**Figure 1. Workflow of quality control for the LAM GWAS discovery data set.** Multiple standard quality controls were performed for both cases (female S-LAM subjects) and controls (healthy women without COPD from COPDGene consortium) to exclude outlier SNPs and subjects.

Definition of abbreviations : LAM = Lymphangioleiomyomatosis; GWAS = Genome-Wide Association Study; S-LAM = Sporadic LAM; COPD = Chronic Obstructive Pulmonary Disease; SNP = Single-Nucleotide Polymorphism; HWE = Hardy-Weinberg Equilibrium test; MAF = Minor Allele Frequency; IBS = Identity-By-State

**Figure 2. Quantile-quantile plot and Manhattan plot for discovery LAM GWAS dataset.** a) The observed distributions of P-values for 549,591 genotyped SNPs are plotted relative to the expected (null) distribution for each of CLR analyses. b) Each dot represents the P-value of a single SNP, plotted on the genome scale at bottom. The Y-axis value is the negative logarithm of the P-value for association between each genotyped SNP and LAM. Two SNPs on 15q26.2 met genome-wide significance (P < 5 x 10-8) by CLR.

Definition of abbreviations : LAM = Lymphangioleiomyomatosis; CLR = Conditional Logistic Regression; SNP = Single-Nucleotide Polymorphism

**Figure 3. Genomic region on chromosome 15 containing the SNPs associated with LAM.** a) Ideogram of chromosome 15. b) Three Mb region containing the SNPs associated with LAM. Manhattan plot at top shows P-values for SNPs in this region, including the two SNPs meeting genome-wide significance (red dots). There are 3 protein-coding genes *NR2F2*, *MCTP2*, and *SPATA8* which were represented in yellow shaded boxs, and many lncRNAs in this region. c) Expanded Manhattan plot of the 250kb region containing the genotyped and imputed SNPs showing association with LAM. SNP rs41374846 is indicated by purple, and other SNPs are colored according to their *r*2 value in relation to rs41374846.

Definition of abbreviations : SNP = Single-Nucleotide Polymorphism; LAM = Lymphangioleiomyomatosis

**Figure 4. Comparison of *NR2F2* expression in kidney angiomyolipoma/LAM with cancer (TCGA) and normal tissues (GTEx).**

Boxplot figures are shown to compare expression of *NR2F2* in 4 angiomyolipoma and one abdominal LAM lesion with 2463 cancers of 27 types (from TCGA, brackets on x-axis include the number of samples analyzed per tumor type; abbreviations are explained in Supplementary Table 5) in RSEM units (a); and with ~7,000 samples of 47 normal tissues (from GTEx) in RPKM units (b). Remarkably, NR2F2 gene expression is the highest compared to all TCGA tumors and higher compared to most GTEx normal tissues; similar to cervix, fallopian tubes, uterus and ovaries. The median value, interquartile range, and 95% ranges are shown, with outliers indicated by circles. In the X axis, the each number in brackets is the number of samples corresponding each tissue.

Definition of abbreviations: TCGA = The Cancer Genome Atlas; GTEx = Genotype-Tissue Expression; LAM = Lymphangioleiomyomatosis; RSEM =RNA-Seq by Expectation Maximization; RPKM = Reads Per Kilobase Million; Full terms for TCGA tumor abbreviations are explained in Supplementary Table 5.

**Figure 5. Immunohistochemistry for *NR2F2* in LAM and kidney angiomyolipoma.** Strong nuclear staining is seen in lung LAM cells (a) and angiomyolipoma cells (b) (brown stain). Some other cells also have nuclear staining for *NR2F2* but most do not.

Definition of abbreviations: LAM = Lymphangioleiomyomatosis