Gmail Shicheng Guo <shihcheng.guo@gmail.com>

Final Decision made for SREP-16-22400

1 message

scientificreports@nature.com <scientificreports@nature.com> Mon, Jul 25, 2016 at 6:44 AM
Reply-To: scientificreports@nature.com
To: scguo@ucsd.edu

Dear Dr Guo:

Thank you for your help with manuscript SREP-16-22400, "Identification of Methylation-Driven, Differentially Expressed STXBP6 as a Novel Biomarker in Lung Adenocarcinoma", which you recently reviewed for Scientific Reports.

For your records, the decision for this manuscript, based partly on your input, was Major revision. A full copy of the comments to authors is appended, below.

Your assistance and participation in the review process for Scientific Reports is greatly appreciated.

Best regards,

Manuscript Administration
Scientific Reports
4 Crinan Street
London N1 9XW
E-mail: scientificreports@nature.com

Referee comments to the authors:

Reviewer #1:
Remarks to the Author:
Comments for the Author

In the present study, Dr. Chuang and colleagues identified a novel biomarker STXBP6 for lung adenocarcinoma. STXBP6 was demonstrated to be methylation-driven and therefore to be differentially expressed in lung adenocarcinoma. The methylation (HM27K) and gene expression (LabChip6000) status were measured in 32 pair-wide cancer-adjacent samples by genome-wide beadchip and compressive bioinformatics, biological function experiments and clinical relevance association analysis were conducted. The study was performed rigorously and the findings are quite interesting.
Major Compulsory Revisions
1, Figure 1B should be checked again. The main difference from the data is not derived from case-control in PC1. I am wondering which factors could interpret this main difference? Batch effect?
2, From Figure 2D, we can find the methylation status in both cancer and normal cell line are quite low (<30%), that means, in majority cancer cells, the methylation status might be non-methylation, right?
3, The prognosis analysis was based on microarray data therefore the bias and false-positive/negative cannot be guaranteed. The result should be validated or else the credit is limited. The author could collect another dataset from GEO to validate this conclusion if you don't want to do it in your own samples.

Minor Essential Revisions
1, in the abstract section, 5-aza-2'-deoxycytidine treatment cannot be used to showed hyper-methylation of STXBP6 in tumor cells. 5-aza-2'-deoxycytidine treatment could be applied to show the relationship between the methylation and expression, However, 5-aza-2'-deoxycytidine would change the methylation profile genome-widely, therefore, it is also hard to show the relationship between the methylation status of the promoter and the corresponding gene expression. More rigorous method should be introduced such as reporting system to demonstrate the slicing were caused by DNA methylation of promoter regions.
2, When the author claim STXBP6 to be a novel biomarker, please introduce the currently most lasted potential biomarkers for lung adenocarcinoma, such as APC, SHOX2, so that the reader could be have more context for this

study. Meanwhile, the current paper could be connected with previous studies based on these hub genes by computational search strategy.
3, For the Figure 1E, same color scale but reverse value could be applied to show the relationship between methylation and gene expression. In the present figure, the color is different and it is little hard to get a clear comparison.
4, In the Table 1, please change M-value to Beta-value so that the reader could get the point quickly. In addition, what does 4600 in Me_probeb represents? Why not use the cg number of the probe?
5, in the section of 'Overexpression of STXBP6 in lung cancer cells', only one primer was applied. How many transcripts does STXBP6 have? Do you think this primer could measure the expression accurately?
6, In PCA and Cluster analysis, it is not clear whether whole probes or only differential probes were enrolled? Please make explicit note in the manuscript. Meanwhile, I didn't find the result of cluster analysis to the samples, they are forgot to attached in the manuscript? I don't know the function/aim of the clusters analysis to the genes or probes?
7, In the Figure 2C, please don't have space between Chr and 14, they are together.
8, Please make explicit annotation to the population, age, race of the samples in current study.
9, please provide genome assembly version of the coordination (hg19, hg18 or hg38).
10, The scatter plot should be provided to show the relationship between the methylation and expression of STXBP6 in the beadchip data.

Reviewer #2:
Technical Comments to the Author:
1,the manuscript chose 32 pairs non-smoking women case-control sample combine the bioinformatics analysis with further experiment evidence to support the authors' result.

2、the weakness whole story just duplicate some previous work on DNA methylation and gene expression on some other cancers, although, they use their own array data.

3、 The combination of the public microarray data and the author's own ones were used to filter the candidate biomarker in lung adenocarcinoma. However, a limited number of probe in microarray data, why not add some second-generation sequencing data in lung adenocarcinoma? Most of the time, a single biomarker is not stable to predict the patients' survival time. If they add much more NGS data, they may find a useful gene set to predict patients survival time.

Remarks to the Author:
1,the manuscript chose 32 pairs non-smoking women case-control sample combine the bioinformatics analysis with further experiment evidence to support the authors' result.

2、the weakness whole story just duplicate some previous work on DNA methylation and gene expression on some other cancers, although, they use their own array data.

3、 The combination of the public microarray data and the author's own ones were used to filter the candidate biomarker in lung adenocarcinoma. However, a limited number of probe in microarray data, why not add some second-generation sequencing data in lung adenocarcinoma? Most of the time, a single biomarker is not stable to predict the patients' survival time. If they add much more NGS data, they may find a useful gene set to predict patients survival time.