Comments to the Authors,

Dr. Wang conducted a genetic case-control association study and further biological validation to demonstrate a functional SNP which is highly linked with the GWAS identified NSCLC associated SNP were involved with cancer susceptibility by influencing the transcript binding factor YY1. The idea and the strategy were excellent and it would give a great help to understand the interaction between SNP and TF binding in the complex disease susceptibility. However, I still have several consideration on the study design and the statistic method.

**Major Compulsory Revisions**

1, The author estimated the LD between GWAS SNPs with other adjacent SNPs with Hapmap dataset, However, 1000 Genome data have been released several years ago, the LD block analysis to 1000 Genome data would be more powerful to detect more SNPs with is linked to the interest SNP in this study.

2, In the association study, as an accurate study design the smoking, BMI between case and control population should be almost same or no significantly difference so that the genetic difference could be estimated. Why the authors didn’t control these confounders? Do you think these effects can be adjusted 100% with the statistic model? What’s worse, I cannot understand why the author show the Chi-square P-value in Table 3, rather than logistic P-value? Finally, although the number of candidate SNPs in the study were not too many, the P-value will turn not to be significant after multiple test correction.

3, It is still very difficult to understand the logic of the study. Such as 1) *YY1* is low expressed in lung cancer cell lines. 2) Expression of *DCBLD1* was not significantly different among different genotypes, even though the authors gave several hypothesis. 3) Another difficult question is that for the SNPs or genotype, the distribution was only little different between cancer and normal population and you can find the proportion of the genotype was almost very close, even the P-value was significantly, in such way, the molecular function interpretation should be carefully. I suggest that the eQTL analysis can be repeated again with large sample size to check whether the difference between different genotype were significant, if so, then everything is clear, or else, the author should be give more comprehensive and reasonable explanation.

**Minor Revisions**

1, In table 2, the location of the SNPs should be shown and the order should be same with the genomic position order.

2, In table 3, the test should be adjusted with BMI, Smoking, gender and age.

3, In Figure 2B, the scale should be provided.

4, Figure 2C, the location of the star should be changed.