Author's response to reviews

Title: An integrative multi-platform analysis for biomarker discovery of osteosarcoma

Authors:

Guodong Li (litrue2004@yahoo.com.cn)
Wenjuan Zhang (forrestzhangcn@gmail.com)
Huazong Zeng (zhz@sagc.org.cn)
Lei Chen (chenlein39@hotmail.com)
Wenjing Wang (wwjcde@hotmail.com)
Jilong Liu (Jilong.liu@ymail.com)
Zhiyu Zhang (yfj0352@yahoo.com.cn)
Zhengdong Cai (czd856@vip.sohu.com)

Version: 4 Date: 23 March 2009

Author's response to reviews: see over
Dear Sir/Madam,

Thanks for transferring the reviewers’ comments to us. We appreciate Toni whistler and David O’Gorman very much for their patience and nice advice to help us improve our manuscript.

We have revised our manuscript following the reviewers’ suggestion point-by-point. The major revisions in latest version of manuscript are displayed with annotations. This cover letter also contains the point-by-point description of revised content and answers to reviewers.

Again, Dr Yili Yang from Cancer and Developmental Biology Laboratory, National Cancer Institute at Frederick, National Institutes of Health and Dr Rob Waterland from Baylor College of Medicine helped us for English usage of the manuscript.

In addition, with the cover letter, we also upload the revised version of the manuscript, figures and supplement materials.

The following is the point-by-point reply:

A. To editor,
   1. We ask that you copyedit the manuscript for English language usage. We recommend that you ask a native English speaking colleague to help with this, or if that is not possible, you may need to use a professional copyediting service.
      Re: Dr Yili Yang from Cancer and Developmental Biology Laboratory, National Cancer Institute at Frederick, National Institutes of Health and Dr Rob Waterland from Baylor College of Medicine have helped us for improving the readability of the manuscript.

   2. Ethics - Experimental research that is reported in the manuscript must have been performed with the approval of an appropriate ethics committee.
      Re: We have added this item at the end of the first paragraph of “Specimens” in “Methods”. The Approval file number is 2009LL002 by Second Military Medical University Ethics Committee.

   3. Competing interests - Please include a 'Competing interests' section between the Conclusions and Authors’ contributions. If there are none to declare, please write 'The authors declare that they have no competing interests'.
      Re: We have added this item in manuscript, before “Authors' contributions”.

B. To reviewer_ Toni whistler,
   Major compulsory revisions:
   1. The first sentence in the SELDI-TOF-MS data reads “The mass spectrometry
amples were from serum while the microarray samples were from cell lines.” I personally found this confusing. I feel it would be helpful if the specimen description were put into a separate paragraph under the heading “Specimens”. This could then include more information on serum processing. The authors state that serum samples were collected into “glass tubes” – what tubes with what additives? How was the serum processed? Was it aliquoted and frozen immediately upon storage? What temperature were they stored at? Were there any freeze/thaw cycles on the sera prior to their use in this protocol? How long had the sera been in storage? All of this is important information for the reader as these are important factors in interpreting the SELDI data.

Re: Specimen description has been put into a separate paragraph, see the first paragraph of “Specimens”. The preoperative serum samples of 27 patients diagnosed as osteosarcoma without chemotherapy were collected for SELDI-TOF-MS analysis. After operation, the serum samples of these patients were obtained again for comparative western-blot experiment. As for each patient, 10ml peripheral vein blood from cubital vein was collected in axenic, dry, and additives-free tubes. The samples were centrifuged at 4,000 rcf for 10 minutes after being placed at 4°C for 4 hours and serum being separated out. Then, serum samples were stored in -80°C refrigeratory. When the experiment was implemented, the serum samples were thawed on ice, and centrifuged at 4°C, at 12,000rcf for 10 minutes to wipe off the indiscernible materials. The same method was also applied to collect serum samples of the 47 normal controls.

2. The healthy controls – how were they selected? What was the population from which they were drawn? No where in the paper did I see an explanation as to why cell lines were used for the gene expression profiling. I assume that peripheral blood mononuclear cells or tissue samples were not available from these same patients? Using cell lines for biomarker detection has many drawbacks. The authors have tried to mitigate this by using 3 different osteosarcoma cell lines. I feel these needs to be addressed in the paper.

Re: All the healthy controls were the volunteers from a routine-health examination and selected matched for age and sex. All participants of this study signed the informed consents. Studies were performed according to the rules of the Medical Ethics Committee and approved by the local institutional review boards of participating institutions.

The reason why the cell lines were chosen rather than peripheral blood mononuclear cells and tissue samples was based on the intrinsic characters of osteosarcoma which are different from other cancers such as liver cancer and prostate cancer:

1. osteosarcoma originates from stromal osteoblast in human bone marrow. It is difficult to obtain the control corresponding to the osteosarcoma tissue.
2. Unlike the uniformity of cancer tissues such as liver and prostate cancers, the components of osteosarcoma tissue are obviously influenced by sample location.
The advantage of using standard cell lines is not only to get excellent control, but also reduce the bias by sampling. Using the coincident expression difference results among three ATCC different osteosarcoma cell lines will help to reduce the influence of individual variation.

The items above have been added into manuscript in the first paragraph of “Specimens” and the fourth paragraph of “Discussion”.

3. Sample processing for the SELDI samples. Not nearly enough information is provided. There is a minimum amount that would be required for a reader to be able to repeat this experiment. For example I assume that the serum was in no way fractionated prior to application to the ProteinChip surface? What methodology was used for application – were the manufacturer’s instructions followed exactly? Were sera run in duplicate? Was the processing performed using automation (i.e. the Biomek platform) or was it done manually? Were the samples randomly placed on the ProteinChips? How much EAM was applied? Which EAM was applied? What version of the mass spectrometer was used – an older PBSIIc or a newer model? 

Re: We have supplied more details for sample information. Please see the answer of question 1.

Yes, the serum was in no way fractionated prior to application to the ProteinChip surface. We implement all the analysis procedure followed the instructions strictly. Sera were run in duplicate for twice. The processing performed manually. 0.5 ul saturated solution of SPA in 50% CAN and 0.15% TFA was applied. The version of the mass spectrometer was Ciphergen ProteinChip Reader PBSIIc.

The items above have been added into manuscript in the first paragraph of “SELDI-TOF-MS analysis” in “Methods”.

4. The order of information supplied needs to be checked. For example, in the second paragraph, the second sentence gives information on exporting the collected spectra with baseline subtracted. The very next sentence then provides information on the acquisition of the spectral data. This happens before spectra can be exported.

Re: We have revised the content of this part. See the first paragraph in “SELDI-TOF-MS data”.

5. Were all the sera run in a single batch? Were any quality control samples (say of pooled normal sera) run on each chip to determine the variability of the system and to monitor performance?

Re: Yes. Sera were run in one batch and we used pooled normal sera as control
sample. We have revised the content of this item. See the first paragraph in “SELDI-TOF-MS data”.

6. Paragraph 3 starts to address the post-processing of the spectral data. What QC was done to insure data was of a good enough quality to be included in the study? Were there replicate samples? How were these dealt with? Peak detection is mentioned, but none of the parameters used are given. Were spectra normalized? Were the spectra calibrated and aligned?

Re: Both samples collection and treatment procedures are standard. Chips were from one single batch. And the instrument was normalized. For each sample, there were two points. Each point was collected for 130 times. Spectra were normalized, calibrated and aligned.

We have added content of this item. See the first paragraph in “SELDI-TOF-MS data”.

7. It is not clear what MZ range was used for the analysis. This is important as there is a marked EAM signature at the beginning of the spectra. The optimization range is 2000-10000, with a max at 50000. The next sentence states that the MZ range from 0 to 20000. No information is given as to the number of peaks detected.

Re: In this study, the m/z values range from 0 to 20,000. 96 peaks were detected per sample. The total number of peaks detected was 7104 for 27 samples from patients and 47 ones from healthy controls.

We have revised content of this item. See the first paragraph in “SELDI-TOF-MS data”.

8. The description of the statistical testing performed on the protein peak data set is a little confusing. “The statistical significance of the protein differences between the sample sets and control sets was calculated by ANOVA. For each peak, the difference of means between two datasets was tested using wilcoxon rank sum test. Wilcoxon rank sum test was an independent, non-parametric test for assessing whether two samples of observations come from the same distribution.” By “protein differences” are you referring to differences in the area under the MZ peak? Or height of peak? Why an ANOVA and a Wilcoxon Rank sum test? The Kruskal-Wallis test is a non-parametric analogue of the one-way analysis of variance (ANOVA). The “difference of the means” the mean of what? What package was used for the statistical analysis? Was this part of the Biomarker Wizard?

Re: We have corrected the description in the statistical method section. ANOVA was not used in our data analysis process any more. We tried non-parametric test (Wilcoxon rank sum test) at first. Then, we found that classical t-test produced very
similar results. In the end, we chose t-test. In the t-test, mean values (height of peaks) of OS and control samples were compared. To minimize false discovery rate, Benjamini and Hochberg multiple test correction was used.

We have revised content of this item. See the second paragraph in “SELDI-TOF-MS data”.

9. Moving the description of the cell lines to the description of the samples section would be helpful. I would like to see information included on passage level of the cell lines and also how cell lines were maintained – medium, split ratios, growth conditions. At what point would they be harvested? Confluence of monolayer?

Re: Three OS cell lines (MG-63, Saos-2 and U-2 OS) and one osteoblastic cell line (hFOB1.19) were collected for gene microarray analysis. MG-63 was kindly provided by Dr Agi Grigoriadis from University College London. Saos-2, U-2 OS and hFOB1.19 were purchased from ATCC. All the cell lines were cultured under the corresponding guidelines of ATCC standard conditions as following:

hFOB 1.19: ATCC complete growth medium: The base medium for this cell line is a 1:1 mixture of Ham's F12 Medium Dulbecco's Modified Eagle's Medium, with 2.5 mM L-glutamine (without phenol red). To make the complete growth medium, add the following components to the base medium: 0.3 mg/ml G418; fetal bovine serum to a final concentration of 10%. Temperature: 34.0°C

MG-63: ATCC complete growth medium: The base medium for this cell line is ATCC-formulated Eagle's Minimum Essential Medium, Catalog No. 30-2003. To make the complete growth medium, add the following components to the base medium: heat-inactivated fetal bovine serum to a final concentration of 10%. Atmosphere: air, 95%; carbon dioxide (CO2), 5%. Temperature: 37.0°C

U-2 OS: ATCC complete growth medium: The base medium for this cell line is ATCC-formulated McCoy’s 5a Medium Modified, Catalog No. 30-2007. To make the complete growth medium, add the following components to the base medium: fetal bovine serum to a final concentration of 10%. Atmosphere: air, 95%; carbon dioxide (CO2), 5%. Temperature: 37.0°C

Saos-2: ATCC complete growth medium: The base medium for this cell line is ATCC-formulated McCoy’s 5a Medium Modified, Catalog No. 30-2007. To make the complete growth medium, add the following components to the base medium: fetal bovine serum to a final concentration of 15%. Atmosphere: air, 95%; carbon dioxide (CO2), 5%. Temperature: 37.0°C

In order to satisfy the requirement of Affymetrix gene microarray hybridization (i.e. at least 1µg total RNA can be extracted), all the cell lines were cultured at least for three generations and the total cell amount >10^7. All cells used were stayed in exponential phase of growth and harvested when confluence of monolayer being observed. Split ratios were all 1:4.

The items above have been added into manuscript in the last two paragraphs of “Specimens”.
10. Very little detail is provided as to the sample processing for the gene expression arrays. What amount of total RNA was used for labeling? The authors’ talk of assessing quality and quantity of total RNA extracted using the Agilent Bioanalyzer system – what were the quality criteria for inclusion? Did any samples not meet these criteria? What kit from what company was used for labeling the RNA for the 3’expression arrays? There are many different ones out there, so this needs to be specified.

Re: 5µg total RNA was used for labeling. The criteria for Agilent bioanalyzer QC is RIN >=7 and 28S/18S >=0.7. Labeling kit is Affymetrix Genechip expression 3’ amplification one-cycle target labeling and control reagent.

The items above have been added into manuscript in the first paragraphs of “Gene microarray data” in “Methods”.

11. Details of the array format – PM and MM probe sets appears in the middle of the description of how the arrays were processed. Is this really important to put in here? Discussing the criteria for what is a positive signal – before the fact that the arrays were scanned is out of order and confusing.

Re: Done. See “Gene microarray data” in “Methods”.

12. Clarification needs to be made as to data analysis and data pre-processing. MAS5 is the algorithm used for pre-processing the array data, and a “pre-standard procedure” is not helpful. Was MAS5 implemented using Affy Power Tools? What was done to look for array outliers, to check for data quality? Was all data processed together, or each cell line compared to the osteblastic cell line separately? The fact that a t-test was used would imply this.

Re: We have modified the description here to make it clear. Please see “Gene microarray data” in “Methods”. MAS5 was implemented using Affy Power Tools. Each cell line compared to the osteblastic cell line separately.

13. What software was used to for the t-Test and to calculate the 2-fold change in signal between cell lines?

Re: MATLAB 7.5 was used to do the calculation. Please see the second paragraph of “Gene microarray data” in “Methods”.

14. It is interesting that in both the proteomic and genomic data no attempt is made to correct for multiple testing. I would think this would add another layer of stringency and add confidence to the data you are generating and increase chances of validation.

Re: Yes, the reviewer is right. We have realized the problem of multiple testing. The p values of t-test have been re-calculated by the Benjamini and Hochberg multiple test correction to control false discovery rate (adjusted p values).
15. No mention is made of making the microarray data to the public by deposition in a public repository.

**Re:** Our microarray data have been submitted to NCBI GEO database (http://www.ncbi.nlm.nih.gov/geo/) under the GEO number GSE14789 (including GSM369298, GSM369299, GSM369300 and GSM369301, a total of 4 samples). We have supplied this description at the end of text.

16. What positive and negative controls were run with the blot? Figure 4 shows the blots with a single band identified as CYC1, what is the doublet above this? If a negative isotype control was run it would answer the question whether this was non-specific binding. Does CYC1 run as a singlet? The positive control would help with this question. It would also be helpful to have molecular weight markers on the photograph. To be convinced it would also be helpful to have the Ponceau Red stained photograph alongside, showing equal loading of protein.

**Re:** The doublet above CYC-1 band was identified as non-specific band by test with irrespective antibody like antibody of BCL-2. According references, CYC-1 band should be one single band in western-blot experiment. Molecular weights were marked at the corresponding positions on the modified pictures of western-blot results to distinguish the specific and non-specific bands.

In order to keep the equal loading of protein, first, BCA Protein Assay Kit from Pierce is used to quantify protein concentration in serum samples. Second, after SDS-PAGE electrophoresis using equal amount of samples, the expression levels of samples were normalized according to the gray analysis of IgH that molecular weight was around 50KD.

The items above have been revised into manuscript in “Immuno-blot analysis” in “Methods” and “Experimental validation” in “Results”.

17. I think this data would be more convincing if it was not just a purely visual interpretation of which samples have more and which less. It would be possible to do relative quantitation, using a ratio of area under the curves for the protein bands. A simple statistical test could then be applied to show differences.

**Re:** The gray values of expression levels for the specific bands on western-blot pictures have been calculated. The normalization of gray values for specific bands using IgH as criterion was shown in histogram.

The items above have been revised into manuscript in “Immuno-blot analysis” in “Methods” and “Experimental validation” in “Results”.

18. It is stated that there were 96 differently expressed peaks identified between the 1500 and 20000 MZ values (are these really Da?) at a cutoff of 1e-3, (I assume that
is the p-value). How many peaks were identified in total in this range, if these are the differentially expressed? Was any form of multiple testing applied to the data – this would add a level of confidence and really help in selection of those peaks that are truly different. The next sentence then states “Differences of 9 peaks owned the statistical significances with p<0.001.” Please clarify.

Re: We have corrected and polished the text. We also adjusted p values of t-test by the Benjamini and Hochberg multiple test correction method. As a result, 6 out 96 peaks range from 2000 to 20000 MZ values were identified as differentially expressed peaks (adjusted p value<0.05).

We have revised content of this item. See the second paragraph in “SELDI-TOF-MS data”.

19. Fig 1b (I believe is labeled Fig2 in the PDF). No information is given as to the methodology/software used for the clustering. Was this hierarchical clustering? What was the distance metric used (Euclidean, Pearson etc.) was average linkage used? This should be put into the methodology. This information helps in the interpretation of the heat map. Also perhaps centering the data on zero may help in showing the differences in expression levels of the peptides. The heat map as is, is not all that convincing.

Re: We have corrected the descriptions in both text and figure legend. Hierarchical clustering analysis was implemented for significantly changed peaks. Hierarchical clustering was conducted with MATLAB 7.5(Mathworks, inc. USA) using default parameters (Euclidean distance metric and average linkage method). Raw values were log2-transformed, centered relative to the median. Relative changes in their expression level were indicated by a color code. Red indicates that the level of gene expression is higher than median, and green indicates that the level is lower than median.

The items above have been revised into manuscript in the second paragraphs of “Gene microarray data” in “Methods”.

20. It would be helpful if the number of genes up and down-regulated in each of the 3 cell lines was given before the reader is asked to look at Fig 2. Move the sentence “The cluster analysis and GO annotation of differently expressed genes were presented in Fig.2” to the end of the next sentence.

Re: Done. See ‘Gene markers from microarray’.

21. It would be more useful to have the mean intensities and the fold change given in an excel file rather than the heat map for the 655 differentially expressed genes in the supplemental figure. I was not able to search the figure and find the genes mentioned in the results – and doing this manually took a lot of time. No
information is given as to the number of replicates done for each cell line. **Re:** 653 differentially expressed genes from three osteosarcoma cell lines and one osteoblastic cell line were screened by microarray analysis. Two control genes have been removed from previous figure and gene list. Limited to time and expensive cost, no replicates was performed. The gene list comparable to heatmap can be found in supplementary file gene_list.xls.

22. I do not understand the “confirmation” of the candidate genes by matching to the GO database. Please clarify.  
**Re:** There is an improper expression here. We have improved the sentence. See ‘Gene markers from microarray’.

23. There were 18 potential biomarkers identified in the Link-test analysis (Table 2). None of these relate to those described in the results of the microarray data. It would be helpful to include the fold change values in this table for both the proteomic and genomic data.  
**Re:** You are right. We have redone Table 2. The fold change values for both the proteomic and genomic data are included in this new version of Table 2. After running multiple testing, the potential biomarkers identified in link-test were corrected to 13.

24. Experimental validation. No explanation is given as to why CYC1 was chosen for validation? It did not have the highest p-value in the Link-test, it was has been associated with other cancers in the literature, which would make it a less-specific candidate biomarker. There were also 5 other candidate protein fragments with the same MZ (Table 2).  
**Re:** The most important purpose of experiment validation is to verify the veracity and reliability of our approach that considering results from both microarray and SELDI-TOF rather than disease specialty. CYC1 has been reported associated with other cancers. And it is also existed in our candidate gene list. In microarray analysis, the average difference of expression level of cyc-1 between OS cell lines and control was largest. Thus, we can expect that it will be a positive result in experiment. As to our purpose of experiment, CYC1 is chosen as a creditable biomarker.  
In addition, the MZ absolute value reflects the molecular weight of peptide. The peak with larger MZ value are more creditable because it is more possible correspond to protein. Thus, the candidate biomarkers with larger MZ values were considered for validation experiment. Another reason is that IgH, Igκ and Igλ will affect the identification of other proteins at the same location in western-blot for immunoglobulin enriched in serum samples. Hence, from six candidate biomarkers with same MZ values of 8769.12, the proteins with molecular weight around that of IgH, Igκ and Igλ were excluded for western-blot experiment.  
Moreover, we also examined the expression of PITPNC1 and PDCD2, whose molecular weights and MZ value were similar with CYC-1, and found no
significant difference between OS and control samples.

25. Mention should also be made that of the 9 potential markers in the proteomic data, 2 were not associated with a corresponding peptide (4476.07 and 4820.49 Table 1). Re: According to link-test, we do not discover the consist result between microarray and SELDI data at these two m/z values.

The items above have been revised into manuscript in “Biomarkers exaction from link-test” in “Results”.

Conclusions
26. I think the conclusion should focus a little more on the results presented in this paper, explaining the short-comings and many relevant issues (most of which I have addressed in my previous comments). Re: Done. We have revised the “Discussion” to give more attention on our results and short-comings.

27. Fig 4 and the 3D structure of CYC1 needs to be removed, it brings nothing to the paper. Re: Done.

28. The authors keep mentioning “specific”, “early”, “reliable” with reference to biomarkers. However, they do not address this directly with regard to the possible markers they identified in this study. Re: Yes, we need more and further experiments to directly verify and explore the specific biological function of these candidate genes effect on osteosarcoma. In this article, we mainly present an analysis strategy to discover these genes.

29. This was particularly confusing. The figure legends do not match the figures identified. Fig1 legend annotates an a) and a b). For the figure labeled Fig1. The values given of peak intensities are for both cases and controls combined? No mention is made of what the red plus signs indicate. Were these all the peaks identified in the 2000 to 20000 MZ range, or just the differentially expressed ones? Perhaps the 9 that are discussed in the text could be highlighted? Re: In order to draw an overall profile of peak intensities, cases and controls were plotted in Fig 1. Six differentially expressed peaks (p<0.05) were marked with gray rectangles. In the box plot, the length of the box represents the difference between the 25th and 75th percentiles and the horizontal line inside the box represents the median. Outliers are label with red crosses.

In Fig 2, the 6 peaks were selected from all peaks by t-test. These peaks were considered differentially expressed ones with statistical support.

30. Fig 2. I believe this is Fig 1b, mentioned in the legend and in the manuscript? Re: Yes, Fig 2 is Fig 1b. We are terribly sorry for the fact that there is an
unexpected misplay in notation during uploading the figures.

31. Fig 3. I personally find it helpful if all the information required for interpreting the figure is given in the legend. It helps the reader not to have to go back and forth to the text. For example here it would be helpful to mention that this is the enrichment analysis for the 655 genes common to the 3 cell lines that were differentially expressed. What was the background list used to determine the enrichment? What was the program used to do the enrichment analysis? The abscissa of the bar plot – genes percentage. Percentage of what? This is not clear, and values of 0.2% are giving me no guidance.

**Re:** Gene set enrichment analysis was performed using EASE software. All annotated genes in the human genome were provided as background gene list. The abscissa of the bar plot was the percentage of genes in up-regulated and down-regulated gene list, respectively. Besides, 0.2 was meant to be 20%, a ratio value actually.

32. Table 2. It would be helpful to have the candidate biomarkers ordered by MZ value. In the present table there are 4 candidate protein fragments with an MZ = 13761.7 and they are at the top, middle and bottom of the table. It would also be helpful to include the data from which these were extrapolated – the gene expression and protein fold changes.

**Re:** Yes, the reviewer is right. We have redone Table 2. Please refer to Q23.

33. I see no relevance to the inclusion of the structure of CYC1 to this paper.

**Re:** We have canceled the discussion on structure of CYC1.

34. Legend of Fig 5. It is misleading to state “Finally, the candidate biomarkers were screened and validated by the experiments.” Only ONE biomarker was validated.

**Re:** There is an improper expression here. We have improved the sentence.

Minor essential revisions:

**Background:**

1. “Mass spectrometry and microarray technologies have been commonly used in studying genomics and proteomics, respectively”. The genomics and proteomics need to be swapped – mass spectrometry with proteomics and microarray associated to genomics.

**Re:** Done.

2. “Mass spectrometry is considerably faster, cheaper, and more accessible than microarray.” My experience tells me this is not a true statement on any of the 3 points. There are probably more labs with microarray capacity than there are labs with mass spec capability. I do not believe this is the reason that SELDI has gained attention as a biomarker discovery platform. Either reference this statement or make adjustments.
Re: This saying has been canceled in our statement. See ‘Background’.

3. The abbreviation SAA is used at the beginning of a paragraph and then the in the next sentence the abbreviation is defined. This needs to be swapped.
   Re: Done.

4. In that same paragraph SELDI-MS is used as an abbreviation where SELDI-TOF-MS is given as the formal abbreviation.
   Re: Done. We have unify the abbreviation as ‘SELDI-TOF-MS’.

5. 2-D DIGE is used but nowhere is the abbreviation defined. Not all readers will be familiar with the technology.
   Re: Done.

6. Protein Chip reader parameters: change “sensibility 7” to “sensitivity 7”.
   Re: Done.

7. “The data were settled by the software named Ciphergen ProteinChip 3.1.1.” should read “Data was collected using the Ciphergen ProteinChip software v3.1.1.” “The quality and purity of the products were controlled by Agilent 2100.” Change “controlled” to “assessed”.
   Re: Done.

8. ”The 300ul buffer system contained”. I believe the authors are referring to the hybridization buffer. Such details are not really necessary – this is the standard methodology used for the Affymetrix system – it is probably best just to state that this was performed according to the manufacturer’s instructions (give the kit name) and then reference the user manual.
   Re: Done.

9. Minipore water change to Millipore.
   Re: Done.

Discretionary revisions:
1. Abstract and in background: “has been developed fast as an attractive approach” perhaps use “has rapidly developed”.
   Re: Done.

2. It may be helpful to get this manuscript reviewed by an English-speaking scientific writer. The authors have done a commendable job of writing. However, there are a few places where clarification is needed. A better choice of words may help the reader understand what the authors are trying to convey.
   Re: Dr Yili Yang from Cancer and Developmental Biology Laboratory, National Cancer Institute at Frederick, National Institutes of Health and Dr Rob Waterland
from Baylor College of Medicine has helped us for improving the readability of the manuscript.

C. To reviewer David O’Gorman,

1. The figure legends and the figures themselves do not always correspond, notably figures 4 and 5.
   Re: We are terribly sorry for there is an unexpected misplay in notation during uploading the figures. We have corrected the problem in revised version.

2. The authors used the data to identify CYC-1 as a potential serum marker of osteosarcoma. In figure 3 (text, marked Figure 4, a western immunoblot for CYC-1), the text describes “whole cell lysates from serum samples”. This is confusing, since serum by definition is the liquid component derived from blood after clotting and centrifugation and does not contain whole cells. The authors need to clarify precisely what these samples are. If the SELDI analysis was performed on serum without “cell lysis” then there should be no need for any cell lysis for the confirmatory western immunoblotting.
   Re: We are so sorry for making such confusing. The serum samples were used in western-blot experiment. The legend of figure 3 has been modified.

3. While they claim that equal protein levels were loaded on each gel, no loading controls are shown (such as Ponceau staining). A much more convincing blot would contain serum samples from patients with osteosarcoma and controls on one gel rather than two, as we have no way of ensuring that the ECL exposure of the two gels was identical. Each blot appears to display at least three bands of differing molecular weight yet there is no comment or explanation in the text. All these points should be addressed prior to publication.
   Re: As the answer to question 16 of Toni whistler. The doublet above CYC-1 band was identified as non-specific band by test with irrespective antibody like antibody of BCL-2. According references, CYC-1 band should be one single band in western-blot experiment. Molecular weights were marked at the corresponding positions on the modified pictures of western-blot results to distinguish the specific and non-specific bands.

   In order to keep the equal loading of protein, first, BCA Protein Assay Kit from Pierce is used to quantify protein concentration in serum samples. Second, after SDS-PAGE electrophoresis using equal amount of samples, the expression levels of samples were normalized against the density of IgH, which migrate around 50 kDa. The gray values of expression levels for the specific bands on western-blot pictures have been calculated. The normalization of gray values for specific bands using IgH as criterion was shown in histogram. The result suggested that the expression levels of CYC-1 in tumors are higher than in healthy controls.

The items above have been revised into manuscript in “Immuno-blot analysis” in
4. To identify osteosarcoma biomarkers, the authors chose to perform SELDI-TOF-MS on serum from patients and compared these with serum form normal controls to generate a protein biomarker output using Biomarker Wizard software. The use of serum rather than plasma for the analysis means that a significant sub-set of proteins was removed from the sample through clot formation. Can the authors comment on their choice of using serum instead of plasma for this analysis?

**Re:** Both plasma and serum contain various kinds of proteins. However, there is great difference for protein enrichment between plasma and serum. There is still no verdict of which one being more proper for detection. At present, most biomarkers of cancers are proteins in serum with low enrichments. Furthermore, there are limitations for dynamic range and sensitivity of instrument when discovering biomarkers in complicated samples such as plasma and serum. Some simple sample preparation method, like wipe off the proteins with highest enrichment, can help expanding the dynamic range of current instrument. In our study, in order to avoid the interfering from fibrinogen with high enrichment, we selected serum samples to obtain more sensitive outcomes.

The items above have been added into manuscript. Please see the third paragraph in “Discussion”.

5. The authors used the CM-10 chip, a weak cation exchange surface, for SELDI analysis. This surface would therefore select for a sub-set of appropriately charged proteins in the serum. Why was this chip the only surface used? Why not also utilize a weak anion exchange chip such as Q10 to increase the potential yield of differentially expressed proteins?

**Re:** We have tried other chips. Only CM10 can be obtained most protein signals. However, our results may still skew toward potential markers that are positively charged at the experimental conditions. We will solve this problem in future study.

The items above have been added into manuscript. Please see the third paragraph in “Discussion”.

6. For the microarray analysis, the authors used an osteoblast cell line (hFOB1.19) as a control to compare to three osteosarcoma lines, MG-63, Saos-2 and U-2 OS. No details are supplied regarding culture conditions of these cell lines. For example, were they all grown in the same media type? Was fetal bovine serum used in the culture and if so, at what concentration? These details are necessary to allow others to replicate these findings.

**Re:** The details have been given in text. Three OS cell lines (MG-63, Saos-2 and
U-2 OS) and one osteoblastic cell line (hFOB1.19) were collected for gene microarray analysis. MG-63 was kindly provided by Dr Agi Grigoriadis from University College London. Saos-2, U-2 OS and hFOB1.19 were purchased from ATCC. All the cell lines were cultured under the corresponding guidelines of ATCC standard conditions as following:

HFOB 1.19: ATCC complete growth medium: The base medium for this cell line is a 1:1 mixture of Ham's F12 Medium Dulbecco's Modified Eagle's Medium, with 2.5 mM L-glutamine (without phenol red). To make the complete growth medium, add the following components to the base medium: 0.3 mg/ml G418; fetal bovine serum to a final concentration of 10%. Temperature: 34.0°C.

MG-63: ATCC complete growth medium: The base medium for this cell line is ATCC-formulated Eagle's Minimum Essential Medium, Catalog No. 30-2003. To make the complete growth medium, add the following components to the base medium: heat-inactivated fetal bovine serum to a final concentration of 10%. Atmosphere: air, 95%; carbon dioxide (CO2), 5%. Temperature: 37.0°C.

U-2 OS: ATCC complete growth medium: The base medium for this cell line is ATCC-formulated McCoy's 5a Medium Modified, Catalog No. 30-2007. To make the complete growth medium, add the following components to the base medium: fetal bovine serum to a final concentration of 10%. Temperature: 37.0°C.

Saos-2: ATCC complete growth medium: The base medium for this cell line is ATCC-formulated McCoy's 5a Medium Modified, Catalog No. 30-2007. To make the complete growth medium, add the following components to the base medium: fetal bovine serum to a final concentration of 15%. Atmosphere: air, 95%; carbon dioxide (CO2), 5%. Temperature: 37.0°C.

In order to satisfy the requirement of Affymetrix gene microarray hybridization (i.e. 1µg total RNA can be extracted), all the cell lines were cultured at least for three generations and the total cell amount >10^7. All cells used were stayed in exponential phase of growth and harvested when confluence of monolayer being observed. Split ratios were all 1:4.

The items above have been added into manuscript in the last two paragraphs of “Specimens”.

7. The study incorporates a central assumption that the genes expressed by immortalized cell lines grown in tissue culture medium in plastic trays will reflect the proteins secreted into the serum by an osteosarcoma in vivo. Perhaps the authors can comment on this limitation, would surgically resected osteosarcoma tissue be a better approach? Can the latter be obtained, i.e. would this be a practical improvement?

Re: The protein expression level of tumor cells could be different in vivo and in vitro. We are not sure if the surgically resected osteosarcoma tissue will be better. Maybe it is a good suggestion for future study.

On the other hand, the reason why the cell lines were chosen was based on the
intrinsic characters of osteosarcoma which are different from other cancers such as liver cancer and prostate cancer:

1. osteosarcoma originates from stromal osteoblast in human bone marrow. It is difficult to obtain the control corresponding to the osteosarcoma tissue.

2. Unlike the uniformity of cancer tissues such as liver and prostate cancers, the components of osteosarcoma tissue are obviously influenced by sample location. The advantage of using standard cell lines is not only to get excellent control, but also reduce the bias by sampling.

The items above have been added into manuscript in the fourth paragraph of “Discussion”.

8. The link-test algorithm is reported to utilize the Swiss-Prot database. A drawback of the current approach is that it is likely that many of the differentially expressed genes identified in the microarray analysis may give rise to protein products that are differently modified post-translation in osteosarcoma cell lines relative to the same proteins derived from osteoblast cell lines in culture. The authors do comment on the possibility of incorporating post-translational modification information from the Swiss-Prot database. This reviewer believes that such data would be a great improvement on the current approach. Is this data not currently available? A more detailed explanation of how link-test works in this context would also be helpful.

Re: Yes. For protein modification, over 200 forms have been identified. What is more, multi-site modification and cell-type-specific modification make it even harder to deal with in MS data analysis. We have been considered introducing protein modification into our link-test method. RESID Database (http://www.ebi.ac.uk/RESID/) seems to be a good start point, yet there is still a long way to go. We are planning to include protein modification options in our next release of link-test tools.

The items above have been added into manuscript in the first paragraph of “Discussion”.