The Vitamin D Binding Protein axis modifies disease severity in Lymphangioleiomyomatosis

Authors: Suzanne Miller[1], Clare Coveney[2], Janice Johnson[1], Aliki-Eleni Farmaki[3], Nishant Gupta[4], Martin D. Tobin[3,5], Louise V. Wain[3,5], Francis X. McCormack[4], David J. Boocock[2] and Simon R. Johnson[1,6,7].

Affiliations:

[1]Division of Respiratory Medicine, University of Nottingham and National Institute for Health Research, Respiratory Biomedical Research Centre, Nottingham, United Kingdom.

[2]John van Geest Cancer Research Centre, Nottingham Trent University, Nottingham. United Kingdom.

[3]Department of Health Sciences, College of Life Sciences, University of Leicester, Leicester, United Kingdom.

[4]Division of Pulmonary, Critical Care and Sleep Medicine, University of Cincinnati, Ohio, USA.

[5]National Institute for Health Research Leicester Respiratory Biomedical Research Centre, Glenfield Hospital, Leicester, United Kingdom.

[6]National Centre for LAM, Queen's Medical Centre, Nottingham, United Kingdom.

[7]Nottingham Molecular Pathology Node, Nottingham, United Kingdom.

Corresponding authors: Professor Simon Johnson. Division of Respiratory Medicine, University of Nottingham, Nottingham, NG7 2UH. United Kingdom.

simon.johnson@nottingham.ac.uk

Author contributions: SRJ conceived and designed the study. SM, SRJ, JJ, NG and FXM collected clinical information and samples. SM, SRJ, CC, AEF, MDT, LVW and DB analysed and interpreted the data. SM and SRJ wrote the manuscript. All authors critically reviewed and approved the final manuscript.

Take home message: The vitamin D binding protein and *GC* genotype are associated with lung function and survival in women with LAM.

Abstract:

Background: Lymphangioleiomyomatosis (LAM) is a rare disease of women. Decline in lung function is variable making appropriate targeting of therapy difficult. We used unbiased serum proteomics to identify markers associated with outcome in LAM.

Methods: 101 women with LAM and 22 healthy controls were recruited from the National Centre for LAM (Nottingham, UK). 152 DNA and serum samples with linked lung function and outcome data were obtained from patients in the NHLBI LAM Registry (USA). Proteomic analysis was performed on a discovery cohort of 50 LAM and 20 control sera using a SCIEX SWATH mass spectrometric workflow. Protein levels were quantitated by ELISA and SNPs in *GC* encoding Vitamin D Binding Protein (VTDB) genotyped.

Results: Proteomic analysis showed VTDB was 2.6 fold lower in LAM than controls. Serum VTDB was lower in progressive compared with stable LAM (p=0.001) and correlated with diffusing capacity (p=0.01). Median time to death or lung transplant was reduced by 46 months in those with *CC* genotypes at rs4588 and 38 months in those with non-A containing haplotypes at rs7041/4588 (p=0.014 and 0.008 respectively).

Conclusions: The VTDB axis is associated with disease severity and outcome, and *GC* genotype could help predict transplant free survival in LAM.

Abstract word count: 200

Introduction:

Lymphangioleiomyomatosis (LAM) is a rare multi-system disease characterised by lung cysts and lymphatic abnormalities. The disease is almost exclusively restricted to women, of whom it affects around 9 per million and can occur both sporadically and in those with tuberous sclerosis complex (TSC)[1, 2]. In LAM, cysts progressively replace the lung parenchyma leading to recurrent pneumothorax and often respiratory failure over a variable period of years[3]. Lymphatic obstruction leads to chyloptysis, chylous effusions and ascites. Around half of patients with sporadic LAM and most with TSC-LAM also have angiomyolipomas, a benign tumour, generally occurring in the kidneys[2]. The lungs and lymphatics of patients are infiltrated by LAM cells: a clonal, metastatic cell with a combined smooth muscle and melanocyte phenotype characteristic of perivascular epithelioid cell neoplasms[4]. LAM cells have bi-allelic TSC mutations[5] which lead to hyper-activation of the mechanistic target of rapamycin (mTOR), a component of two multi-protein complexes, controlling proliferation, migration, autophagy and metabolism[6].

Most women with LAM lose lung function at an accelerated rate with FEV_1 declining by 70-140ml per year[7, 8] however, some progress rapidly whilst others can remain stable for many years[3, 9]. Treatment with mTOR inhibitors prevents loss of lung function in most with progressive disease[8-10]. Recognising progressive disease in individuals with mild lung function impairment is important, although generally requires multiple measurements over a prolonged period[7]. Markers of disease activity are therefore required to predict those at risk of loss of lung function to allow treatment before this occurs. Further, stratification of patients with active disease could reduce the size, duration, cost and feasibility of phase II studies of new therapies.

A number of clinical and serum prognostic factors have been identified. Elevated serum Vascular Endothelial Growth Factor-D (VEGF-D) is associated with both the presence of LAM[11] and more rapid loss of lung function. Presentation with dyspnoea rather than pneumothorax and a response to bronchodilators have been associated with worse outcomes[12-14] whereas post-menopausal status is associated with slower lung function loss[7, 15]. Despite this, it is not possible to accurately predict prognosis within individuals. Here we have used serum proteomics to identify proteins associated with the presence and severity of LAM and identified that changes in Vitamin D Binding Protein (VTDB) and it gene, *Group-Specific Component* (*GC*), are associated with disease severity and survival in LAM.

Materials and Methods:

Patients and sample collection

101 women with LAM and 22 healthy control women were recruited between 2011 and 2016 from the National Centre for LAM, Nottingham, UK, (Figure 1). Ethical approval was obtained from the East Midlands Research Ethics Committee (13/EM/0264). All subjects provided written informed consent. A second cohort of 152 women with LAM recruited between 1998 and 2001 in the National Heart Lung and Blood Institute (NHLBI) LAM Registry was used for replication and to study long term survival[16] (Figure 1). Baseline chest and abdominal CT, serial lung function serum and DNA at recruitment was obtained for all subjects. Clinical assessment, lung function and sample analysis for both cohorts are described in the online supplement. Due to duration of follow-up, all-cause mortality or the need for lung transplant, was only available for NHLBI Registry cohort and was obtained by querying the United States National Death Index and the United Network for Organ Sharing

databases. As data on the use of rapamycin was not available for this cohort, outcome data were censored at 2010 before rapamycin was widely used for the treatment of LAM in the USA.

Proteomics

70 serum samples (50 LAM and 20 controls) were analysed on a SCIEX TripleTOF 6600 mass spectrometer hyphenated to an Eksigent nanoLC 425 system using the SCIEX SWATH mass spectrometric workflow[17]. Tandem mass spectrometry (MS/MS) spectra were searched using ProteinPilot 5.0 (SCIEX) with the Swissprot human database (Jan 2015) at 1 % false discovery rate with an identification focus on biological modifications. SWATH data were aligned to library files in PeakView (SCIEX), uploaded and processed using the SCIEX OneOmics platform[18]. Full details are given in the online supplement.

Serum protein quantification

Serum VTDB, Alpha-1 acid glycoprotein 1 (A1AG1) and VEGF-D were determined in the UK cohort using Quantikine ELISA kits DVDBP0, DAGP00 and DVED00 respectively (R & D Systems, UK). VTDB in the NHLBI Registry was measured using Quantikine ELISA kit DVDBP0B (R & D Systems, UK).

Genotyping

DNA was extracted from whole blood using the Wizard Genomic DNA Purification Kit (Promega, UK). As *GC* genotype varies across populations, genetic analysis was confined to those of European ancestry. 65 UK LAM samples and 168,141 unrelated control women of European ancestry from the UK Biobank were genotyped using the Affymetrix Axiom UK Biobank array. Ancestry was determined from k-means clustering of the first two principal components from the genome-wide SNP data[19]. The NHLBI LAM registry cohort were genotyped using KASP PCR genotyping (LGC Genomics service. Herts, UK) with ancestry obtained by questionnaire.

Statistical analysis

Proteins identified by proteomics were considered differentially expressed if they were ≤ -2 or ≥ 2 Log_2 fold different between groups with a confidence ≥ 0.7 as described[18]. Welch's t-test or Mann-Whitney U tests were used for categorical data, linear regression and Spearman's correlation for continuous data. *GC* allele frequencies for women with LAM and UK Biobank controls were compared using Chi-squared tests[20]. Survival analyses were performed using Kaplan-Meier plots with differences analysed by Mantel-Cox log rank test. Analyses were performed using GraphPad Prism v7 and SPSS v24 (IBM).

Results:

Discovery cohort and serum proteomics

The first 50 UK women with LAM enrolled who were not treated with an mTOR inhibitor and 20 healthy control women formed the discovery cohort. The cohort was divided into more progressive and stable disease based upon a retrospective loss of FEV_1 of more than 50ml/yr over a mean period of observation of 11 (± 4) years. Those with progressive disease had lower FEV_1, DL_{CO} and higher serum VEGF-D values but were of similar age and disease duration as those with stable disease (Table 1).

Mass spectrometry of the 70 serum samples identified 126 proteins including the serum proteins albumin, haemopexin, acid glycoprotein, immunoglobulins, complement components, clotting factors, proteases and protease inhibitors (Table E1). VTDB levels were 2.6-fold lower (confidence 0.65) in LAM than healthy control women (Table E2). To identify markers of severity we compared the proteomic profiles of those with stable and progressive disease. A1AG1 levels were 3.6-fold higher (confidence 0.70) in those with progressive compared with stable disease. Comparison of pre- and post-menopausal women with LAM did not identify differentially expressed proteins at the pre-specified confidence level.

Serum protein quantification

Mass spectrometry findings were validated using ELISAs for VTDB and A1AG1. Consistent with the proteomic findings, serum VTDB was lower in 50 women with LAM in the UK

discovery cohort and 27 women with LAM in the UK replication cohort than in controls (p=0.007 and p=0.002, respectively). For the 77 women in the UK discovery (50) and replication cohorts (27) combined, VTDB was 273 ± 96 μg/ml in LAM and 347 ± 92 μg/ml in control women (p=0.002, Figure 2a). When assessed by ELISA, A1AG1 was higher in women with LAM in the discovery and replication cohorts than control women (p=0.04 and p=0.0001, respectively). For all women with LAM, A1AG1 was 910 ± 478 μg/ml and 619 ± 270 μg/ml in control women (p=0.004, Figure 2b).

VTDB is associated with disease severity

VTDB was significantly lower in those with more progressive, compared with more stable lung disease at recruitment (progressive 221 ± 89 μg/ml, stable 299 ± 90 μg/ml, p=0.001, Figure 3a). VTDB level was positively associated with percent predicted DL_{CO} (p=0.01) but not forced vital capacity (p=0.09) or FEV_1 (p=0.23, Figure 3b-d). A1AG1 was higher in those with stable, compared with progressive disease (stable 1004 ± 525 μg/ml, progressive 753 ± 341 μg/ml, p=0.01, Figure E1) but was not related to lung function. Levels of VTDB were not associated with age, age at diagnosis, menopausal status, nature of presenting symptom, the presence of tuberous sclerosis, angiomyolipomas, lymphatic disease or serum VEGF-D level (data not shown). The distribution of VTDB was similar in the 77 untreated and 24 women receiving treatment with rapamycin for LAM, whereas A1AG1 was higher in the rapamycin treated group (rapamycin treated 1132 ± 474 μg/ml, untreated 910 ± 478 μg/ml (p=0.031, Figure E2).

Association of *GC* genotypes with LAM and serum VTDB

As *GC* genotype varies according to ancestry, genetic analyses were restricted to the 65 individuals in the UK and 145 individuals in the NHLBI LAM cohorts of European origin. Two SNPs within *GC* at rs7041 and rs4588 define the major *GC* haplotypes (i) *GC2* where rs7041 (G) and rs4588 (C), (ii) *GC1F* where rs7041(G) and rs4588 (A) and (iii) *GC1S* where rs7041 (T) and rs4588 (A). The allele frequencies at these SNPs in the UK and NHLBI LAM cohorts did not differ from control women in the UK Biobank or each other (Table E3). In both LAM cohorts, as in the general population, serum VTDB was dependent on *GC* genotype[21] (Figure E3).

Association of VTDB protein and genotype with outcome

From the UK cohort, 91 women with LAM had lung function measured over greater than one year after enrolment (64 untreated and 27 receiving rapamycin for LAM). The mean period of observation was 19 months, corresponding to 144 patient years of observation. Within the NHLBI LAM cohort, 136 women with untreated LAM had lung function measured over greater than one year after enrolment with a mean period of observation of 40 months, corresponding to 500 patient years of observation. Serum VTDB was not associated with prospective change in lung function in either cohort (Table 2).

Within the NHLBI LAM Registry cohort, those with low serum VTDB, the *AA* genotype at rs4588 and *TT* at rs7041 had the highest rates of loss of FEV_1 and DL_{CO}, although not significantly so (Table 3). We then examined the relationship of the VTDB axis with time to death or lung transplant in the NHLBI LAM Registry cohort. Although time to death or

transplant was not associated with serum VTDB level (p=0.76, Figure 4a) or rs7041 genotype, there was an association with rs4588 genotype. Median time to death or transplant for the *AA* or *AC* genotype at rs4588 was 150 months compared with 104 months for *CC* (p=0.014, Figure 4b). Median time to death or transplant for all haplotypes with an *A* allele at rs4588 (including *GC1F* and *GC1S* haplotypes) was 150 months compared with 112 months for haplotypes with no *A* allele present (including *GC2*, p=0.008, Figure 4c).

Discussion

We have shown for the first time the VTDB axis is associated with both severity and outcome in women with LAM. VTDB levels were associated with DL_{CO} and disease activity at assessment. Those with progressive disease, defined by a loss of FEV_1 of more than 50 ml/yr, tended to have lower levels of VTDB than those with more stable disease with a loss of FEV_1 of less than 50 ml/yr, despite being matched for age and other clinical manifestations. Haplotypes of *GC* were associated with the time to death or lung transplant. As such, *GC* genotype is the first genetic host factor found to influence transplant free survival in LAM.

VTDB is a glycosylated alpha-globulin produced by the liver, kidneys, adipose tissue and neutrophils. Coded for by the *GC* gene on chromosome 4q, two SNPs in exon 11; rs7041 (Glu416Asp) and rs4588 (Thr420Lys) define the three major haplotypes of VTDB: *GC1F*: 416Asp/420Thr, *GC1S*: 416Glu/420Lys and *GC2*: 416Asp/420Lys with serum VTDB level related to these SNPs[21]. VTDB binds 25(OH)-vitamin D and $1,25(OH)_2$-vitamin D, although vitamin D levels are far exceeded by the transport capacity of VTDB. Serum levels of VTDB and vitamin D are unrelated in many diseases studied including Chronic Obstructive

Pulmonary Disease (COPD)[22]. The *GC* variants have differing affinities for vitamin D: the complexities of the VTDB isoforms, vitamin D and their impact on lung disease are not yet clear[23].

The mechanism relating *GC* genotype and serum VTDB is also unknown: rs7041 and rs4588 are intronic SNPs and neither are in linkage disequilibrium with known promotor or enhancer SNPs, nor are they known to affect protein stability. Factors other than *GC* genotype, including epigenetics, may also influence serum VTDB levels, as although serum VTDB is lower in women with LAM than controls, *GC* genotype in our study was not different.

Our findings reflect the complexity of both the VTDB axis and LAM. We observed that lower serum VTDB was associated with lower lung function and more active lung disease at presentation. As VTDB is not associated with other aspects of the LAM phenotype including the presence of angiomyolipoma or lymphatic disease it is likely that VTDB axis is not related to LAM per se, but as in other lung diseases may alter the tissue response to disease.

Importantly, *GC* genotype was associated with time to death or lung transplantation. The strongest effect being for the *GC1F* and *GC1S* haplotypes, which were associated with an increase in median survival of over three years. Interestingly, this and other *GC* variants associated with improved survival were not those associated with the lowest serum VTDB levels. VTDB is a multifunctional protein which may impact upon the response to lung damage in a number of ways. *GC1F* and *GC1S* are associated with increased macrophage activation over *GC2*[22] and increased macrophage activation may be protective in LAM, either by enhancing protective neutrophil responses or enhancing the chemotactic effect of complement-derived C5a[24][25]. VTDB also acts as an actin scavenging protein and therefore has the potential to influence disease by different mechanisms including altered

innate immunity and tissue repair. Different *GC* haplotypes are already associated with susceptibility to lung disease: *GC1F* being associated with an enhanced risk of COPD over *GC1S* and *GC2*[26].

These observations underscore the multiple potential functions of VTDB, how these functions may be related to genotype and the complex relationship with lung disease. The complexity of LAM, a multisystem disease, is also likely to be important. For example, VTDB protein is associated with DL_{CO} but not FEV_1, FVC or event-free survival. Whilst FEV_1 is generally used to study the natural history of LAM, DL_{CO} is usually impaired before FEV_1 and may better reflect early parenchymal damage in LAM with loss of FEV_1 occurring later due to loss of elastic recoil and premature airway closure brought about by parenchymal damage. Pulmonary vascular disease, host defence, peripheral muscle function and other processes potentially affected by VTDB function may also contribute to survival.

One of the strengths of our study was the use of an unbiased proteomic method that identified VTDB as a protein of interest in LAM. The involvement of the vitamin D axis in other diseases associated with tissue remodelling make our findings biologically plausible[27]. Our study also has limitations however, including the low number of control samples, technical limitations and those inherent in studying rare diseases. Firstly, VTDB was one of only two proteins differentially expressed in the serum of women with LAM and the proteomic methodology used did not identify other LAM markers such as VEGF-D. VEGF-D is expressed at picomolar levels[28], whereas VTDB is present a micromolar levels suggesting that only relatively abundant serum proteins with robust differences between women with LAM and healthy controls could be detected using this proteomic strategy. It is therefore likely that other potentially useful biomarkers remain undiscovered. Consistent with this, A1AG1, also known as orosomucoid, the other protein linked to the presence of LAM in our

proteomic screen, is another relatively abundant plasma alpha globulin, comprising 1-3% of plasma proteins. As A1AG1 is an acute phase protein, already recognised as a biomarker of overall survival in many populations we did not study it further[29]. As LAM is very rare, studying the disease relies upon cohorts accumulated over longer periods of time. Although both cohorts studied used protocol driven assessments to capture key data including lung function, there are some differences in the data available for these groups. Although the two cohorts used were similar in terms of age and lung function, prospective change in lung function differed, probably due to the use of rapamycin in the UK cohort resulting in reduced loss of FEV_1. Conversely, due to time of recruitment, long term survival prior to Rapamycin use can now only be studied in the NHLBI registry cohort. Current individuals with progressive disease, including those in the UK cohort studied here, tend to be treated with rapamycin[30] and longer periods of observation are needed to study the effect of the VTDB protein or genotype on survival in women with LAM treated with rapamycin.

In conclusion, low levels of VTDB are associated with poor lung function in LAM and *GC* genotypes are associated with long-term outcome. Our findings suggest that the VTDB axis is a host factor that may protect against lung damage in LAM and could be of prognostic significance. Further studies are required to validate our findings and understand how the VTDB isoforms modulate lung damage in LAM and other diseases.

Acknowledgements: We are grateful to the investigators who contributed to the NHLBI LAM Registry, and to the women with LAM who contributed to the research. This research used the ALICE High Performance Computing Facility at the University of Leicester. LVW holds a GSK /British Lung Foundation Chair in Respiratory Research. MDT holds a Wellcome

Trust Investigator Award (WT 202849/Z/16/Z). This research has been performed using the UK Biobank Resource under application 648.

Funding: This study was funded by the National Institute for Health Research Rare Disease Translational Research Collaboration, a LAM Foundation Biomarker Innovation Grant and a LAM Foundation project grant.

References

1. Harknett EC, Chang WYC, Byrnes S, Johnson J, Lazor R, Cohen MM, Gray B, Geiling S, Telford H, Tattersfield AE, Hubbard RB, Johnson SR. Regional and National Variability Suggests Underestimation of Prevalence of Lymphangioleiomyomatosis. *Quarterly Journal of Medicine* 2011: 104(11): 971-979.
2. Johnson SR. Lymphangioleiomyomatosis. *The European respiratory journal* 2006: 27(5): 1056-1065.
3. Johnson SR, Whale CI, Hubbard RB, Lewis SA, Tattersfield AE. Survival and disease progression in UK patients with lymphangioleiomyomatosis. *Thorax* 2004: 59(9): 800-803.
4. Pea M, Martignoni G, Bonetti F, Zamboni G, Colombari C, Manfrin E, Lestani M, Mombello A, Mariuzzi GM. Tumors characterized by the presence of HMB45-positive perivascular epithelioid cell (PEC) - A novel entity in surgical pathology. *Electronic Journal of Pathology & Histology* 1997: 3(2): 28-40.
5. Carsillo T, Astrinidis A, Henske EP. Mutations in the tuberous sclerosis complex gene TSC2 are a cause of sporadic pulmonary lymphangioleiomyomatosis. *Proc Natl Acad Sci U S A* 2000: 97(11): 6085-6090.

6. Saxton RA, Sabatini DM. mTOR Signaling in Growth, Metabolism, and Disease. *Cell* 2017: 169(2): 361-371.
7. Johnson SR, Tattersfield AE. Decline in lung function in lymphangioleiomyomatosis: relation to menopause and progesterone treatment. *Am J Respir Crit Care Med* 1999: 160(2): 628-633.
8. McCormack FX, Inoue Y, Moss J, Singer LG, Strange C, Nakata K, Barker AF, Chapman JT, Brantly ML, Stocks JM, Brown KK, Lynch JP, Goldberg HJ, Young LR, Kinder BW, Downey GP, Sullivan EJ, Colby TV, McKay RT, Cohen MM, Korbee L, Taveira-DaSilva AM, Lee H-S, Krischer JP, Trapnell BC. Efficacy and Safety of Sirolimus in Lymphangioleiomyomatosis. *New England Journal of Medicine* 2011: 364: 1595-1606.
9. Bee J, Bhatt R, McCafferty I, Johnson S. Audit, research and guideline update: A 4-year prospective evaluation of protocols to improve clinical outcomes for patients with lymphangioleiomyomatosis in a national clinical centre. *Thorax* 2015.
10. Bee J, Fuller S, Miller S, Johnson SR. Lung function response and side effects to rapamycin for lymphangioleiomyomatosis: a prospective national cohort study. *Thorax* 2017.
11. Young LR, VanDyke R, Gulleman PM, Inoue Y, Brown KK, Schmidt LS, Linehan WM, Hajjar F, Kinder BW, Trapnell BC, Bissler JJ, Franz DN, McCormack FX. Serum Vascular Endothelial Growth Factor-D Prospectively Distinguishes Lymphangioleiomyomatosis From Other Diseases. *Chest* 2010: 138(3): 674-681.
12. Oprescu N, McCormack FX, Byrnes S, Kinder BW. Clinical Predictors of Mortality and Cause of Death in Lymphangioleiomyomatosis: A Population-based Registry. *Lung* 2013: 191(1): 35-42.
13. Cohen MM, Pollock-BarZiv S, Johnson SR. Emerging clinical picture of lymphangioleiomyomatosis. *Thorax* 2005: 60(10): 875-879.
14. Taveira-DaSilva AM, Steagall WK, Rabel A, Hathaway O, Harari S, Cassandro R, Stylianou M, Moss J. Reversible airflow obstruction in lymphangioleiomyomatosis. *CHEST Journal* 2009: 136(6): 1596-1603.

15. Taveira-DaSilva AM, Stylianou MP, Hedin CJ, Hathaway O, Moss J. Decline in Lung Function in Patients With Lymphangioleiomyomatosis Treated With or Without Progesterone. *Chest* 2004: 126(6): 1867-1874.

16. Ryu JH, Moss J, Beck GJ, Lee J-C, Brown KK, Chapman JT, Finlay GA, Olson EJ, Ruoss SJ, Maurer JR, Raffin TA, Peavy HH, McCarthy K, Taveira-DaSilva A, McCormack FX, Avila NA, DeCastro RM, Jacobs SS, Stylianou M, Fanburg BL, for the NHLBI LAM Registry Group. The NHLBI Lymphangioleiomyomatosis Registry: Characteristics of 230 Patients at Enrollment. *Am J Respir Crit Care Med* 2006: 173(1): 105-111.

17. Sajic T, Liu Y, Aebersold R. Using data-independent, high-resolution mass spectrometry in protein biomarker research: Perspectives and clinical applications. *PROTEOMICS – Clinical Applications* 2015: 9(3-4): 307-321.

18. Lambert J-P, Ivosev G, Couzens AL, Larsen B, Taipale M, Lin Z-Y, Zhong Q, Lindquist S, Vidal M, Aebersold R, Pawson T, Bonner R, Tate S, Gingras A-C. Mapping differential interactomes by affinity purification coupled with data-independent mass spectrometry acquisition. *Nature Methods* 2013: 10: 1239.

19. Shrine N, Guyatt AL, Erzurumluoglu AM, Jackson VE, Hobbs BD, Melbourne C, Batini C, Fawcett KA, Song K, Sakornsakolpat P, Li X, Boxall R, Reeve NF, Obeidat M, en, Zhao JH, Wielscher M, Weiss S, Kentistou KA, Cook JP, Sun BB, Zhou J, Hui J, Karrasch S, Imboden M, Harris SE, Marten J, Enroth S, Kerr SM, Surakka I, Vitart V, Lehtimäki T, Allen RJ, Bakke PS, Beaty TH, Bleecker ER, Bossé Y, Brandsma C-A, Chen Z, Crapo JD, Danesh J, DeMeo DL, Dudbridge F, Ewert R, Gieger C, Gulsvik A, Hansell AL, Hao K, Hoffman JD, Hokanson J, Homuth G, Joshi PK, Joubert P, Langenberg C, Li X, Li L, Lin K, Lind L, Locantore N, Luan J, an, Mahajan A, Maranville JC, Murray A, Nickle DC, Packer R, Parker MM, Paynton ML, Porteous D, Prokopenko D, Qiao D, Rawal R, Runz H, Sayers I, Sin DD, Smith BH, Soler Artigas M, Sparrow D, Tal-Singer R, Timmers PRHJ, Van den Berge M, Whittaker JC, Woodruff P, Yerges Armstrong LM, Troyanskaya OG, Raitakari OT, Kähönen M, Polasek O, Gyllensten U, Rudan I, Deary IJ, Probst-Hensch NM, Schulz H, James AL,

Wilson JF, Stubbe B, Zeggini E, Jarvelin M-R, Wareham N, Silverman EK, Hayward C, Morris AP, Butterworth AS, Scott RA, Walters RG, Meyers DA, Cho MH, Strachan DP, Hall IP, Tobin MD, Wain LV. New genetic signals for lung function highlight pathways and pleiotropy, and chronic obstructive pulmonary disease associations across multiple ancestries. *bioRxiv* 2018.

20. Wain LV, Shrine N, Miller S, Jackson VE, Ntalla I, Artigas MS, Billington CK, Kheirallah AK, Allen R, Cook JP, Probert K, Obeidat Me, Bossé Y, Hao K, Postma DS, Paré PD, Ramasamy A, Mägi R, Mihailov E, Reinmaa E, Melén E, O'Connell J, Frangou E, Delaneau O, Freeman C, Petkova D, McCarthy M, Sayers I, Deloukas P, Hubbard R, Pavord I, Hansell AL, Thomson NC, Zeggini E, Morris AP, Marchini J, Strachan DP, Tobin MD, Hall IP. Novel insights into the genetics of smoking behaviour, lung function, and chronic obstructive pulmonary disease (UK BiLEVE): a genetic association study in UK Biobank. *The Lancet Respiratory Medicine* 2015: 3(10): 769-781.
21. Moy KA, Mondul AM, Zhang H, Weinstein SJ, Wheeler W, Chung CC, Männistö S, Yu K, Chanock SJ, Albanes D. Genome-wide association study of circulating vitamin D–binding protein. *The American Journal of Clinical Nutrition* 2014: 99(6): 1424-1431.
22. Wood AM, Bassford C, Webster D, Newby P, Rajesh P, Stockley RA, Thickett DR. Vitamin D-binding protein contributes to COPD by activation of alveolar macrophages. *Thorax* 2011: 66(3): 205-210.
23. Arnaud J, Constans J. Affinity differences for vitamin D metabolites associated with the genetic isoforms of the human serum carrier protein (DBP). *Human Genetics* 1993: 92(2): 183-188.
24. Zhang J, Kew RR. Identification of a Region in the Vitamin D-binding Protein that Mediates Its C5a Chemotactic Cofactor Function. *Journal of Biological Chemistry* 2004: 279(51): 53282-53287.
25. Yamamoto N, Homma S. Vitamin D3 binding protein (group-specific component) is a precursor for the macrophage-activating signal factor from lysophosphatidylcholine-treated lymphocytes. *Proceedings of the National Academy of Sciences* 1991: 88(19): 8539.

26. Horita N, Miyazawa N, Tomaru K, Inoue M, Ishigatsubo Y, Kaneko T. Vitamin D binding protein genotype variants and risk of chronic obstructive pulmonary disease: A meta-analysis. *Respirology* 2015: 20(2): 219-225.
27. Chishimba L, Thickett DR, Stockley RA, Wood AM. The vitamin D axis in the lung: a key role for vitamin D-binding protein. *Thorax* 2010: 65(5): 456-462.
28. Young LR, Lee H-S, Inoue Y, Moss J, Singer LG, Strange C, Nakata K, Barker AF, Chapman JT, Brantly ML, Stocks JM, Brown KK, Lynch JP, Goldberg HJ, Downey GP, Swigris JJ, Taveira-DaSilva AM, Krischer JP, Trapnell BC, McCormack FX. Serum VEGF-D concentration as a biomarker of lymphangioleiomyomatosis severity and treatment response: a prospective analysis of the Multicenter International Lymphangioleiomyomatosis Efficacy of Sirolimus (MILES) trial. *The Lancet Respiratory Medicine* 2013.
29. Fischer K, Kettunen J, Würtz P, Haller T, Havulinna AS, Kangas AJ, Soininen P, Esko T, Tammesoo M-L, Mägi R, Smit S, Palotie A, Ripatti S, Salomaa V, Ala-Korpela M, Perola M, Metspalu A. Biomarker Profiling by Nuclear Magnetic Resonance Spectroscopy for the Prediction of All-Cause Mortality: An Observational Study of 17,345 Persons. *PLoS Med* 2014: 11(2): e1001606.
30. Bee J, Fuller S, Miller S, Johnson SR. Lung function response and side effects to rapamycin for lymphangioleiomyomatosis: a prospective national cohort study. *Thorax* 2018: 73(4): 369.

Table 1. Clinical data for cohorts studied.

	UK Discovery cohort			UK Replication cohort	UK Rapamycin treated	USA NHLBI cohort	Healthy controls
	All	Stable	Progressive				
n	50	26	24	27	24	152	22
Age (years)*	50.6 ± 10.9	50.9 ± 11.8	50.3 ± 10.0	49.4 ± 13.9	46.4 ± 9.7	45.4 ± 9.0	35.0 ± 11.7
Disease duration (years)*	13.9 ± 11.1	14.2 ± 11.4	13.5 ± 11.1	9.1 ± 9.5	13.1 ± 9.5	4.6 ± 4.3	N/A
Angiomyolipoma†	72	77	67	55	54	N/T	N/A
Lymphatic disease†	16	15	17	23	25	N/T	N/A
TSC†	14	19	8	15	21	N/T	N/A
Pneumothorax†	48	50	46	40	46	N/T	N/A
Post-menopause†	34	42	25	30	25	48	N/A
FEV_1 (% predicted)*	68.9 ± 20.6	76.4 ± 18.9	60.8 ± 19.5	77.4 ± 23.4	46.7 ± 14.8	74.1 ± 27.5	N/A
DL_{CO} (% predicted)*	59.8 ± 15.8	68.9 ± 12.7	50.0 ± 12.9	62.9 ± 17.1	43.3 ± 12.3	55.7 ± 25.6	N/A
VEGF-D (pg/ml)*	1327 ± 1187	985 ± 833	1698 ± 1405	1275 ± 1527	1082 ± 1257	N/T	397 ± 125

*: mean ± standard deviation at recruitment. †: present at any time in disease course (%).

N/A: not applicable. N/T: not available for testing. Disease duration in the UK LAM cohort was first symptom to enrolment whilst in the NHLBI cohort disease duration was from diagnosis to enrolment. In the NHLBI cohort, menopause was assumed if ≥ 50 years of age.

Table 2. Prospective change in FEV_1 and DL_{CO} and relationship to VTDB

	UK				NHLBI	
	untreated	**p**	**rapamycin**	**p**	**untreated**	**p**
n	64		27		136	
ΔFEV_1 (ml/yr)*	-32.6 ± 111.2	n/s	24.3 ± 141.4	n/s	-94.7 ± 96.2	n/s
ΔDL_{CO} (mmol/min/kPa/yr)*	-0.2 ± 0.40	n/s	-0.17 ± 0.23	n/s	-0.23 ± 0.31	n/s
VTDB (µg/ml)*	273 ± 96	-	281 ± 105	-	255 ± 53.4	-

* mean (standard deviation). p value for Spearman's correlation with serum VTDB.

Table 3. Relationship of VTDB genotype with clinical features, serum VTDB and change in lung function in the NHLBI LAM Registry cohort

SNP	**rs4588**				**rs7041**			
Genotype	**AA**	**CA**	**CC**	**p**	**TT**	**GT**	**GG**	**p**
n	11	46	74		25	57	48	
Age at diagnosis	37.4 (6.7)	42.1 (9.9)	40.6 (9.2)	n/s	39.9 (7.6)	41.8 (9.7)	40.5 (9.6)	n/s
Age at recruitment	40.9 (6.4)	47.2 (9.4)	45.3 (8.9)	-	43.8 (7.5)	46.4 (9.4)	45.4 (9.3)	-
FEV_1 (% predicted)	88.0 (21.0)	78.8 (25.2)	72.9 (29.4)	n/s	79.9 (26.8)	79.0 (30.2)	72.0 (24.0)	n/s
DL_{CO} (% predicted)	58.3 (17.5)	59.5 (22.5)	57.0 (29.3)	n/s	56.3 (18.2)	58.1 (30.8)	58.9 (23.2)	n/s
VTDB (µg/ml)	220 (36)	245 (57)	266 (52)	0.022	233 (43)	250 (57)	270 (53)	0.026
ΔFEV_1 (ml/yr)	-125 (142)	-78 (81)	-99 (97)	n/s	-135 (126)	-80 (84)	-94 (94)	n/s
ΔDL_{CO} (mmol/min/kPa/yr)	-0.35 (0.23)	-0.21 (0.36)	-0.22 (0.3)	n/s	-0.26 (0.27)	-0.20 (0.35)	-0.26 (0.27)	n/s

Mean (and standard deviation) are shown for women with LAM of European ancestry. Data for percent predicted FEV_1, DL_{CO}, VTDB and age at recruitment were all at entry to the study. ΔFEV_1 and ΔDL_{CO} are prospective changes from recruitment. Linear regression was used to model the relationship between genotype, clinical factors and VTDB.

Figure legends:

Figure 1. Enrolment and samples tested. Recruitment and access to samples and lung function data in the UK and NHLBI LAM cohorts. The UK discovery cohort consisted of 50 serum samples from individuals with LAM, the UK replication cohort comprised 27 LAM serum samples and the USA NHLBI LAM cohort 152. PFT: pulmonary function test.

Figure 2. Serum VTDB and A1AG1 in LAM and healthy controls. (a) Women with LAM had lower levels of serum Vitamin D Binding Protein (VTDB) compared with healthy control women (p=0.002). (b) Women with LAM had higher levels of serum Alpha-1-Acid Glycoprotein (A1AG1) compared with healthy control women (p=0.004).

Figure 3. VTDB is associated with disease severity. (a) Lower levels of serum VTDB were associated with progressive, compared with stable LAM (p=0.001). (b) VTDB level is positively correlated with percent predicted DL_{CO} (p=0.01). (c) VTDB was not associated with percent predicted FVC (p=0.09) or (d) percent predicted FEV_1 (p=0.23).

Figure 4. Survival analysis for VTDB level and *GC* genotype in the NHLBI LAM Registry cohort. (a) Overall time to death or transplant did not differ with serum VTDB level (low (147 - 221 µg/ml), medium (222 - 275 µg/ml) and high (276 - 413 µg/ml. (p=0.76). (b) Individuals with the AA or AC genotype at rs4588 had greater time to death or transplant

than those with the genotype CC (p=0.014). (c) Haplotypes with an A allele at rs4588 (*GC1F* and *GC1S*) were associated with longer time to death or transplant (p=0.008).

Figures:

Figure 1.

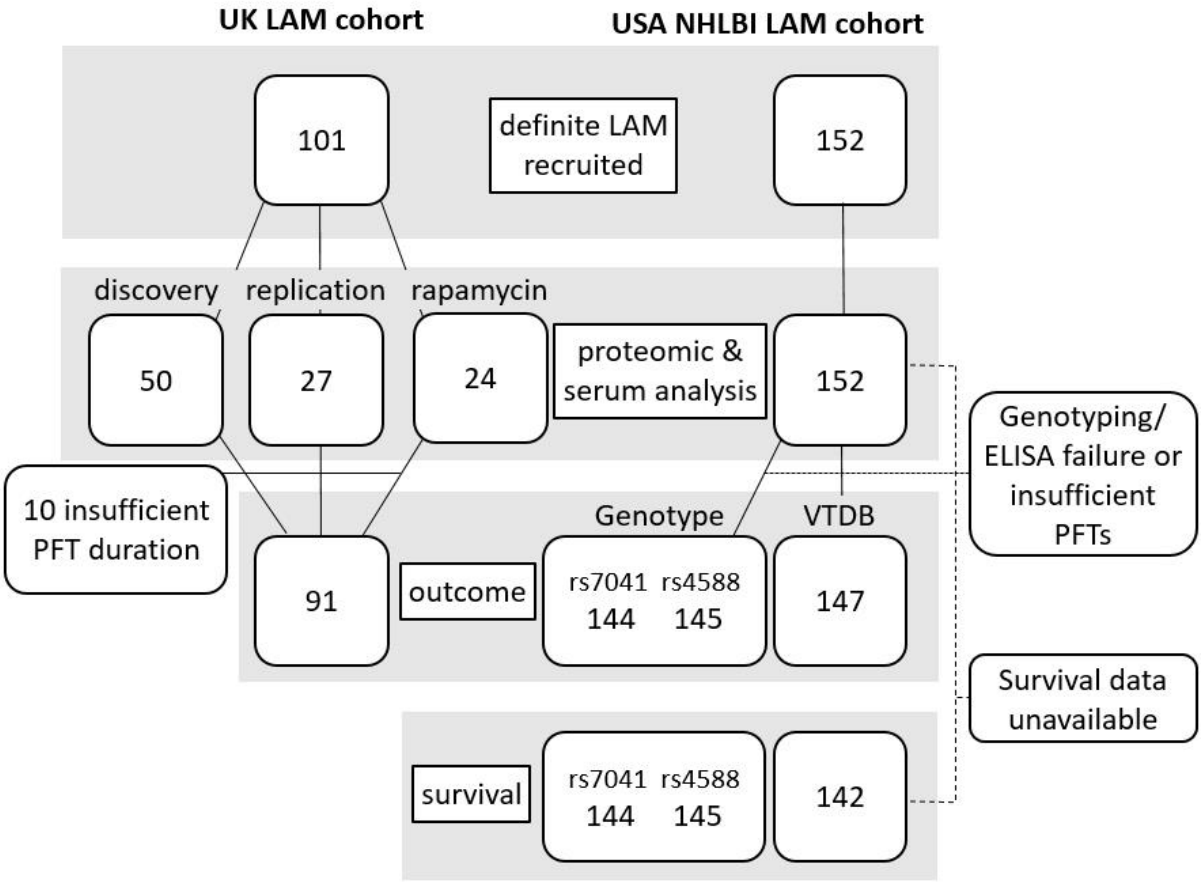


Figure 2.

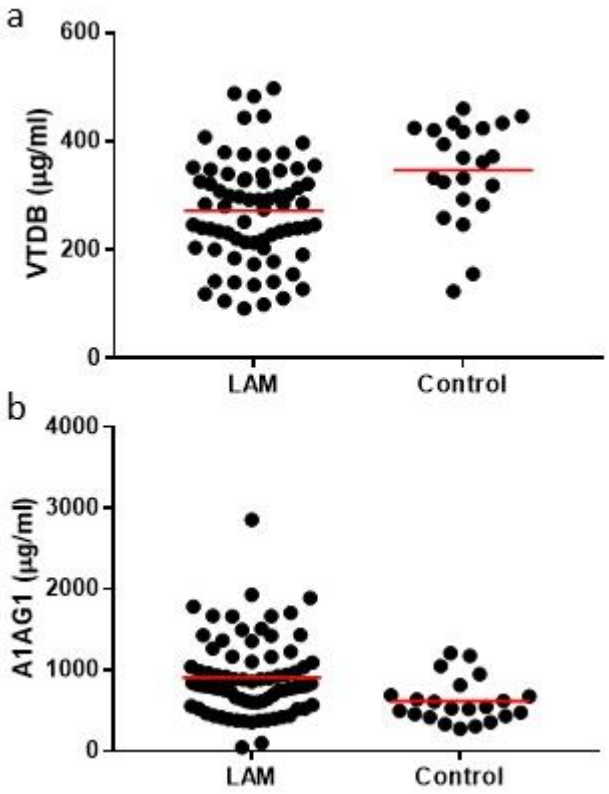

a
600
400
200
0
VTDB (µg/ml)
LAM
Control
b
4000
3000
2000
1000
0
A1AG1 (µg/ml)
LAM
Control

Figure 3.

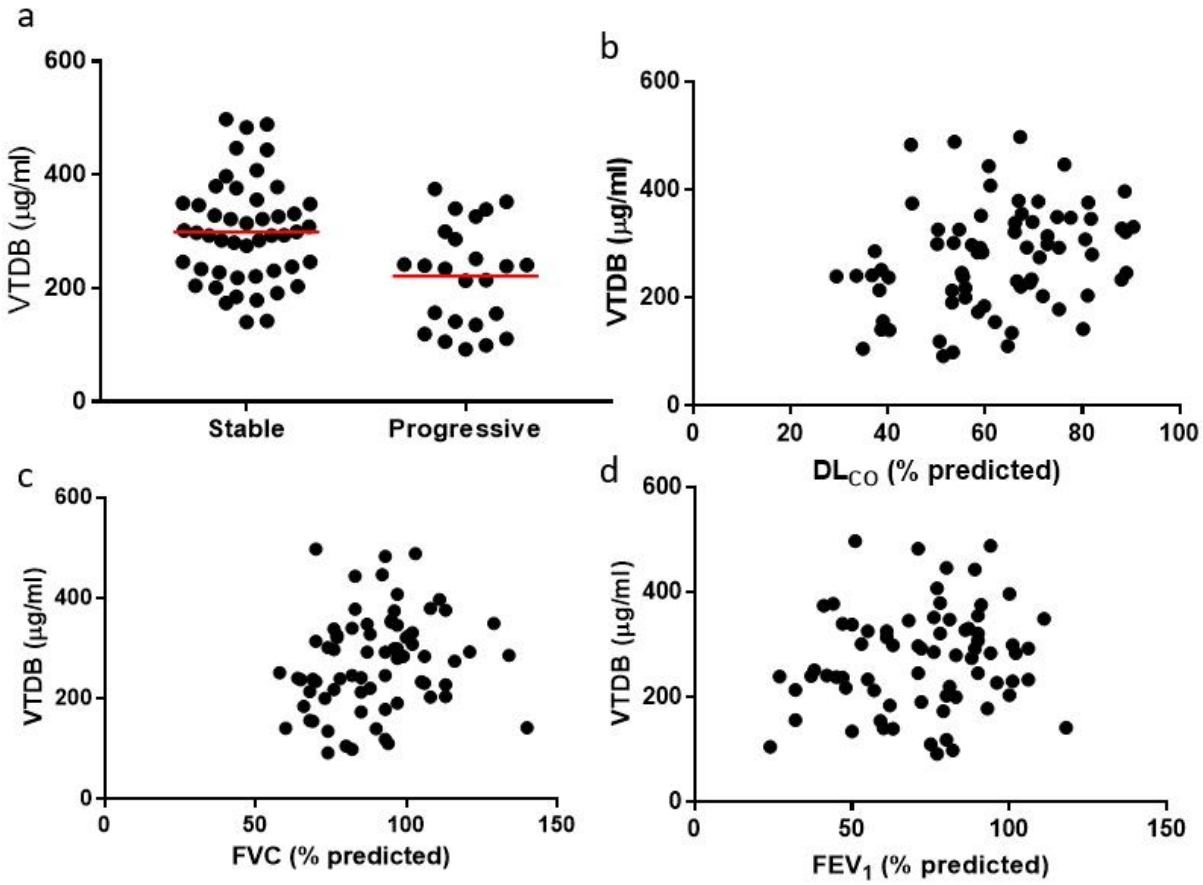
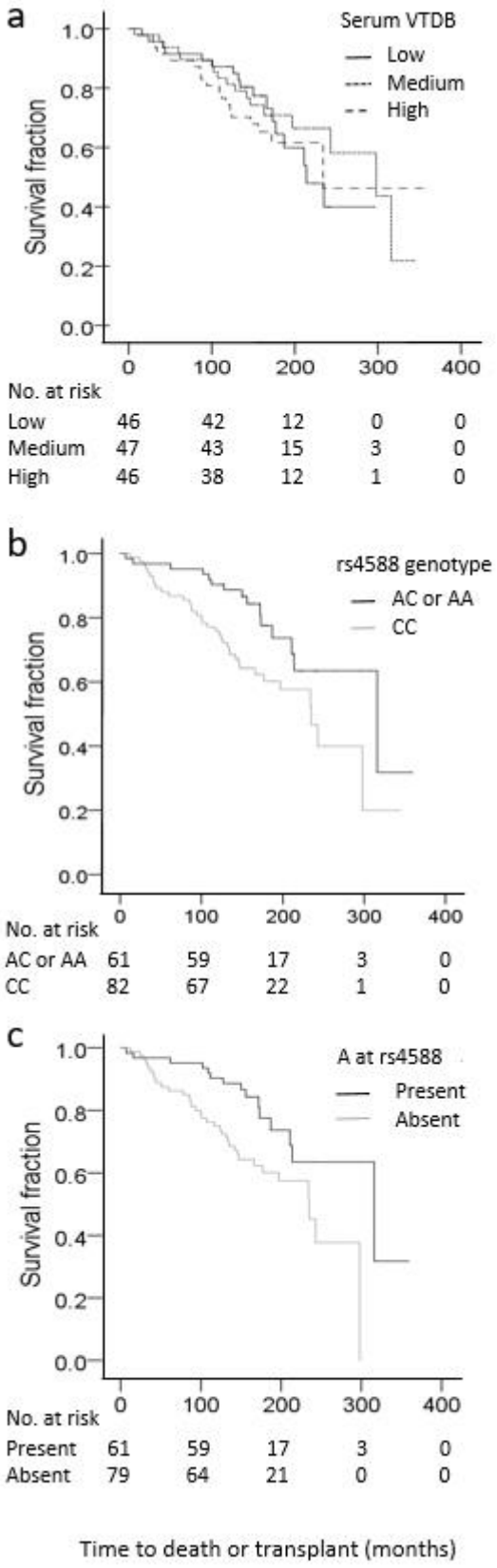

a
VTDB (μg/ml)
600
400
200
0
Stable
Progressive
b
VTDB (μg/ml)
600
400
200
0
0
20
40
60
80
100
DLCO (% predicted)
c
VTDB (μg/ml)
600
400
200
0
0
50
100
150
FVC (% predicted)
d
VTDB (μg/ml)
600
400
200
0
0
50
100
150
FEV1 (% predicted)

Figure 4.


a
Serum VTDB
Low
Medium
High
Survival fraction
1.0
0.8
0.6
0.4
0.2
0.0
0
100
200
300
400
No. at risk
Low 46 42 12 0 0
Medium 47 43 15 3 0
High 46 38 12 1 0
b
rs4588 genotype
AC or AA
CC
Survival fraction
No. at risk
AC or AA 61 59 17 3 0
CC 82 67 22 1 0
c
A at rs4588
Present
Absent
Survival fraction
No. at risk
Present 61 59 17 3 0
Absent 79 64 21 0 0
Time to death or transplant (months)

Supplementary Methods

Patients and sample collection

101 women with LAM and 22 healthy control women were recruited between 2011 and 2016 from the National Centre for LAM, Nottingham, UK, hence-forth termed the UK cohort. Ethical approval was obtained from the East Midlands Research Ethics Committee (13/EM/0264). For healthy controls, age and ethnicity were recorded. All subjects provided written informed consent. Clinical history, presence of TSC, angiomyolipoma, lymphatic disease, menopausal status and drug treatment were recorded. Lymphatic disease was defined as the presence of chylous collections in the chest or abdomen, lymphangioleiomyomas or lymphadenopathy due to LAM visible on CT scanning of the chest abdomen and pelvis. Disease duration was calculated as the time from first symptom attributable to LAM as previously described[7]. Blood samples were taken at enrolment and processed within one hour of phlebotomy. Whole blood collected in serum separator tubes were allowed to clot for 1 hour at room temperature and separated by centrifugation at 1000 g for 10 minutes. Blood and serum samples were stored at -80 °C until analysis.

A second cohort of 243 subjects recruited between 1998 and 2001 in the National Heart Lung and Blood Institute (NHLBI) LAM Registry to study the natural history of LAM was used as a replication cohort and to study long term survival. This cohort has been described in detail previously[18]. Serum and DNA samples at enrolment, available from 152 of these 243 subjects, along with clinical and prospective lung function data were obtained from the National Disease Research Interchange who now curate the resource. Outcome data, either all-cause mortality or the need for lung transplant in the period following baseline assessment, were obtained by querying the United States National Death Index and the United Network for Organ Sharing databases. As data on the use of rapamycin was not available for this cohort, outcome data were censored at 2010 before rapamycin was widely used for the treatment of LAM in the USA. Suitable samples for serum protein measurement, genotyping or lung function over greater than one year's duration was not available in all subjects (Figure 1). The numbers included in individual analyses are stated in the results.

Lung function was measured at either Nottingham University Hospitals NHS Trust or the referring centre in the USA according to ERS/ATS standards. Prospective change in lung function was calculated as the difference between FEV_1 and DL_{CO} measured at recruitment and last follow up visit expressed in ml/year for FEV_1 (ΔFEV_1) and as ml/min/kPa/yr for DL_{CO} (ΔDL_{CO}). To reduce variation in measurement of disease progression, only values spanning one year or longer were used for this analysis[7]. Classification into stable or progressive disease at presentation was performed by calculating retrospective loss of FEV_1 until the time of enrolment between the first recorded FEV_1 value and the FEV_1 at study enrolment divided by the time interval and expressed in ml/year. Those with a retrospective ΔFEV_1 of less than -50 ml/yr were arbitrarily classified as more stable and greater than -50 ml/yr as more progressive.

Fifty women with LAM from the UK cohort who had not been treated with rapamycin formed the initial proteomic discovery population. A further 27 untreated patients from the UK were used as a replication cohort and 24 who were receiving rapamycin for treatment of LAM at

recruitment were also studied (Table 1). The discovery cohort was subdivided into those with stable and more progressive disease based upon retrospective loss of lung function.

Proteomics

Serum was diluted and filtered at 2 µm to remove any particulates to a final 1 in 45 dilution in 100 mM Triethylammonium bicarbonate buffer. An alkylation and reduction step adding 2µL 0.5 mM Dithiothreitol (DTT) with 45 min shaking at 56 °C followed by 7.15 µL of 140 mM Iodoacetamide and 30 min incubation in darkness at room temperature. The reaction was then quenched using 1.95 µL of 0.5 mM DTT. Samples were digested with 1.95 µL of 1 µg/µL trypsin (T656720UG, Sigma, UK) over 17 hrs at 37°C while shaking/agitating after which samples were lyophilised in a speed vac and re-suspended at decreasing concentrations of acetonitrile (ACN) to a final mixture of 40 µL and 5% ACN 0.1% formic acid (FA). After high speed centrifugation, supernatants were transferred to appropriate tubes for mass spectrometric analysis. The Biognosis HRM retention time standard was added for downstream alignment. Samples were analysed on a SCIEX TripleTOF 6600 mass spectrometer hyphenated to an Eksigent nanoLC 425 system operating in micro flow (5 µL/min). The SCIEX SWATH mass spectrometric workflow[17] was utilised for relative protein quantitation, wherein data acquired from a quantitative data independent (DIA) SWATH, are assembled against libraries of protein identified using a data dependent acquisition. Chromatographic separation for protein identification (Information Dependent Acquisition/IDA) was over an 87 min gradient, 4 µL direct injection on a YMC 25 cm x 0.3mm Triart-C18 column (12nm, 3µm particle size) with a gradient of 3 % mobile phase B (2% acetonitrile, 5% DMSO in 0.1% FA) to 30 % over 38 min; to 40% B at 73 min, 80 % B at 75 min, held for 3 min then returned to 3 % over 1 min. Chromatographic separation for SWATH runs was conducted as above but on a 57 min gradient of 3 % mobile phase B (2% ACN, 5% DMSO in 0.1% FA) to 30 % over 38 min; to 40 % B at 43 min, 80 % B at 45 min held for 3 min then returned to 3 % over 1 min. The mass spectrometer set up and method settings consisted of a Duospray™ source (SCIEX) with a 50 µm electrode at +5500V (gas settings GS1 15; GS2 0; CUR 25; TEMP 0). IDA was carried out using parameters of Top 30 (TOFMS 250 ms accumulation time, production 60 ms, total cycle time 2.1 s); charge state 2 - 4 above a threshold of 200 cps; dynamic exclusion for 10 seconds using rolling collision energy (optimised for m/z of target ion). SWATH methods consisted of 100 variable windows optimised for serum and cell lysate proteins. MS/MS spectra were searched using ProteinPilot 5.0 (SCIEX) with the Swissprot human database (Jan 2015) at 1 % false discovery rate with an identification focus on biological modifications. SWATH data were aligned to the library files in PeakView (SCIEX) and uploaded to the SCIEX OneOmics platform for processing, compilation, assembly and annotation of SWATH data.

Supplementary results

Supplementary table E1. Serum proteins identified by proteomic screen in LAM and control serum.

Protein Name	UniProt ID	Full name
A1AG1	P02763	Alpha-1-acid glycoprotein 1
A1AT	P01009	Alpha-1-antitrypsin
A1BG	P04217	Alpha-1B-glycoprotein
A2GL	P02750	Leucine-rich alpha-2-glycoprotein
A2MG	P01023	Alpha-2-macroglobulin
ACTG	P63261	Actin, cytoplasmic 2
AFAM	P43652	Afamin
ALBU	P02768	Serum albumin
AMBP	P02760	Protein AMBP
ANGT	P01019	Angiotensinogen
ANT3	P01008	Antithrombin-III
APOA	P08519	Apolipoprotein(a)
APOA1	P02647	Apolipoprotein A-I
APOA2	P02652	Apolipoprotein A-II
APOA4	P06727	Apolipoprotein A-IV
APOB	P04114	Apolipoprotein B-100
APOC2	P02655	Apolipoprotein C-II
APOC3	P02656	Apolipoprotein C-III
APOD	P05090	Apolipoprotein D
APOE	P02649	Apolipoprotein E
APOF	Q13790	Apolipoprotein F
APOH	P02749	Beta-2-glycoprotein 1
APOL1	O14791	Apolipoprotein L1
APOM	O95445	Apolipoprotein M
C1QC	P02747	Complement C1q subcomponent subunit C
C1R	P00736	Complement C1r subcomponent
C1S	P09871	Complement C1s subcomponent
C4BPA	P04003	C4b-binding protein alpha chain
CAMP	P49913	Cathelicidin antimicrobial peptide
CBPN	P15169	Carboxypeptidase N catalytic chain
CD44	P16070	CD44 antigen
CD5L	O43866	CD5 antigen-like
CERU	P00450	Ceruloplasmin
CFAB	P00751	Complement factor B

CFAH	P08603	Complement factor H
CFAI	P05156	Complement factor I
CLUS	P10909	Clusterin
CO2	P06681	Complement C2
CO3	P01024	Complement C3
CO4B	P0C0L5	Complement C4-B
CO5	P01031	Complement C5
CO6	P13671	Complement component C6
CO8A	P07357	Complement component C8 alpha chain
CO8G	P07360	Complement component C8 gamma chain
CO9	P02748	Complement component C9
CXCL7	P02775	Platelet basic protein
FA12	P00748	Coagulation factor XII
FBLN1	P23142	Fibulin-1
FCG3A	P08637	Low affinity immunoglobulin gamma Fc region receptor III-A
FCN2	Q15485	Ficolin-2
FETUA	P02765	Alpha-2-HS-glycoprotein
FHR3	Q02985	Complement factor H-related protein 3
FHR5	Q9BXR6	Complement factor H-related protein 5
FIBA	P02671	Fibrinogen alpha chain
FINC	P02751	Fibronectin
FOXN3	O00409	Forkhead box protein N3
GELS	P06396	Gelsolin
H2AX	P16104	Histone H2AX
HBA	P69905	Hemoglobin subunit alpha
HBB	P68871	Hemoglobin subunit beta
HDAC1	Q13547	Histone deacetylase 1
HEMO	P02790	Hemopexin
HMMR	O75330	Hyaluronan mediated motility receptor
HPT	P00738	Haptoglobin
HPTR	P00739	Haptoglobin-related protein
HRG	P04196	Histidine-rich glycoprotein
HS12B	Q96MM6	Heat shock 70 kDa protein 12B
HS90B	P08238	Heat shock protein HSP 90-beta
HV101	P01742	Immunoglobulin heavy variable 1-69
HV304	P01765	Immunoglobulin heavy variable 3-23
HV305	P01766	Immunoglobulin heavy variable 3-13
HV306	P01767	Immunoglobulin heavy variable 3-53
HV311	P01772	Immunoglobulin heavy variable 3-33
IBP3	P17936	Insulin-like growth factor-binding protein 3
IC1	P05155	Plasma protease C1 inhibitor
IGHA1	P01876	Ig alpha-1 chain C region
IGHA2	P01877	Ig alpha-2 chain C region
IGHD	P01880	Ig delta chain C region
IGHG1	P01857	Ig gamma-1 chain C region

IGHG2	P01859	Ig gamma-2 chain C region
IGHG3	P01860	Ig gamma-3 chain C region
IGHM	P01871	Ig mu chain C region
IGJ	P01591	Immunoglobulin J chain
IGKC	P01834	Ig kappa chain C region
IGLL5	B9A064	Immunoglobulin lambda-like polypeptide 5
ITIH1	P19827	Inter-alpha-trypsin inhibitor heavy chain H1
ITIH2	P19823	Inter-alpha-trypsin inhibitor heavy chain H2
ITIH4	Q14624	Inter-alpha-trypsin inhibitor heavy chain H4
K1024	Q9UPX6	UPF0258 protein KIAA1024
K1C19	P08727	Keratin, type I cytoskeletal 19
KANK3	Q6NY19	KN motif and ankyrin repeat domain-containing protein 3
KCNC4	Q03721	Potassium voltage-gated channel subfamily C member 4
KI67	P46013	Proliferation marker protein Ki-67
KIFC2	Q96AC6	Kinesin-like protein KIFC2
KLKB1	P03952	Plasma kallikrein
KNG1	P01042	Kininogen-1
KV102	P01594	Immunoglobulin kappa variable 1-33
KV106	P01598	Immunoglobulin kappa variable 1-5
KV305	P01623	Immunoglobulin kappa variable 3-20
KV308	P04207	Immunoglobulin kappa variable 3-15
KV309	P04433	Immunoglobulin kappa variable 3-11
KV404	P06314	Immunoglobulin kappa variable 4-1
LAC2	P0CG05	Ig lambda-2 chain C regions
LG3BP	Q08380	Galectin-3-binding protein
LIPB2	Q8ND30	Liprin-beta-2
LV106	P04208	Immunoglobulin lambda variable 1-47
LV302	P80748	Immunoglobulin lambda variable 3-21
LV403	P01717	Immunoglobulin lambda variable 3-25
PEDF	P36955	Pigment epithelium-derived factor
PGRP2	Q96PD5	N-acetylmuramoyl-L-alanine amidase
PK3CG	P48736	Phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit gamma isoform
PLF4	P02776	Platelet factor 4
PLMN	P00747	Plasminogen
PON1	P27169	Serum paraoxonase/arylesterase 1
PROP	P27918	Properdin
PROS	P07225	Vitamin K-dependent protein S
RET4	P02753	Retinol-binding protein 4
SAA4	P35542	Serum amyloid A-4 protein
SHBG	P04278	Sex hormone-binding globulin
SHC1	P29353	SHC-transforming protein 1
THRB	P00734	Prothrombin
TRFE	P02787	Serotransferrin
TRPV2	Q9Y5S1	Transient receptor potential cation channel subfamily V member 2

TSP1	P07996	Thrombospondin-1
VTDB	P02774	Vitamin D-binding protein
VTNC	P04004	Vitronectin

Supplementary table E2. Comparison of protein expression between women with LAM and control women.

		LAM vs control	
Protein	UniProt ID	Fold change (Log2)	Confidence
VTDB	P02774	-2.6	0.65
ITIH4	Q14624	-0.5	0.34
HMMR	O75330	0.5	0.31
FETUA	P02765	-2.8	0.30
AMBP	P02760	-0.5	0.28
TRFE	P02787	-2.0	0.22
ALBU	P02768	0.5	0.21
FIBA	P02671	-0.4	0.17
ITIH2	P19823	0.5	0.16
IGHG3	P01860	3.1	0.16
SAA4	P35542	0.7	0.15
KI67	P46013	-1.3	0.15
HEMO	P02790	0.5	0.13
APOL1	O14791	-0.5	0.13
VTNC	P04004	-0.5	0.12
APOA1	P02647	0.8	0.12
CERU	P00450	-0.5	0.12
CXCL7	P02775	-0.5	0.12
GELS	P06396	2.1	0.12
LV106	P04208	-1.0	0.11
CFAB	P00751	-0.5	0.11
KIFC2	Q96AC6	1.7	0.10
A1AT	P01009	-0.6	0.09
CO2	P06681	-0.5	0.09
CFAH	P08603	-0.3	0.09
APOH	P02749	2.8	0.09
THRB	P00734	-0.3	0.08
CO3	P01024	-0.3	0.08
IGHM	P01871	0.6	0.08
CBPN	P15169	-0.4	0.08
IGHA1	P01876	-0.7	0.07
CLUS	P10909	0.3	0.07
C4BPA	P04003	0.5	0.07

KNG1	P01042	0.3	0.07
TRPV2	Q9Y5S1	-1.0	0.06
PK3CG	P48736	-1.2	0.06
CD5L	O43866	2.4	0.05
KV106	P01598	-1.4	0.05
IGJ	P01591	-0.5	0.05
APOA2	P02652	0.3	0.05
A2MG	P01023	4.0	0.05
HRG	P04196	0.4	0.05
AFAM	P43652	0.4	0.05
LAC2	P0CG05	-3.0	0.05
APOA4	P06727	0.4	0.05
LV302	P80748	0.5	0.05
A2GL	P02750	0.8	0.05
CO4B	P0C0L5	2.3	0.05
APOA	P08519	-1.1	0.05
APOD	P05090	-0.4	0.05
HBA	P69905	0.9	0.05
PLMN	P00747	-0.3	0.04
HV101	P01742	-2.0	0.04
KV309	P04433	0.5	0.04
KV102	P01594	0.4	0.04
IC1	P05155	0.5	0.04
CO8A	P07357	-0.3	0.04
PROS	P07225	0.4	0.04
IGKC	P01834	0.5	0.04
PON1	P27169	0.9	0.04
SHC1	P29353	0.3	0.04
HPT	P00738	1.1	0.04
IBP3	P17936	-0.5	0.04
RET4	P02753	-0.7	0.03
HDAC1	Q13547	-0.4	0.03
C1S	P09871	0.2	0.03
C1QC	P02747	0.4	0.03
KCNC4	Q03721	-0.9	0.03
CAMP	P49913	-0.5	0.03
LG3BP	Q08380	-0.7	0.03
ANT3	P01008	-0.3	0.03
A1AG1	P02763	0.8	0.03
H2AX	P16104	-1.2	0.03
FA12	P00748	-0.6	0.03
PROP	P27918	-0.5	0.03
CFAI	P05156	-0.3	0.03
A1BG	P04217	0.3	0.03

IGLL5	B9A064	0.5	0.03
IGHG1	P01857	-0.7	0.02
CO9	P02748	-0.3	0.02
ACTG	P63261	-0.8	0.02
HBB	P68871	0.9	0.02
APOB	P04114	-0.5	0.02
APOC3	P02656	-0.4	0.02
FINC	P02751	-0.3	0.02
IGHD	P01880	-1.0	0.02
FBLN1	P23142	-0.5	0.02
APOC2	P02655	0.5	0.02
PGRP2	Q96PD5	-0.4	0.02
APOF	Q13790	-0.7	0.02
ITIH1	P19827	0.5	0.02
FCN2	Q15485	0.5	0.02
FHR3	Q02985	0.6	0.01
KANK3	Q6NY19	-0.9	0.01
FCG3A	P08637	-0.5	0.01
ANGT	P01019	-0.4	0.01
HV311	P01772	-0.6	0.01
C1R	P00736	-0.7	0.01
APOE	P02649	0.4	0.01
CO6	P13671	-0.7	0.01
HS12B	Q96MM6	-0.3	0.01
APOM	O95445	0.2	0.01
IGHA2	P01877	0.3	0.01
TSP1	P07996	0.5	0.01
IGHG2	P01859	0.9	0.01
SHBG	P04278	-0.8	0.01
K1024	Q9UPX6	-0.4	0.01
LV403	P01717	-0.2	0.01
HPTR	P00739	-0.6	0.01
FOXN3	O00409	0.7	0.01
KLKB1	P03952	-0.9	0.01
CO5	P01031	-0.6	0.01
HS90B	P08238	-0.6	0.01
HV304	P01765	-0.6	0.01
HV305	P01766	0.6	0.01
PLF4	P02776	-0.2	0.00
KV305	P01623	0.7	0.00
PEDF	P36955	0.6	0.00

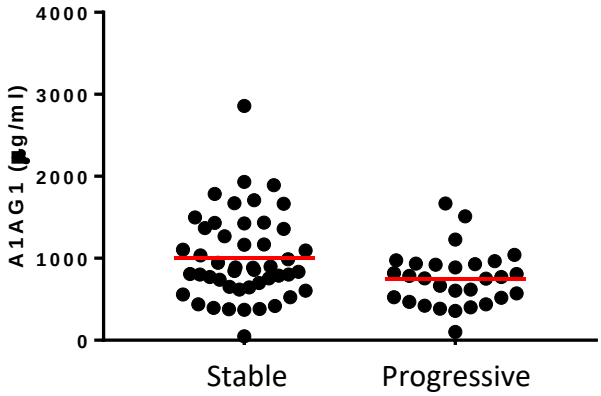
Protein differences are expressed as fold change ranked by significance.

Supplementary table E3. *GC* allele frequencies in control women, UK and NHLBI LAM cohorts

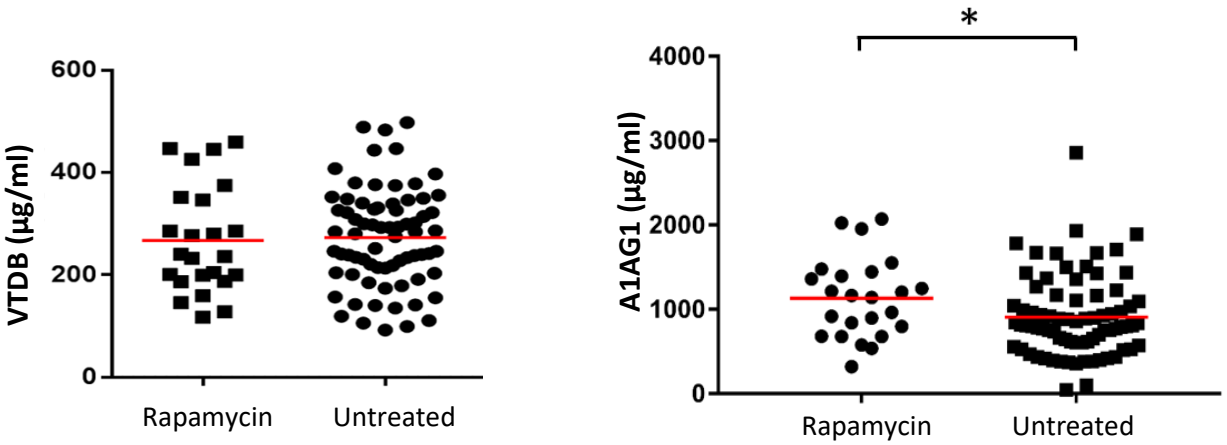
Genotype	Cohort				p value		
	Controls	UK LAM	NHLBI LAM		control vs. UK LAM	control vs. NHLBI LAM	UK LAM vs. NHLBI LAM
n	168141	65	rs7041	rs4588			
			145	146			
rs7041							
GG	31	22	37	-	0.076	0.29	0.075
GT	50	49	43				
TT	19	29	20				
rs4588							
AA	8	14	-	8	0.20	0.17	0.43
AC	42	34		34			
CC	50	52		58			

Percentage allele frequencies are shown for women of European ancestry in the three cohorts. Allele frequencies were compared using the Chi-squared test.

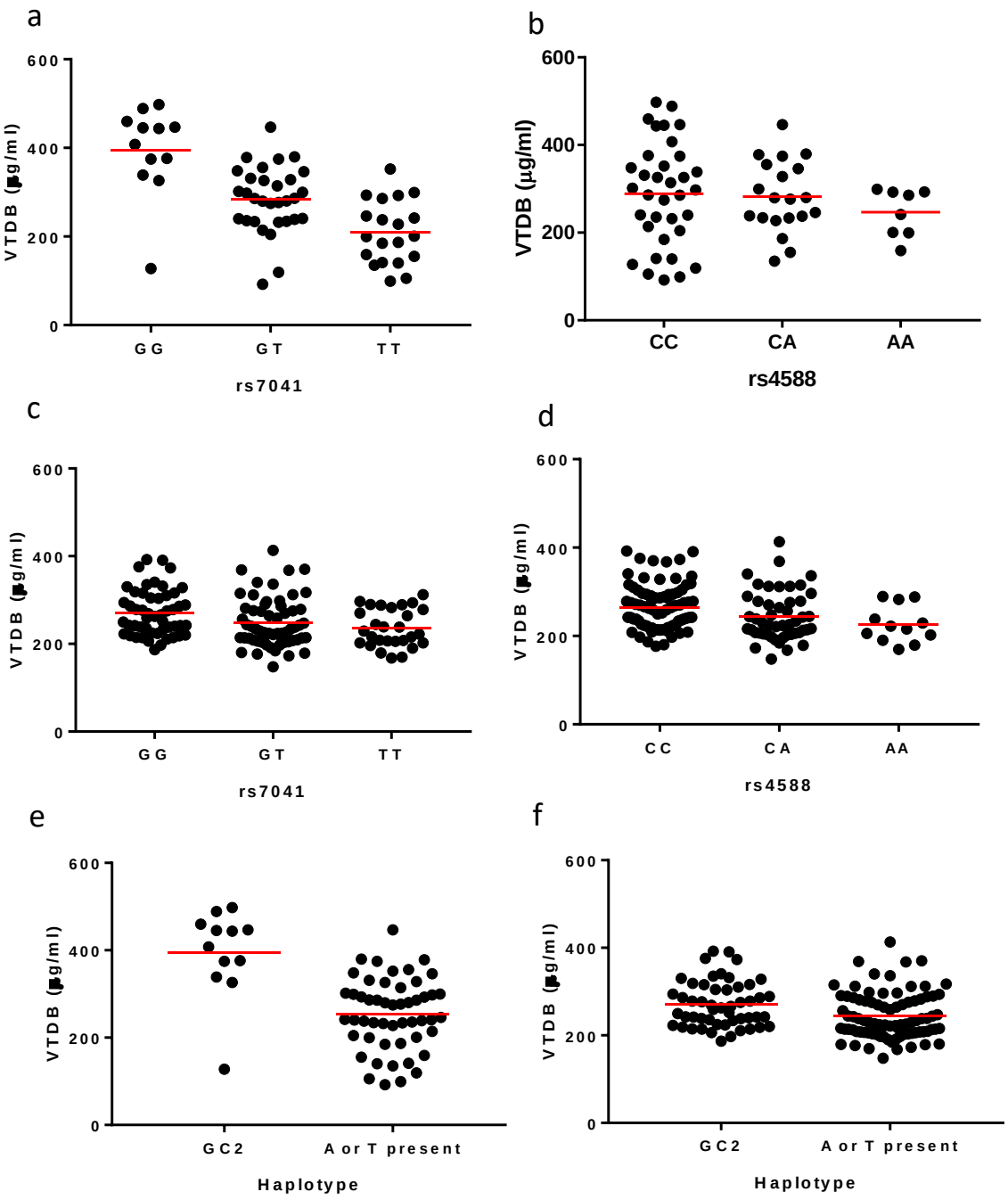
Supplementary figures



Supplementary figure E1. Relationship between serum Alpha1-acid glycoprotein (A1AG1) and disease activity. Serum A1AG1 of women with stable LAM is significantly higher than those with progressive disease (p=0.01).



Supplementary figure E2. Serum Vitamin D Binding Protein (VTDB) levels in patients with LAM untreated or treated with rapamycin. Serum A1AG1 levels in rapamycin treated LAM compared with untreated LAM, (*p=0.031).



Supplementary figure E3. Relationship between *GC* genotype and serum VTDB. In the UK LAM cohort, the presence of the T allele at rs7041 was dose dependently associated with lower serum VTDB levels (n=63, p<0.0001, panel a) although rs4588 was not associated with serum VTDB level (p=0.57, panel b). In the NHLBI Registry cohort, the T allele at rs7041 and the A allele at rs4588 were dose dependently associated with lower serum VTDB levels (n=139, p=0.010 and n=140, p=0.035 respectively, panels c and d). Haplotype analysis combining the allelic information at both SNPs showed the presence of the minor alleles at either rs7041 or rs4588 (T and A respectively) were associated with lower serum VTDB levels in both the UK and NHLBI cohorts, p<0.0001 and p=0.0018, respectively, panels e and f).