coagulation pathways, essential components of the immune system used in the eradication of pathogens and clot formation; such diminished immune system activity is essential for continued survival of tumor cells. Laminin interactions and cytoplasmic membrane-bounded vesicle lumen were also downregulated, in accord with the alterations in the extracellular matrix to compromise the basement membrane that acts as a barrier for cancer cell invasion during epithelial-mesenchymal transition (EMT)[19].

To identify hubs in the network, MCL clustering was performed (Figure S2). Five modules were identified from the clusters and the hubs were pulled together and placed at the center of the regulatory network (Figure 2). Exploring further, the cancer gene indices (CGI) of the modules were loaded using Reactome. CGI contains data on 6,955 human genes, nearly 12,000 NCI Thesaurus cancer disease terms, and 2,180 unique pharmacological compounds from the NCI Thesaurus[20]. As shown in Figure S3, the highlighted nodes (yellow) indicate proteins annotated with the term "neoplasm". Neoplasms are abnormal growths of tissue that may or may not necessarily produce a mass[21]. Expectedly, majority of these hubs have well established roles in the regulation of cancer processes, as supported by their CGI annotations. For example, fibronectin 1 (FN1) is known for its role in suppressing PC cell migration via fibronectin matrix-mediated cohesion[22]. Of the five hubs identified, only glutamate oxaloacetate transferase 2 (GOT2) did not possess CGI annotations, suggesting that this protein is probably involved in non-neoplastic processes. GOT2 encodes for the mitochondrial isozyme of aspartate aminotransferase (AAT), and catalyzes the conversion of

oxaloacetate to aspartate. GOT2, together with mitochondrial malate dehydrogenase 2 (MDH2), comprises the malate-aspartate shuttle of the glycolysis pathway. The malate-aspartate shuttle is known to be active in neoplastic cells of several tumor types and believed to account for about 20% of the total respiratory rate[23].

Conclusions

The current work provided a simple means to evaluate protein expression data by making use of a combination of MS imaging, PAM and LC-MS methods to sample and analyze regions directly from tissue specimens without extensive sample preparation or purification. A combination of rigid protein identification and spectral counting parameters followed by robust statistical methods were then applied to obtain high quality differential protein expression data. Examination of the differentially expressed proteins showed established biomarkers for PC diagnosis, as well as potential candidates for further verification. Apart from individual protein expression analysis, we also used network and GO analyses to identify FN1, GOT2, NCL, PPP2R1A and C1QBP as key regulators of PC among the hundreds of proteins that we identified to have differential expression in the tumor region. This network-based approach supports existing evidence on the active roles of these proteins in carcinogenesis, and allowed us to visualize the relative importance of the protein interactions based on well-defined topological parameters that may aid in the better understanding of the underlying mechanisms of the disease and provide avenues for specific protein therapeutic targets. Results of this work need to be further validated to realize their potential in the clinical setting. These experiments highlight the advantage of localized microproteomic analysis offered by the PAM strategy; we thus expect to find more applications of this strategy in our future proteomics studies.

Acknowledgements

This work was funded by the by grants from SIRIC ONCOLille (IF), Grant INCa-DGOS-Inserm 6041a, the Canadian Cancer Society (RD), Prostate Cancer Canada and the Movember Foundation (RD an RS).

Notes and references

[a]Laboratoire PRISM : Protéomique, Réponse Inflammatoire, Spectrométrie de masse, INSERM, Université de Lille 1, Bât. SN3, 1er étage, F-59655 Villeneuve D'Ascq, France.
[b]Institut de Pharmacologie, Département de Chirurgie/Service d'Urologie, Faculté de Médecine et des Sciences de la Santé, Université de Sherbrooke, Sherbrooke, J1H 5N4 Québec, Canada.
[c]SIRIC ONCOLILLE, Maison Régionale de la Recherche Clinique, 6 Rue du Professeur Laguesse, 59037 Lille, France.
Electronic Supplementary Information (ESI) available: [Supplemental Figures and Tables, and Experimental Procedures]. See DOI: 10.1039/c000000x/

1. D. M. Saman, A. M. Lemieux, M. Nawal Lutfiyya and M. S. Lipsky, *Dis Mon*, 2014, **60**, 150-154.
2. A. M. Wolf, R. C. Wender, R. B. Etzioni, I. M. Thompson, A. V. D'Amico, R. J. Volk, D. D. Brooks, C. Dash, I. Guessous, K. Andrews, C. DeSantis and R. A. Smith, *CA Cancer J Clin*, 2010, **60**, 70-98.
3. J. Quanico, J. Franck, C. Dauly, K. Strupat, J. Dupuy, R. Day, M. Salzet, I. Fournier and M. Wisztorski, *J Proteomics*, 2013, **79**, 200-218.
4. J. Liu and Z. Ouyang, *Anal Bioanal Chem*, 2013, **405**, 5645-5653.
5. D. C. Anderson and K. Kodukula, *Biochem Pharmacol*, 2014, **87**, 172-188.
6. J. Franck, J. Quanico, M. Wisztorski, R. Day, M. Salzet and I. Fournier, *Anal Chem*, 2013, **85**, 8127-8134.
7. J. Dai, J. Shen, W. Pan, S. Shen and U. N. Das, *Lipids Health Dis*, 2013, **12**, 71.
8. L. Chen, T. Munoz-Antonia and W. D. Cress, *PLoS One*, 2014, **9**, e101040.
9. R. Medeiros, A. Vasconcelos, S. Costa, D. Pinto, P. Ferreira, F. Lobo, A. Morais, J. Oliveira and C. Lopes, *Prostate*, 2004, **58**, 414-420.
10. C. A. Haiman, D. O. Stram, A. J. Vickers, L. R. Wilkens, K. Braun, C. Valtonen-Andre, M. Peltola, K. Pettersson, K. M. Waters, L. L. Marchand, L. N. Kolonel, B. E. Henderson and H. Lilja, *J Natl Cancer Inst*, 2012, **105**, 237-243.
11. M. T. Dittrich, G. W. Klau, A. Rosenwald, T. Dandekar and T. Muller, *Bioinformatics*, 2008, **24**, i223-231.
12. R. Yang, B. J. Daigle, Jr., L. R. Petzold and F. J. Doyle, 3rd, *BMC Bioinformatics*, 2012, **13**, 12.
13. H. Jeong, S. P. Mason, A. L. Barabasi and Z. N. Oltvai, *Nature*, 2001, **411**, 41-42.
14. N. M. Penrod and J. H. Moore, *BMC Syst Biol*, 2014, **8**, 12.
15. G. Bindea, B. Mlecnik, H. Hackl, P. Charoentong, M. Tosolini, A. Kirilovsky, W. H. Fridman, F. Pages, Z. Trajanoski and J. Galon, *Bioinformatics*, 2009, **25**, 1091-1093.
16. T. Teng, G. Thomas and C. A. Mercer, *Curr Opin Genet Dev*, 2013, **23**, 63-71.
17. A. Ziemiecki, R. G. Muller, X. C. Fu, N. E. Hynes and S. Kozma, *EMBO J*, 1990, **9**, 191-196.
18. M. G. Vander Heiden, L. C. Cantley and C. B. Thompson, *Science*, 2009, **324**, 1029-1033.
19. P. Lu, V. M. Weaver and Z. Werb, *J Cell Biol*, 2012, **196**, 395-406.
20. M. Milacic, R. Haw, K. Rothfels, G. Wu, D. Croft, H. Hermjakob, P. D'Eustachio and L. Stein, *Cancers (Basel)*, 2012, **4**, 1180-1211.
21. G. Fragoso, S. de Coronado, M. Haber, F. Hartel and L. Wright, *Comparative and functional genomics*, 2004, **5**, 648-654.
22. D. Jia, I. Entersz, C. Butler and R. A. Foty, *BMC Cancer*, 2012, **12**, 94.
23. J. M. Thornburg, K. K. Nelson, B. F. Clem, A. N. Lane, S. Arumugam, A. Simmons, J. W. Eaton, S. Telang and J. Chesney, *Breast cancer research : BCR*, 2008, **10**, R84.

