Dear Editors,

We appreciate the insightful comments provided by the reviewers. They have helped us significantly improve the manuscript. In the revised version, we have addressed all the comments and modified minor typographical errors.

Reviewer Comments:

**# Reviewer 1**Comments to the Author  
This clearly written manuscript details the first GWAS of lymphangioeiomyomatosis (LAM), which used 479 LAM cases and compared them with 1261 controls from the COPDGene cohort, which are well-characterized for lung phenotypes. Cases and controls were genotyped on the same array, which reduces potential bias. Genome-wide significant association was observed for two genotyped SNPs in a novel locus on chr. 15 near the NR2F2 gene; importantly, the SNP associations were replicated in an independent cohort. Follow-up analyses were carried out to aid the biological interpretation of the statistical associations, including using gene-based association tests and chromatin TADs in the region surrounding the genome-wide significant SNPs that both pointed to NR2F2 as the most likely gene influenced by the associated SNPs. Additional functional data added further support to the plausibility of NR2F2, including RNA-seq showing higher expression levels of NR2F2 in LAM and related tumors compared to other cancers and immunohistochemistry showing strong nuclear expression of NR2F2 in tumor samples. A caveat to the functional connection is that there is presently no evidence of the LAM-associated SNPs exerting eQTL effects on NR2F2, but the author appropriately address this limitation in the Discussion.   
  
Major Comments  
1. Given that there is “no precedent for genetic influence on the development of S-LAM,” the authors might consider determining the SNP heritability of S-LAM using their GWAS results to provide the field with this foundational information.

**(RESPONSE)** We appreciate this thoughtful suggestion. We calculated SNP heritability of S-LAM based on both entire genotyped SNPs and SNPs within the LD block for the top associated SNPs using GCTA, and we have added the following paragraph:

→ We calculated the proportion of phenotypic variance explained by the genotyped SNPs, h\_SNP^2. Estimates of h\_SNP^2 vary according to disease prevalence (Supplementary Figure 4). With prevalence set at 1 in 100,000 women, h\_SNP^2 was 15% (0.3% on the observed 0-1 scale).

2. Further details are needed on the source of cases used for both discovery and replication.

**(RESPONSE)**

3. Running GWAS using conditional logistic regression is unique, and it isn’t clear what is being conditioned in the model.  Please clarify the model and provide the rationale for conditioning.

**(RESPONSE)** We are sorry for this uncertainty. LAM is a very rare disease, and cases were selected from different countries. Thus, to minimize the genetic heterogeneity between cases and controls, each case was matched with those corresponding controls with PC scores. We have clarified our CLR by adding the following sentences:

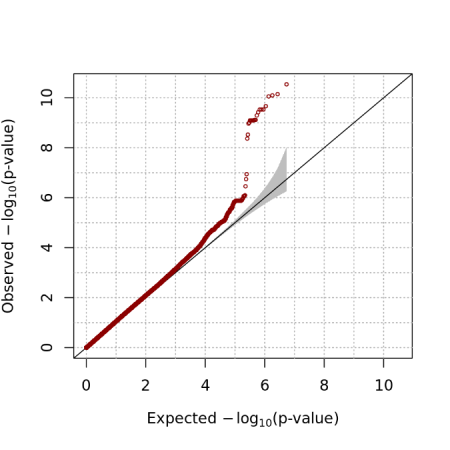
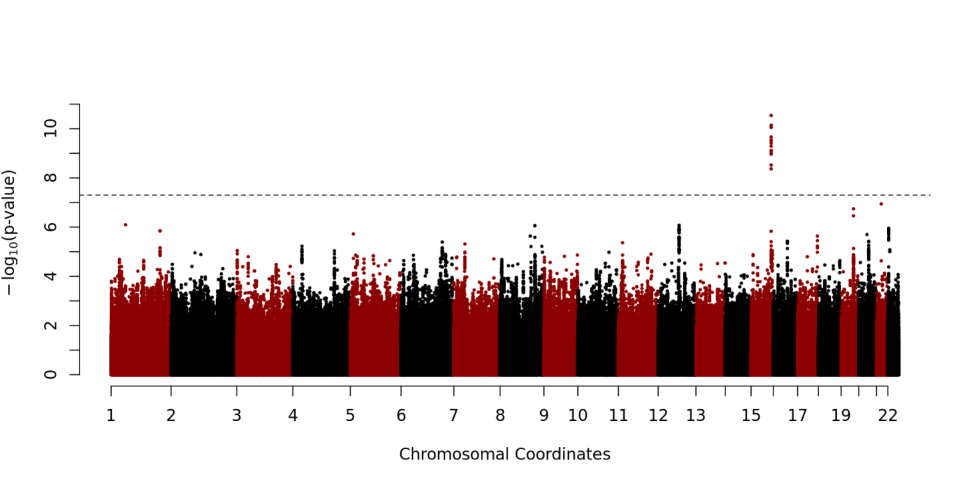
→ Thus CLR was conducted conditioning on the matched cases and controls, by the first 2 PC scores. Our CLR can be expressed as follows:

for th strata,

where and indicate the phenotype and genotype of *j*th subject in the *i*th matched strata, respectively.

4. GWAS analyses focused only on ~550,000 genotyped SNPs, with imputation using Haplotype Reference Consortium for fine-mapping the genome-wide significant locus. It is unclear why imputation was not carried out for the discovery GWAS, as it is feasible to perform with cases vs. controls obtained from a different dataset. Lack of imputation doesn’t invalidate the current findings but means that the study may have missed other novel loci across the genome.

**(RESPONSE)** We appreciate your insightful comment. We performed whole genome imputation by using Haplotype Reference Consortium and 5,427,337 SNPs remained after quality control. At the genome-wide significant level of 510-8, we could not find any other significant SNPs except the SNPs identified in our previous analyses (See the following figures). We updated the manuscript accordingly.



Minor comment  
1. There are several instances in the manuscript where the text are incorrectly shown in sub/superscript (see for example Table 2).

**(RESPONSE)**

2. Adding gene annotations to the SNPs presented in Supplementary Table 2 would be helpful. Also, the SNP on chr. 4 might not be listed correctly (rs10517300 is not found in the GWAS catalogue).   
**(RESPONSE)** We really appreciate this careful observation and are sorry for this mistake and confusion. It was removed from the table. SNPs associated with nicotine addiction were chosen from the GWAS catalogue, and those SNPs and SNPs correlated with those were removed. rs10491551 was not found from the GWAS catalogue but chosen because of its high correlation with rs12348139 (r2 = 1). We added the mapped gene information and SNP and added the following sentence:

→ rs10491551 is included due to its high correlation with rs12348139 in the GWAS catalogue (r2 = 1).

**# Reviewer 2**Comments to the Author  
I appreciate the opportunity to review the manuscript submitted by Kim and colleagues entitled “A Genome-wide Association Study implicates NR2F2 in Lymphangioleiomyomatosis Pathogenesis” for the European Respiratory Journal. This manuscript summarizes a GWAS of 549,000 SNPs (MAF>0.05) in 479 European descent and White (unclear of Hispanic ethnicity although there were US and Puerto Rican samples) LAM female cases and 852 matched controls with replication in 196 non-Hispanic White females and 409 controls. GWAS in the discovery cohort identified two SNPs in a gene desert that reached genome-wide significance for LAM susceptibility closest to the gene encoding NR2F2 which also replicated successfully. Fine mapping of imputed genotype data within the LD blocks for the top associated SNPs identified a possible causative locus associated with LAM while gene-based SKAT-O testing within the Chr15 region only found a significant association for NR2F2 and no other genes. Additional database searches for genotype and RNA expression data revealed that (1) the associated SNPs were within the boundaries of a topologically-associated domain for NR2F2 and no other genes, (2) the allele frequency of these SNPs in these US and European LAM cohorts were remarkably different from that of different general populations, and (3) NR2F2 gene expression was significantly higher in LAM-affected tissues compared to other cancer and normal tissue from TCGA and GTEx. Finally, immunohistochemistry showed strong nuclear expression of NR2F2 in LAM lung and renal angiomyolipoma cells.   
This is the largest genetic study of LAM cases and unaffected controls assembled to date from an international effort which identified a locus not related to the TSC1 and 2 genes. In addition, this study used appropriate fine-mapping of genotyped and imputed data complemented by different bioinformatics and molecular phenotyping approaches on LAM-affected tissues to validate the role of the top GWAS associations that reach genome-wide significance. This study provides insight into a novel pathogenic mechanism for a largely fatal pulmonary disease which was well substantiated in a relatively large cohort of individuals with a rare disease that primarily impacts women. This study is also a good and uncommon example of a GWAS of individuals with a rare disease related to a known gene (TSC1/2) that was able to find a strongly associated novel gene locus. There are several concerns which I recommend the authors address or clarify.

I recommend the following major revisions:  
1) The authors should clarify what was the “careful scrutiny” used to prevent overlapping subjects between the discovery and replication cohorts since GWAS data wasn’t available in the replication cohort to assess IBD.

**(RESPONSE)**

2) It is unclear whether the cases and controls in the discovery cohort were NON-Hispanic White? It is clear from Figure 1, supplementary table 1, and the text that the cohort consisted of European Whites, but the issue of Hispanic ethnicity was not clearly outlined for the discovery cohort and could have implications on environmental/cultural exposures and ancestral background confounding genetic associations. The authors should at least provide the proportion of Hispanic subjects, if this data is available and an explanation of how well the PC approach has accounted for confounding by variable ancestral backgrounds which I suspect it did based on the detailed descriptions of the PC’s for this cohort.

**(RESPONSE)**

3) It was not clear whether fine-mapping covered lower frequency variants across the Chr 15q LD blocks containing the top associations. The authors outline in Figure 1, that allele frequencies <0.05 were excluded from the initial discovery GWAS, but should outline whether low frequency or rare variants were considered during the fine mapping steps, including SKAT-O. Was there a MAF threshold? If these variants were otherwise excluded, specify the obvious reasons why a MAF threshold was selected (ie poor imputation quality, inability to genotype any associations identified, etc.). Since LAM is considered to be caused by rare pathogenic TSC1/2 gene variants and LAM is a rare disease, this is an important consideration.  
**(RESPONSE)** We are sorry for this confusion. As you pointed out, we excluded rare variants in the initial discovery GWAS but both rare and common SNPs in three genes were considered for SKAT-O analysis. We added the following sentences to clarify our SKAT-O analysis:

→ We also conducted gene-based analyses for genes nearly located with the genome-wide significant SNPs from the discovery GWAS using the SKAT-O statistic [23]. Gene-based analysis usually includes rare variants. Thus QC with MAF 5% criteria was not applied, and both rare and common SNPs were used for SKAT-O analysis. Age, squared age and 10 PC scores were included as covariates.

23. Lee S, Emond MJ, Bamshad MJ, Barnes KC, Rieder MJ, Nickerson DA, Team ELP, Christiani DC, Wurfel MM, Lin X. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. The American Journal of Human Genetics 2012: 91(2): 224-237.

4) Diagnostic criteria for a diagnosis of LAM were not outlined in the manuscript nor summarized in a baseline characteristics table. This data would be good to see in the methods as diagnostic criteria and in a table.

**(RESPONSE)**

5) It’s unclear how variation in the TSC genes were considered in this GWAS. Were patients with known variants excluded from analysis or were these genes sequenced/genotyped for known pathogenic variants? If not, was TSC variation associated with LAM at the level of any individual SNPs or at the gene-level using imputed data or SNP data not filtered for a low allele frequency? I presume that identifying associations at these known loci would be tough since pathogenic variants were rare and excluded from GWAS, but also difficult to impute. However, more common SNPs in LD with pathogenic variants might be associated or gene-level associations could also represent LD-tagged SNP associations for strongly association pathogenic loci (TSC genes). These questions are more to address whether the NR2F2 associates with LAM risk through interactions with TSC genes or independently and might not be addressable in a study based on GWAS genotyping and imputed data alone (a limitation to outline in the discussion section). Candidate gene/SNP studies are performed many time with GWAS and should be considered here if the quality of data is appropriate (ie genotyping coverage, imputation quality).

**(RESPONSE)***.*

5) Increased expression of *NR2F2* was found in a small number of LAM-related tissues compared to tissue from TCGA and GTEx which is not the same thing as an eQTL analysis of the top associated SNPs. In the discussion, the authors states that the novel SNP loci were not eQTL’s when evaluated. However, the authors should outline if and how this analysis was performed in more detail the methods or results sections.

**(RESPONSE)** We are sorry for the uncertainty. Due to lack of RNA-seq samples for LAM, we used GTEx database in order to find eQTL of SNPs in LD block. We added several sentences in method and results for eQTL mapping as follows:

→ We also searched the single-tissue *cis*-expression quantitative trait loci (eQTL) of all SNPs in the LD block from public database provided by GTEx in release v7. They provide results of eQTL analysis for each SNP-gene pair consisting of all SNPs within 1 Mb upstream and downstream of transcription start site of each gene. FastQTL was used for *cis-*eQTL mapping [34] with covariate adjustment of top three PC scores, genotyping platform, sex and a set of relavant variables identified using PEER method [35]. The detailed workflow can be found on the GTEx website (https://www.gtexportal.org/home).

→ We queried the GTEx database for SNPs in the same LD block to find the eQTL SNPs. Unfortunately, there was no significant association with an FDR threshold of 0.05.

34. Ongen H, Buil A, Brown AA, Dermitzakis ET, Delaneau O. Fast and efficient QTL mapper for thousands of molecular phenotypes. Bioinformatics 2015: 32(10): 1479-1485.

35. Stegle O, Parts L, Durbin R, Winn J. A Bayesian framework to account for complex non-genetic factors in gene expression levels greatly increases power in eQTL studies. PLoS computational biology 2010: 6(5): e1000770.

Minor revisions:   
1) Some of the tables and text throughout have font that is unnecessarily superscripted.

**(RESPONSE)**

2) The term long-non coding RNAs should be defined before the use of the term “lncRNA” on page 11.

**(RESPONSE)** Thanks for this comment. We have added the definition of IncRNA in the manuscript.

3) Number of SNPs for GWAS on page 7 of the methods does not match the numbers on Figure 1. This might be related to the exclusion of nicotine dependence genes.

**(RESPONSE)** We are sorry for this typo. We updated our manuscript accordingly.

4) In the abstract, please provide a sample size for the replication cohort and state that the controls in the discovery were matched.  
**(RESPONSE)** We have added the information to abstract.

**# Reviewer 3 (Statistical Reviewer Comments)**:  
1. The lack of genome-wide imputation is a major limitation – what was the justification for not performing this prior to the GWAS?

**(RESPONSE)** We would like to thank the reviewer for pointing out this important issue. In our previous manuscript, we focused on the targeted imputation for the genome-wide significant region. However, we agree that the lack of genome-wide imputation is a major limitation, and performed genome-wide imputation. The same method was applied to the imputed genotypes, and method and result have been updated with these results accordingly.

2. Page 8 of 47, line 14: How were “non-white” LAM subjects excluded?

**(RESPONSE)**

3. Page 9 of 47, line 23: Controls from a source entirely independent of the discovery would add more confidence to the results of the replication. Although the authors provide strong evidence that smoking behaviour is unlikely to account for the signals, there is still uncertainty as to whether other differences between the case and control populations, besides LAM status, might be accounting for the differences.

**(RESPONSE)** We appreciate this thoughtful comment. Cases have been selected all over the world, and thus there may exist genetic differences between cases and controls. To minimize such genetic differences, only non-hispanic whites were considered, and controls were matched with genetic background. Nonetheless, we agree with the concern of this reviewer that potential effects may not be completely removed. Therefore, we used two additional datasets, MESA and UK Biobank, as controls for replication studies, and conducted the further association analyses. The following sentences have been added:

→ Replication analysis was performed for the 2 SNPs with association with LAM using 196 additional non-Hispanic white (NHW) S-LAM patients for three control datasets; 1) 409 NHW healthy females from COPDGene participants who were not used for discovery analyses, 2) 1,121 Hispanic white females from MESA dataset, and 3) 225,731 white British females in UK Biobank dataset. Similar ORs for association of the minor allele of these SNPs with S-LAM were seen in the each replication dataset (Table 1).

4. Page 9 of 47, line 24: what is the pre-specified threshold for replication?

**(RESPONSE)** We are sorry for this confusion. We used almost same quality control procedure for replication as for the discovery analyses. We selected NHW females first, and SNPs were removed if the Hardy-Weinberg equilibrium (HWE) test gave P < 1×10-5; minor allele frequency (MAF) was < 0.05; or genotype call rates were less than 95%. We added the following paragraph to the Method section:

→ For each control dataset, we used genotyped or imputed data depending on the genotype availability of the genome-wide significant SNPs. Imputation was conducted by using 1000 Genomes Project [20] as reference genotypes. We included only NHW female and excluded variants if P-value of HWE test is less than 1×10-5, MAF is less than 0.05 or missing genotype rate is greater than 5%.

20. Consortium GP. An integrated map of genetic variation from 1,092 human genomes. Nature 2012: 491(7422): 56.

5. Page 10 of 47, line 24 and Table 2 legend – please clarify the method used for imputation (was it IMPUTE2?)

**(RESPONSE)** We are sorry for this uncertainty. Prior to imputation with PBWT, EAGLE2 was utilized for pre-phasing according to the recommendation of the Sanger Imputation Service in terms of speed and accuracy. Then we used the Positional Burrows–Wheeler Transform (PBWT) package for imputation. We modified sentences as follows:

→ Before imputation, pre-phasing was conducted with EAGLE2 v2.0.5 [25], and then the Positional Burrows–Wheeler Transform (PBWT) package [26] was used for the imputation according to the imputation pipeline recommended by Sanger Imputation Service in terms of speed and accuracy.

→ **Table 2. Statistical analyses of imputed SNPs with CLR.** Imputation was conducted using EAGEL2 and PBWT for pre-phasing and imputation with the Haplotype Reference Consortium as reference panel.

25. Loh P-R, Danecek P, Palamara PF, Fuchsberger C, Reshef YA, Finucane HK, Schoenherr S, Forer L, McCarthy S, Abecasis GR. Reference-based phasing using the Haplotype Reference Consortium panel. Nature genetics 2016: 48(11): 1443.

26. Durbin R. Efficient haplotype matching and storage using the positional Burrows–Wheeler transform (PBWT). Bioinformatics 2014: 30(9): 1266-1272.

6. Page 10 of 47, line10 – please justify the use of CLR as this is non-standard in GWAS.

**(RESPONSE)** We appreciate this comment, and agree on this comment. In the discovery study, cases were collected from several countries, and effects of population substructure can be substantial. It has been shown that CLR is more efficient than unconditional logistic regression (LR) if the matching variables are true confounders and the number of controls excluded from the analysis due to matching is moderate. According to this comment, we included the results from LR, and we added the following paragraph:

→ There are potential limitations to our study. Although our cohort of samples was large for a rare disease like S-LAM, it was of only moderate size for GWAS. In order to overcome this issue, we employed a worldwide recruitment strategy for S-LAM patients of European origin. Although our controls were all from the USA, they were selected for European ancestry to minimize population stratification issues. In addition, we employed EIGENSTRAT to remove genetic outliers from both our S-LAM and control cohorts to further reduce genetic heterogeneity. With PC scores, each case was matched with two controls to minimize the genetic heterogeneity between cases and controls, and CLR was utilized for matched case and control design. According to some literatures, CLR can improve the efficiency of the study rather than unconditional logistic regression (LR) if the variables used for matching are true confounding variables and moderate number of controls were excluded from the analysis as a result of matching [57-63]. Even though we dropped some controls out after matching, we found that CLR generated much significant result, compared to LR for top two significant SNPs on chromosome 15 (Supplemental Table 6).

57. Breslow N, Day N, Halvorsen K, Prentice R, Sabai C. Estimation of multiple relative risk functions in matched case-control studies. American Journal of Epidemiology 1978: 108(4): 299-307.

58. Kupper LL, Karon JM, Kleinbaum DG, Morgenstern H, Lewis DK. Matching in epidemiologic studies: validity and efficiency considerations. Biometrics 1981: 271-291.

59. McKinlay SM. Pair-matching--A reappraisal of a popular technique. Biometrics 1977: 725-735.

60. THOMPSON WD, KELSEY JL, WALTER SD. Cost and efficiency in the choice of matched and unmatched case-control study designs. American journal of epidemiology 1982: 116(5): 840-851.

61. Thomas DC, Greenland S. The relative efficiencies of matched and independent sample designs for case-control studies. Journal of chronic diseases 1983: 36(10): 685-697.

62. Miettinen OS. Estimation of relative risk from individually matched series. Biometrics 1970: 75-86.

63. Luca D, Ringquist S, Klei L, Lee AB, Gieger C, Wichmann H-E, Schreiber S, Krawczak M, Lu Y, Styche A. On the use of general control samples for genome-wide association studies: genetic matching highlights causal variants. The American Journal of Human Genetics 2008: 82(2): 453-463.

7. Page 12 of 47, line 24: Please make clear here that these two SNPs are highly correlated and represent one signal.

**(RESPONSE)** We appreciate this comment. We have added the following sentence to clarify that two significant SNPs are derived from one signal:

→ Based on the location of the two SNPs and their LD relationship, these two associations might be derived from a unique signal.

8. Page 13 of 47, lines 9 and 13: Nicotine addiction cannot be a confounder in a genetic study as it does not influence both exposure (SNP) and outcome. Nicotine addiction can be on the causal pathway between an exposure and outcome (i.e. the SNP is influencing disease risk via nicotine addiction). Please correct this wording in the manuscript.

**(RESPONSE)** We thank the reviewer for pointing out this error. Controls consist of only smokers, and this is a selection bias not a confounder effect. We have modified that sentence as follows:

→ All subjects from COPDGene cohort were smokers, and such selection bias can introduce the significant results for SNPs associated with nicotine addiction. We checked p-values for SNPs associated with nicotine addiction from the GWAS catalog [37] and other SNPs correlated with those (r2 >0.8) (Supplementary Table 2). None of those SNPs showed a significant difference in allele frequency in the LAM and COPDGene cohorts, indicating that our findings are not affected by nicotine addiction SNPs.

37. MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, Junkins H, McMahon A, Milano A, Morales J. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). Nucleic acids research 2016: 45(D1): D896-D901.

9. Page 13 of 47, line 5: Please present the genomic inflation factor lambda. A visual inspection of the QQ plot is insufficient to prove that there is no inflation.

**(RESPONSE)** We appreciate this suggestion. We added the genomic inflation factor as follows:

→ Quantile-quantile plots for CLRs and Manhattan plots demonstrated that the distribution of observed P-values met the expected distribution, with the exception of the two SNPs (Figure 2), indicating that the analyses were free of systematic P-value inflation (genomic inflation factor = 1.07).

10. Page 15 of 47, lines 1 to 9; more clarity is needed here to explain how the data presented in Supplementary Figures 4 to 7 demonstrate an interaction between the LAM-associated SNPs and NR2F2.

**(RESPONSE)** We thank the reviewer for pointing out this problem. We have added the following sentence to explain the meaning of the heatmap:

→ Supplementary Figures 6-9 display Hi-C heatmaps around 1 Mb sub-TADs containing the GWAS SNPs and NR2F2 near the boundary of the sub-TAD within a larger TAD in all four of these cells/tissues, and NR2F2 was the only protein-coding gene.

11. SKAT-O analysis: please clarify the criteria for SNP inclusion. Were both intronic and exonic SNPs included? Was there any MAF filter? Gene-based approaches such as SKAT-O are usually intended to look for the cumulative effect of rare variants within a gene, however, it is stated in the methods that SNPs with MAF<5% were excluded.

**(RESPONSE)** We are sorry for this confusion. As you pointed out, we excluded rare variants in the initial discovery GWAS but both rare and common SNPs in three genes were considered for SKAT-O analysis. We added the following sentences to clarify our SKAT-O analysis:

→ We also conducted gene-based analyses for genes nearly located with the genome-wide significant SNPs from the discovery GWAS using the SKAT-O statistic [23]. Gene-based analysis usually includes rare variants. Thus QC with MAF 5% criteria was not applied, and both rare and common SNPs were used for SKAT-O analysis. Age, squared age and 10 PC scores were included as covariates

23. Lee S, Emond MJ, Bamshad MJ, Barnes KC, Rieder MJ, Nickerson DA, Team ELP, Christiani DC, Wurfel MM, Lin X. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. The American Journal of Human Genetics 2012: 91(2): 224-237.