

Reviewers' comments:

Reviewer #1 (Remarks to the Author):

This manuscript describes quantitative proteomic analysis of 37 prostate tissue samples from three groups of patients using SWATH mass spectrometry and reports expression of 3394 proteins in all samples. The authors then compare the proteomic data with published genomic and transcriptomic data from some of the samples, and identified correlation between miRNA and some proteins. By comparing proteomics data among patient groups, the authors proposed the TCA cycle as a potentially critical process during prostate cancer progression. Two TCA proteins (ACO2 and MDH2) are validated using WB, and further validated using larger sample sets using IHC.

The miRNA-target pairs analysis is interesting however the results are not extensively discussed. The authors are advised to discuss and further validate at least a few of the observed pairs before claiming it is a useful resource and clinically significant result.

The quality of the proteomics data is not convincing unless the following issues are properly addressed.

1, it is unclear whether and how the proteomics experiments were performed with replicates. The authors are requested to provide the protein and peptide quantification data matrices in the supplementary tables, and to demonstrate the quality metrics of the protein matrix, including FDR estimates.

2, the authors are requested to provide more details in Supplementary table 4 to show what samples are common between data sets.

3, Fig5c: the authors should clearly state which representative samples were selected for WB validation. The quality of WB results is not optimal and convincing.

Reviewer #2 (Remarks to the Author):

The manuscript describes mass spectrometry-based global analysis of proteome associated with BPH, localized PCa and locally recurrent lymph node metastasis. Authors have used the SWATH technique to profile about 4000 and odd proteins. In addition, for most of the samples (except BPH) data on matched analysis of genomic alterations, transcriptome, methylome and miRNA that were previously carried out have been used for analysis. The authors compare data from the different omics compartments and describe the degree of correlation (positive vs negative vs none). The concluding part of the manuscript deals with aconitase 2 and malate dehydrogenase that authors select for validation studies (immunoblot and tissue staining) and demonstrate that ACO2 is elevated in PCa vs benign but downregulated in lymph node met vs PCa, while Malate dehydrogenase is progressively elevated from BPH to PCa to lymph node met.

The strength of this study is the global analysis of 4000 proteins in prostate cancer associated tissues and availability of matched OMICs data to cross compare and possibly integrate

A number of weakness exist. First and foremost is their definition of CRPC which in clinical terms is used to designate patients who exhibit biochemical recurrence and are treated with second generation antiandrogens and become refractory to these regimens. Samples used here are non-regional lymph node samples. Lymph node involvement at the time of primary resection is an indicator of poor prognosis but the specimens are not CRPC. This is a major caveat.

Secondly, the concept of data integration is limited to comparison of relative expression levels of analytes in the various compartments. It is well known that the degree of correlation between transcript and protein for example is low and part of this is due to existence of PTM's. This issue is completely ignored in the study. The entire manuscript deals with numerous comparisons with lack

of rationale as to why each of these are performed and no mention about the take home message. It is not clear what is the rationale looking at common set of analytes between BPH and PCa vs PCa and lymph node met. BPH is not a precursor for PCa. The authors should have used benign adjacent tissue as a control to look at PCa. Also it is unclear what is the contribution of different cell types to the various results. Example in the comparison between BPH and PCa, GTPase is regulated solely at the proteome level but the same signaling process is regulated by transcriptome when comparing CRPC vs PCa. Why is this the case? Is this due to heterogeneity in cell population? Also when making correlations, authors donot consider the fact that mass spectrometry-based proteomics is inherently affected by missingness of the data in which case many of the proteins corresponding to transcripts seen in the mRNA dataset for example may be missing. How is this accounted for? None of the proteomics findings have been orthogonally validated except ACO2 and MDH2. Here again it is unclear what is the role of increased ACO2 in PCa. It is well known that zinc transporters regulate ACO2 function and that Zn is an inhibitor of ACO2 activity. What does increase in ACO2 mean? What does progressive change in MDH2 mean? None of these have been addressed.

In summary, there is no evidence presented to state that proteomics data in this case has uncovered driver events in prostate cancer progression as stated by the authors in the abstract.

Response to reviewers:

We thank the reviewers for the insightful comments, and have revised the manuscript accordingly. We have marked additions and changes in the manuscript text in blue colour.

Reviewer #1 (Remarks to the Author):

This manuscript describes quantitative proteomic analysis of 37 prostate tissue samples from three groups of patients using SWATH mass spectrometry and reports expression of 3394 proteins in all samples. The authors then compare the proteomic data with published genomic and transcriptomic data from some of the samples, and identified correlation between miRNA and some proteins. By comparing proteomics data among patient groups, the authors proposed the TCA cycle as a potentially critical process during prostate cancer progression. Two TCA proteins (ACO2 and MDH2) are validated using WB, and further validated using larger sample sets using IHC.

The miRNA-target pairs analysis is interesting however the results are not extensively discussed. The authors are advised to discuss and further validate at least a few of the observed pairs before claiming it is a useful resource and clinically significant result.

We thank the reviewer for the suggestion and have now performed additional analysis to validate miRNA-target pairs. The data is presented in two new supplementary figures. In supplementary figure 11, we show targets for two representative miRNAs that are differentially expressed to opposite directions in CRPC compared to PC. We show verification for both targets that are decreased by miRNA overexpression at the mRNA level, as well as targets whose mRNA levels remain unaltered, but protein levels decrease as expected according to the *in silico* analysis. Furthermore, in Supplementary figure 18 we verify two predicted, differentially expressed miRNAs to target MDH2 with no alteration at the mRNA level, but detectable effect at the protein level and inhibitory ability of the specific miRNAs in MDH2-3-UTR-Luciferase assay. These results confirm direct targeting of the MDH2 by the two miRNAs.

We have also added a paragraph of discussion on the miRNA-target regulation, and discussion on the MDH2 targeting miRNAs to the manuscript text.

The quality of the proteomics data is not convincing unless the following issues are properly addressed.

1, it is unclear whether and how the proteomics experiments were performed with replicates.

Each sample in the proteomics has been analyzed in two replicate MS analyses. We have added this information to the manuscript text in materials and methods section.

The authors are requested to provide the protein and peptide quantification data matrices in the supplementary tables, and to demonstrated the quality metrics of the protein matrix, including FDR estimates.

The quality of the replicate MS analyses was analysed by calculating the intraclass correlation (ICC) and Spearman's rank correlation was used to generate p-values in permutation tests (n = 1000 permutations / replicate MS analyses). The SWATH-MS data was reproducible with mean intraclass correlation (ICC) coefficient of 0.98 between technical replicate MS analyses. Permutation tests (Spearman correlation) showed that 98.6% of the replicate MS analyses had a p-value < 0.05, which suggests that the replicate MS analyses were of excellent quality. We have added these information to the manuscript text.

We have piled the complete list of all proteins and peptides and their respective results, including the FDR quality metrics, as a Supplementary file (see example image below). Due to the size of the file (150MB), the full file is not included in the revised manuscript submission through the electronic submission system, but in case of article acceptance, will be supplied to the journal.

Protein	Peptide	Precursor	Precursor RT	Sample00	Sample00	Sample00	Sample00	Sample00	Sample01	Sample02	Sample00	Sample00	Sample00	Sample00	Sample00	Sample00	Sample00	Sample00
1	sp Q0966 LKGVDV	615,6485	3	55,01485	69919,23	65049,04	52222,7	88755,71	15210,48	17949,03	4751,362	134511,7	135564,2	66245,53	68316,55	152527,3	74086,12	1
2	sp Q0966 DDGVFVQ	917,4343	2	55,22634	23206,27	28489,28	17222,71	1768,318	9715,678	3410,578	5744,424	7165,942	10109,57	849,5833	12179,61	44342,95	81281,37	1
3	sp Q0966 ISMPDIDL	770,9238	2	67,5632	13957,48	31134,65	21995,57	17211,16	6815,47	6460,063	4268,843	25337,66	24687,36	24108,52	4944,204	37266,43	19616,87	3
4	sp Q0966 ISMPDIDL	770,9238	2	67,5632	13957,48	31134,65	21995,57	17211,16	6815,47	6460,063	4268,843	25337,66	24687,36	24108,52	4944,204	37266,43	19616,87	3
5	sp Q0966 ISMPDIDL	770,9238	2	67,5632	13957,48	31134,65	21995,57	17211,16	6815,47	6460,063	4268,843	25337,66	24687,36	24108,52	4944,204	37266,43	19616,87	3
6	sp Q0966 GPGVDVN	449,7533	2	35,43503	39172,76	69759,97	36963,57	23260,32	5347,448	4171,816	18570,61	80678,69	86179,55	4739,595	44050,39	20287,66	79929,25	€
7	sp Q0966 VEGDMQ	729,8609	2	56,83814	93635,92	116603,5	86653,16	87757,92	76715,91	40164,47	41325,93	151765,2	157042,9	59110,25	59256,64	136816,2	106293,6	1
8	sp Q0966 VDINAPD	606,8146	2	44,47918	192112,7	231368,6	174353,7	188216,6	153687,3	3651,167	79156,55	245063,9	236670	148719,9	138072,4	279069	249141,5	2
9	sp Q0966 VKGVDV	386,2239	3	35,6103	397048,1	409484,3	270925,9	410950,3	260566,3	148825,8	96279,6	684031,4	678450,9	317501,1	340881,8	711590,9	618107,1	4
10	sp Q0966 GEGPDVD	620,3144	2	44,6782	201539,3	32143,69	178072,7	180407,6	7235,717	115996,6	107761,1	269656,5	284144,4	116460,2	42844,52	277038,2	249635,2	3
11	sp Q0966 ISMPDVI	423,2287	3	56,28392	320729,7	330243,8	231247,9	294717,4	229850,8	93752,03	83734,45	611885,8	606329,2	234603,7	230035,4	523686,7	378192,9	€
12	sp Q0966 AEGPEVD	634,3301	2	45,45232	490322	554771,2	400141,9	386344,4	318559,4	157139,5	123310,9	590197,9	583345,5	222088,8	321494,2	650948,2	497054,3	7
13	sp Q0966 VDIDAPD	507,751	4	55,43092	185139,9	212730	124831,6	159599	114895,6	43510,49	5129,517	262697,2	264031,3	154503,6	144014,9	250228,1	205742,1	3
14	sp Q0966 VDIN[Dea	507,751	4	55,43092	185139,9	212730	124831,6	159599	114895,6	43510,49	5129,517	262697,2	264031,3	154503,6	144014,9	250228,1	205742,1	3
15	sp Q0966 VDIDVPD	683,0092	3	66,0062	183705,6	169731,6	113372,5	165275,3	117720,8	10689,63	19280,86	243289,1	237858,9	109210,6	120868,1	232707,1	179710,2	2
16	sp P0276 FQNALLV	480,7849	2	44,64179	13144878	14432455	7622703	19929097	10945839	2057573	1534248	21782872	21650064	20511718	2174949	19202942	16847443	1
17	sp P0276 LVNEVTEF	575,3112	2	45,87921	15001283	16129888	10464395	489934,6	16125685	2921171	2082827	26422827	24595095	26548932	23468318	31044492	30437184	3
18	sp P0276 AVMDDFA	671,821	2	69,77415	7998307	8477084	4956174	9429152	6300411	2068416	1698980	15124318	13671047	13623561	13671984	14589812	13715602	2
19	sp P0276 VFDEFKPL	682,37	3	68,76536	18061594	19531490	8066376	23684418	4169283	63893,53	38996,52	19001001	19424443	20852164	20918396	22165891	11882196	1
20	sp P0276 QTALVELV	500,8055	2	48,61327	9806842	10206123	5873723	12622413	5570071	1717608	1346347	11909324	11430501	11633891	11119277	10588677	9917595	3
21	sp P0276 LVAASQA	507,3031	2	57,74651	6073131	6305711	3797495	10329898	5770232	1803360	1678787	14985129	13733471	13213697	11568638	14964399	14059905	1
22	sp P0276 SHC[CAM]	992,1197	3	75,35872	8045089	9000569	6400960	12989868	7728219	2123110	1507245	32160,59	31088,69	17033587	15288816	8107729	6497454	1
23	sp P0276 YIC[CAM]	722,3246	2	30,11189	5208631	6318787	5315526	9289171	7586325	3236914	3058151	9349573	9240298	8636696	8940602	10945738	9884832	2
24	sp P0276 YLYEIAR	464,2504	2	43,44749	792684,2	864637,4	542842,3	9954351	6204976	148422	33828,97	14209644	13963352	10345210	10535579	10139056	9079573	3
25	sp P0276 KVPQVSTI	547,3174	3	47,72121	3787264	21708849	13703634	36377713	24556407	34659,6	28592,73	31973637	37685607	12410818	37567845	135938,7	114883,2	4
26	sp P0276 LVRPEVDI	663,3215	4	65,77509	9321156	9487099	5579498	13313972	7503907	2679,428	3394,102	8205705	7233952	12011310	12752198	11335711	6400747	1
27	sp P0276 QNC[CAM]	829,3799	2	61,59942	6342943	3028207	83693,91	145638,8	119124,2	2457389	1078925	342763,1	167484	8697555	8150934	216842,2	8744517	2
28	sp P0276 ETYGEMAI	717,7703	2	32,92832	6840443	8449220	6215167	7960947	6226352	2984926	2315521	6603930	6606432	6612519	6917440	8627792	7546010	1
29	sp P0276 FKDLGEEEN	409,5399	3	37,77772	1025571	1109572	786146	1633370	1137060	379579,3	227862	3001761	2874619	3001812	2636160	3197466	2895669	3
30	sp P0276 [PGQ]-QN	820,8667	2	71,18302	3176189	3320543	2447534	2390929	1830706	1169800	1200247	3349259	3259048	3383064	3445601	4385089	3269732	2
31	sp P2133 SPFSAVAVS	767,4116	2	62,90857	1119717	1316578	930349,1	2038160	1550787	79673,41	78036,87	978919,9	914141,2	768970,9	834924,8	1533068	1210337	2
32	sp P2133 AFGPGQLQ	715,3628	2	37,44194	701574,3	874381,9	653119,5	1545243	1133709	51222,41	48327,88	629914,4	624296	554777,3	532938,8	943683,7	819580,5	7
33	sp P2133 IANLQTDL	708,3781	2	55,65624	939828,7	1097828	815989,5	1844484	1470649	73595,24	75586,5	823646,1	774512,6	607840,5	598153,3	1479580	1256274	2
34	sp P2133 VTTYPMIA	870,9657	2	68,53505	230410,8	267166,2	191238,8	376188,2	281764,4	2255,618	175,1699	260362,9	234510,8	254117,6	236530,4	238228,8	201282,2	2

Raw data has also been transmitted to Peptide Atlas. The data can currently be viewed with a password that is provided here for review purposes. The data will be made publicly available at the publishing date. The data can be viewed here:

<http://www.peptideatlas.org/PASS/PASS01126> (password:PCnature)

2, the authors are requested to provide more details in Supplementary table 4 to show what samples are common between data sets.

We have included the sample-wise overlaps between the different analyses as a Supplementary Table 4b.

3, Fig5c: the authors should clearly state which representative samples were selected for WB validation. The quality of WB results is not optimal and convincing.

The sample numbers have now been added to Figure 5c. To provide further supportive evidence of protein level changes, we present the chromatograms of the mass spectrometry analysis for ACO2 and MDH2 in Supplementary figures 15 and 16, exhibiting the clear difference in expression of these proteins between the sample groups.

In order to obtain more convincing evidence in terms of protein level quantification, the additional experiments in the revised manuscript (concerning the miRNA targets) have been performed with targeted MS/MS analysis (targeted mass spectrometry method for MicroLC-MSTrap using stable isotope labeled peptide standards for absolute quantification).

Reviewer #2 (Remarks to the Author):

The manuscript describes mass spectrometry-based global analysis of proteome associated with BPH, localized PCa and locally recurrent lymph node metastasis. Authors have used the SWATH technique to profile about 4000 and odd proteins. In addition, for most of the samples (except BPH) data on matched analysis of genomic alterations, transcriptome, methylome and miRNA that were previously carried out have been used for analysis. The authors compare data from the different omics compartments and describe the degree of correlation (positive vs negative vs none). The concluding part of the manuscript deals with aconitase 2 and malate dehydrogenase that authors select for validation studies (immunoblot and tissue staining) and demonstrate that ACO2 is elevated in PCa vs benign but downregulated in lymph node met vs PCa, while Malate dehydrogenase is progressively elevated from BPH to PCa to lymph node met.

The strength of this study is the global analysis of 4000 proteins in prostate cancer associated tissues and availability of matched OMICs data to cross compare and possibly integrate

A number of weakness exist. First and foremost is their definition of CRPC which in clinical terms is used to designate patients who exhibit biochemical recurrence and are treated with second generation antiandrogens and become refractory to these regimens. Samples used here are non-

regional lymph node samples. Lymph node involvement at the time of primary resection is an indicator of poor prognosis but the specimens are not CRPC. This is a major caveat.

The samples used in this study were freshly frozen tissue specimens from BPH obtained from cystoprostatectomies or transurethral resections (TURP), untreated PC obtained from prostatectomies, and CRPC obtained from TURPs. The CRPC samples came from patients that had been treated either by castration and/or antiandrogens and experienced urethral obstruction (ie. local recurrence) during the treatment. No lymph node samples were used in this study.

Secondly, the concept of data integration is limited to comparison of relative expression levels of analytes in the various compartments. It is well known that the degree of correlation between transcript and protein for example is low and part of this is due to existence of PTM's. This issue is completely ignored in the study.

We agree with the reviewer that the PTMs are a biologically very relevant and interesting issue. Regulation by PTMs may explain the discrepancy in transcript and protein abundance for several proteins as the main mechanism, or one of several mechanisms. To address PTMs on a large scale is a highly important aspect of protein regulation, and has been addressed in prostate cancer samples recently by Drake et al. (2013,2016). In this study, we focused on protein abundance and the correlations of the proteomic data to genetic, epigenetic and transcriptomic data. In the future, it would be extremely interesting to obtain quantitative, large-scale data on PTMs from our sample set. We thank the reviewer for raising this important point, and have included discussion of this topic to the manuscript.

The entire manuscript deals with numerous comparisons with lack of rationale as to why each of these are performed and no mention about the take home message. It is not clear what is the rationale looking at common set of analytes between BPH and PCa vs PCa and lymph node met. BPH is not a precursor for PCa. The authors should have used benign adjacent tissue as a control to look at PCa.

We agree with the reviewer that BPH is not a precursor of PCa. As fully normal prostate tissue is difficult to obtain especially from elderly men, the benign tissues are often used instead as the baseline reference for PC samples. We agree with the reviewer that benign adjacent tissue would be a good option to represent the baseline of prostate tissue. As benign adjacent tissue was not available from these samples, we used the other commonly used and widely accepted reference of BPH.

As for the lymph node mets, no such samples were used in this study. The comparison between PC and BPH represents a comparison between primary cancer and benign prostate tissue, which is able to reveal at least partly the alterations taking place during initial cancer formation. The comparison between CRPC and PC represents comparison between untreated and recurrent cancer, revealing alterations occurring during development of treatment resistance and cancer progression.

As to the rationale for the different comparisons, we would emphasize the novelty of our data and approach. Large-scale proteomics data has not been previously reported from both primary and advanced

disease in prostate cancer. In addition to novel information of the proteomic output of the previously reported genetic and RNA expression data, we show for the first time how the increased aberration load in CRPC is not similar at the level of the proteins, showing how genetic or RNA expression aberrations cannot be solely used to interpret the molecular state of cancer cells. Although the information can still be complemented by other large-scale analyses, such as metabolomics and PTMs, as kindly pointed out by the reviewer, our results provide a whole new level of knowledge to the molecular events in prostate cancer. In addition, our dataset provides an enormous resource for the cancer research community, as the datasets will be publically available for future studies.

We have clarified sample comparison set-up and motivations in the manuscript text.

Also it is unclear what is the contribution of different cell types to the various results.

This is a relevant question that concerns most studies using clinical tissue material, and what needs to be addressed in more detailed studies of tissue heterogeneity and microenvironment. We have secured that our samples contain over 70% of cancerous or hyperplastic cells, in order to make sure that most of the signals we detect originate from the desired epithelial compartment. This information is provided in the materials section and now also raised in the discussion.

Example in the comparison between BPH and PCa, GTPase is regulated solely at the proteome level but the same signaling process is regulated by transcriptome when comparing CRPC vs PCa. Why is this the case? Is this due to heterogeneity in cell population?

Our analysis points out the levels of expression that contain significant differences between the sample groups. As described above, the hyperplastic and tumor cell compartments comprise the majority of our samples, but it remains a possibility that other cell types may contribute to the results. While the large scale analysis presented here is aiming to identify pathways worth focusing future efforts on, the detailed conclusions on any altered pathways need to be based on verification experiments by supportive methods, such as IHC for detection of cells of signal origin, in similar manner that we have performed here for the TCA cycle. We have added the point of cellular heterogeneity contributing to the results to the discussion.

Also when making correlations, authors donot consider the fact that mass spectrometry-based proteomics is inherently affected by missingness of the data in which case many of the proteins corresponding to transcripts seen in the mRNA dataset for example may be missing. How is this accounted for?

We agree that, as with any large scale datasets, conclusions cannot be drawn on missing data. In the correlative analyses, we have used data on only those genes/proteins that are common between the datasets used in the analysis, and thus contain the datapoints required to perform the correlations. In the materials and methods section, we have specified the selection criteria and numbers of genes/proteins included in each analysis.

None of the proteomics findings have been orthogonally validated except ACO2 and MDH2.

We have performed additional validations and present them as additional data in the revised manuscript.

While single standards are not feasible in mass spectrometry discovery analyses containing >40 000 peptides/sample, we used a global standard sample in each sample analysis queue to make sure that instrument and sample handling is repeatable and performing in correct level.

We have now performed a validation study with targeted mass spectrometry method for MicroLC-MSTrap using stable isotope labeled peptide standards for absolute quantification for 4 different proteins. The method and peptide information are provided in the materials and methods section, and in Supplementary Table 15. The results are presented in Supplementary figures 11 and 18. Discussion of these results has also been added to the manuscript text. This analysis confirmed our findings and provides a good basis for additional research based on the results of this study.

Here again it is unclear what is the role of increased ACO2 in PCa. It is well known that zinc transporters regulate ACO2 function and that Zn is an inhibitor of ACO2 activity. What does increase in ACO2 mean? What does progressive change in MDH2 mean? None of these have been addressed.

In this study, we identify alterations and pathways of interest from large scale data. We validated our discovery proteomics results by assessing the TCA cycle protein changes in more detail, and suggest that they may be functionally relevant for prostate cancer. The functional significance and the mechanisms of the detected and verified alterations need to be addressed in more detail with experimentation falling beyond the scope of this study.

In summary, there is no evidence presented to state that proteomics data in this case has uncovered driver events in prostate cancer progression as stated by the authors in the abstract.

In this work we have identified previously unknown alterations in prostate cancer both at protein expression, miRNA regulation, and pathway level, and provide a valuable proteomics resource for the community for further studies. However, we agree with the reviewer that the term driver event can be debated without further data on individual events, and thus we have removed the term from the abstract.

REVIEWERS' COMMENTS:

Reviewer #1 (Remarks to the Author):

In the revised manuscript the authors have addressed the issues raised in the initial review. The paper is acceptable.

Reviewer #2 (Remarks to the Author):

The authors have addressed most of the concerns of this reviewer. The authors are encouraged to include the description of CRPC samples they have used in the main text.

Latonen et al. - Response to reviewers' final comments

REVIEWERS' COMMENTS:

Reviewer #1 (Remarks to the Author):

In the revised manuscript the authors have addressed the issues raised in the initial review. The paper is acceptable.

Response: N/A

Reviewer #2 (Remarks to the Author):

The authors have addressed most of the concerns of this reviewer. The authors are encouraged to include the description of CRPC samples they have used in the main text.

Response: We have included a better description of the CRPC samples in the first paragraph of the results section.