



Evidence for shared genetic risk factors between lymphangiomyomatosis and pulmonary function

Xavier Farré^{1,17}, Roderic Espín^{2,17}, Alexandra Baiges^{2,17}, Eline Blommaert², Wonji Kim³, Krinio Giannikou⁴, Carmen Herranz², Antonio Román⁵, Berta Sáez⁵, Álvaro Casanova⁶, Julio Ancochea⁷, Claudia Valenzuela⁷, Piedad Ussetti⁸, Rosalía Laporta⁸, José A. Rodríguez-Portal^{9,10}, Coline H.M. van Moorsel¹¹, Joanne J. van der Vis¹¹, Marian J.R. Quanjel¹¹, Mireia Tena-Garitaonandia^{12,13}, Fermín Sánchez de Medina^{13,14}, Francesca Mateo², María Molina-Molina^{10,15}, Sungho Won¹⁶, David J. Kwiatkowski^{4,18}, Rafael de Cid^{1,18} and Miquel Angel Pujana^{2,10,18}

¹Genomes for Life – GCAT Lab Group, Institut Germans Trias i Pujol, Badalona, Spain. ²ProCURE, Catalan Institute of Oncology, Oncobell, Bellvitge Institute for Biomedical Research (IDIBELL), Barcelona, Spain. ³Channing Division of Network Medicine, Dept of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. ⁴Cancer Genetics Laboratory, Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA. ⁵Lung Transplant Unit, Pneumology Service, Lymphangiomyomatosis Clinic, Vall d'Hebron University Hospital, Vall d'Hebron Research Institute (VHIR), Barcelona, Spain. ⁶Pneumology Service, University Hospital of Henares, University Francisco de Vitoria, Coslada, Spain. ⁷Pneumology Service, La Princesa Research Institute, University Hospital La Princesa, Madrid, Spain. ⁸Pneumology Service, University Hospital Clinica Puerta del Hierro, Majadahonda, Spain. ⁹Medical-Surgical Unit of Respiratory Diseases, University Hospital Virgen del Rocío, Institute of Biomedicine of Seville (IBIS), Seville, Spain. ¹⁰Biomedical Research Network Centre in Respiratory Diseases (CIBERES), Instituto de Salud Carlos III, Madrid, Spain. ¹¹Interstitial Lung Disease Center of Excellence, St Antonius Hospital, Nieuwegein, The Netherlands. ¹²Dept of Biochemistry and Molecular Biology II, School of Pharmacy, Instituto de Investigación Biosanitaria (ibs.GRANADA), University of Granada, Granada, Spain. ¹³Biomedical Research Network Centre in Hepatic and Digestive Diseases (CIBERehd), Instituto de Salud Carlos III, Madrid, Spain. ¹⁴Dept of Pharmacology, School of Pharmacy, ibs.GRANADA, University of Granada, Granada, Spain. ¹⁵Interstitial Lung Disease Unit, Department of Respiratory Medicine, University Hospital of Bellvitge, IDIBELL, Barcelona, Spain. ¹⁶Dept of Public Health Sciences, Interdisciplinary Program of Bioinformatics, Institute of Health and Environment, Seoul National University, Seoul, South Korea. ¹⁷Joint first authors. ¹⁸Joint senior authors.

Corresponding author: Miquel Angel Pujana (miguelangel.pujana@gmail.com)



Shareable abstract (@ERSpublications)

Pleiotropic molecular factors influencing LAM development and pulmonary function
<https://bit.ly/3AE3KgJ>

Cite this article as: Farré X, Espín R, Baiges A, *et al.* Evidence for shared genetic risk factors between lymphangiomyomatosis and pulmonary function. *ERJ Open Res* 2021; 7: 00375-2021 [DOI: 10.1183/23120541.00375-2021].

Copyright ©The authors 2021

This version is distributed under the terms of the Creative Commons Attribution Non-Commercial Licence 4.0. For commercial reproduction rights and permissions contact permissions@ersnet.org

This article has supplementary material available from openres.ersjournals.com

Received: 2 June 2021
Accepted: 17 Sept 2021



Abstract

Introduction Lymphangiomyomatosis (LAM) is a rare low-grade metastasising disease characterised by cystic lung destruction. The genetic basis of LAM remains incompletely determined, and the disease cell-of-origin is uncertain. We analysed the possibility of a shared genetic basis between LAM and cancer, and LAM and pulmonary function.

Methods The results of genome-wide association studies of LAM, 17 cancer types and spirometry measures (forced expiratory volume in 1 s (FEV₁), forced vital capacity (FVC), FEV₁/FVC ratio and peak expiratory flow (PEF)) were analysed for genetic correlations, shared genetic variants and causality. Genomic and transcriptomic data were examined, and immunodetection assays were performed to evaluate pleiotropic genes.

Results There were no significant overall genetic correlations between LAM and cancer, but LAM correlated negatively with FVC and PEF, and a trend in the same direction was observed for FEV₁. 22 shared genetic variants were uncovered between LAM and pulmonary function, while seven shared variants were identified between LAM and cancer. The LAM-pulmonary function shared genetics identified four pleiotropic genes previously recognised in LAM single-cell transcriptomes: *ADAM12*, *BNC2*, *NR2F2* and *SP5*. We had previously associated *NR2F2* variants with LAM, and we identified its functional partner *NR3C1* as another pleiotropic factor. *NR3C1* expression was confirmed in LAM lung lesions. Another candidate pleiotropic factor, *CNTN2*, was found more abundant in plasma of LAM patients than that of healthy women.

Conclusions This study suggests the existence of a common genetic aetiology between LAM and pulmonary function.

Introduction

Lymphangiomyomatosis (LAM) is a rare low-grade progressive neoplasm that affects women almost exclusively and is characterised by cystic lung destruction, which can lead to respiratory failure in severe cases [1, 2]. In addition to cystic lung disease, LAM is also strongly associated with renal angiomyolipomas (AML) and lymphatic alterations, for which reason it may be properly considered a systemic disorder [1, 2]. LAM lung lesions are characterised by low-grade proliferation of “LAM cells” that have both smooth muscle cell-like features, and microphthalmia-associated transcription factor (MITF)-driven gene expression; and by cyst formation, which is likely driven by expression of proteases and cathepsins [3–5]. The tissue of origin of LAM cells is uncertain [6–8].

LAM can occur sporadically (S-LAM) or in the presence of tuberous sclerosis complex (TSC-LAM), an autosomal-dominant multisystem disorder caused by heterozygous germline or mosaic loss-of-function mutations in the tumour suppressor genes *TSC1* and *TSC2* [9, 10]. In S-LAM, somatic inactivation of *TSC2*, or much less commonly *TSC1*, in an unknown cell type(s) appears to be sufficient for disease development [11]. Germline or somatic mutations in *TSC1/TSC2* lead to hyperactivation of the mechanistic target of rapamycin complex 1 (mTORC1), and mTOR allosteric inhibitors (rapamycin/sirolimus and its derivatives, rapalogs) are the main therapeutic approach for LAM [12]. Owing to its central role in metabolism, mTORC1 activity is abnormally enhanced in many cancer types, and generally linked to stem cell-like features, which are also present in LAM cells [13–16]. Indeed, LAM shows several fundamental hallmarks of cancer, including continued cell proliferation and resistance to cell death, expression of factors promoting tissue invasion and metastasis, and immune evasion [6, 17–19].

A recent genome-wide association study (GWAS) we performed identified several common genetic variants on chromosome 15q26.2 for which an allele was associated with risk of S-LAM [20]. Although not certain, these alleles appear to affect LAM development through effects on the nuclear receptor subfamily 2 group F member 2 (*NR2F2*) gene, also known as chicken ovalbumin upstream promoter transcription factor II (COUP-TFII). This transcription factor is widely expressed during embryogenesis and has a role in various endocrine conditions and cancers [21, 22]. Here we sought to explore in further detail the genetic basis of LAM risk, considering the hypothesis that there is a shared genetic basis between LAM and cancer and/or pulmonary function.

Materials and methods

GWAS data

The GWAS summary statistics of LAM [20], 17 cancer types [23] and pulmonary function tests (including of forced expiratory volume in 1 s (FEV₁), forced vital capacity (FVC), FEV₁/FVC ratio and peak expiratory flow (PEF)) [24] were obtained from the corresponding data sources (supplementary Table S1 provides details of the GWASs, including sample sizes and relevant references). This study did not require individual data. The original LAM GWAS included 426 S-LAM subjects that were analysed in comparison with 852 females from the COPDGene study in a matched case–control design [20].

Imputation and data preprocessing

The GWAS summary statistics for LAM were imputed using the SSimp software [25, 26] and European individuals from the 1000 Genomes Project phase 3 reference panel, filtering out single nucleotide polymorphisms (SNPs) with a minor allele frequency (MAF) ≤ 0.01 . The imputation increased the number of SNPs from 681 894 to 9 325 933, but those with poor imputation quality ($r^2 < 0.3$) were removed, providing a set of 7 809 072 SNPs for subsequent analyses. For each of the variant-summary statistics, standard quality controls were applied: removal of SNPs without reference identifier (rs ID), duplicated, poorly imputed (info score < 0.9), with $MAF \leq 0.01$, with strand ambiguous alleles and/or with sample size 5 standard deviations away from the mean. In addition, SNPs in the extended major histocompatibility complex (hg 19 coordinates, chr6: 25 119 106–33 854 733) and 8p23.1 region (chr8: 7 200 000–12 500 000) were excluded given that these regions have complex linkage disequilibrium that could bias pleiotropy analyses [27].

Shared genetic architecture

The estimation of heritability of all phenotypes and computation of genetic correlation between GWASs of LAM and each of the other selected diseases or traits were performed using the high-definition likelihood (HDL) inference method [28]. Genetic correlations estimated using HDL fully account for linkage disequilibrium across the genome, increasing the power to detect connections between complex traits using precomputed linkage disequilibrium matrices for 335 265 Genomic British individuals in the UK Biobank and HapMap3 SNPs [28]. For detection of shared genetic risk factors, we used the pleiotropy-informed conditional false discovery rate approach (FDR) [29] applying pleioFDR software (<https://github.com/>

precimed/pleiofdr) and computing conjunctive false discovery rate (conjFDR) statistics. The conjFDR is given by the maximum between the conditional FDRs (condFDR) for two given conditions; the condFDR method was shown to improve statistical power relative to the conventional approach of using p-value thresholds for detection of shared genetics and was not affected by the direction of the allele effects [30, 31]. A pairwise analysis was performed between LAM GWAS and each other condition. In order to make the results comparable, we analysed a common set of 5684891 SNPs present in all summary statistics. Shared genetic variants were defined by $\text{conjFDR} < 0.05$. We then performed linkage disequilibrium clumping to define independent significant SNPs (PLINK software, $p_1 = 0.05$; $p_2 = 1$, linkage disequilibrium threshold $r^2 = 0.6$, and physical distance threshold for clumping 1000 kb) and lead SNPs (PLINK software, $p_1 = 0.05$, $p_2 = 1$, $r^2 = 0.1$, and distance 1000 kb). Genomic risk loci were found by merging lead SNPs if they were closer than 250 kb. Candidate SNPs were mapped to independent significant SNPs using this clumping strategy. Stratified Q-Q plots were obtained using pleioFDR to visualise shared genetic architecture. In these representations, the p-values of the primary phenotype were plotted against the null distribution. In the same plots, we represented subsets of SNPs of the primary phenotype conditioned by the significance of their association with the secondary phenotype; p-value of the secondary phenotype < 0.1 , 0.01 and 0.001. A leftwards deflection of the lines implies the presence of shared genetic architecture.

Mendelian randomisation

A two-sample Mendelian randomisation (MR) approach was applied to evaluate causality using GWAS summary statistics of cancer risk and spirometry measures as exposure or outcome, and LAM as the outcome or exposure, respectively. A valid instrumental variable should fulfil three core assumptions: the variant is associated with the exposure; the variant is independent of all confounders of the exposure–outcome association; and the variant is only associated with the outcome through its effect on the exposure [32]. Independent genetic variants (linkage disequilibrium $r^2 > 0.001$; genome-wide significance, $p < 5 \times 10^{-8}$, for spirometry measures; and at a suggestive level, $p < 1 \times 10^{-5}$ for cancer and LAM risk) were used as instrumental variables. The three-step MR-Egger method [33] was applied: 1) a test for directional pleiotropy; 2) a test for a causal effect; and 3) estimation of the causal effect. To assess the robustness of the results, causal estimates and p-values were compared using random-effects inverse-variance-weighted and robust adjusted profile score (MR-RAPS) [34] methods. Heterogeneity was calculated from Cochran's Q value [35]. The MR Steiger directionality test, which compares the variance explained by the SNPs in the exposure and outcome, was applied to elucidate the direction of causality [36]. This test estimates directionality leveraging the fact that in most settings the genetic variants will explain more variance of the trait located upstream in the causal chain. Informative scatter and forest plots were generated to examine the results. The analyses were performed using the TwoSampleMR R package [37].

Gene candidate prioritisation

Information of positional gene candidates (up to 1 megabase from the given variant) was integrated with data of expression quantitative trait loci (eQTL) in *cis* identified in non-diseased human tissue (Genotype-Tissue Expression (GTEx) project [38]), and with data of chromatin interactions identified in lung tissue as potential *cis*-regulatory elements [39]. The GTEx evidence was obtained from pan-tissue analyses as well as from tissue expected to be biologically related to LAM disease: blood vessel, lung, kidney and uterus. Information for physical protein interactions and complex membership was taken from the Human Reference Protein Interactome [40] and Biogrid [41] databases. For the evaluation of publicly available single-cell RNA-sequencing (RNA-seq) profiles from LAM1–4 diseased lungs (Gene Expression Omnibus reference GSE135851), the samples were pre-processed and analysed as originally described [8], using the Seurat R package [42]. LAM cells (excluding lung mesenchymal cells) were identified using the LAM^{core} gene expression signature originally described by the authors as the reference [8]. RNA-seq data of kidney AMLs corresponded to 14 published cases from two studies ($n = 5$ [20]; and $n = 9$ [43]) and 14 unpublished cases (internal cohort, unpublished data). RNA-seq data of other cancer types corresponded to The Cancer Genome Atlas (TCGA) [44] and were downloaded from the cBioPortal for Cancer Genomics [45]. The single-cell RNA-seq data of the Human Lung Cell Atlas [46] was downloaded from the Synapse portal (<https://www.synapse.org/#!Synapse:syn21041850/wiki/600865>). For all candidate pleiotropic genes, and for each lung cell type, a percentage of cells with a given gene expression was calculated using the ComplexHeatmap R package with standard parameters [47].

LAM lung samples, immunohistochemistry

LAM patients were recruited and lung tissue samples collected by the participating centres (International LAM Clinic, University Hospital Vall d'Hebron; University Hospital La Princesa; University Hospital Clínica Puerta del Hierro; University Hospital Clínic Barcelona; University Hospital Virgen del Rocío; and University Hospital of Bellvitge) and with the support of the Spanish LAM Association (AELAM). The

patients provided written informed consent, and the study was approved by the Research Ethics Committee of the IDIBELL. The immunohistochemical assays were performed on serial paraffin sections using an EnVision kit (Dako) and antigen retrieved with citrate pH 6.0 buffer. Endogenous peroxidase was blocked by pre-incubation in a solution of 3% H₂O₂, performed in 1× phosphate-buffered saline with 10% goat serum. Slides were incubated overnight at 4°C with primary antibodies (anti-NR3C1, dilution 1:100, D6H2L, Cell Signalling Technology, Danvers, MA, USA; and anti-NTN4, dilution 1:30, HPA049832, Sigma-Aldrich, St. Louis, MO, USA) in blocking solution. Secondary peroxidase-conjugated antibody (Envision+ system-HRP, Dako) was used. Sections were counterstained with haematoxylin and examined under a Nikon Eclipse 80i microscope. For immunohistofluorescence, the slides were incubated with a mixture of the two primary antibodies (anti- α SMA, dilution 1:1000, A2547, Sigma-Aldrich; and anti-NR3C1). The secondary antibodies used were goat anti-mouse Alexa-546 and goat anti-rabbit Alexa-488 (dilution 1:100; Thermo Fisher, Waltham, MA, USA). Sudan black staining was performed to avoid paraffin autofluorescence. The sections were mounted with coverslips in Vectashield containing DAPI and visualised under a Nikon 80i epifluorescence microscope equipped with a DS-Ri1 camera.

Plasma samples and ELISA

LAM blood samples were collected and immediately processed during the 2017 and 2018 annual AELAM patient conferences, so the time between undertaking pulmonary function tests and sample acquisition varied, making it impossible to assess the former relative to the biomarkers. The data collected consisted of age at diagnosis, age at sample extraction, diagnosis of AML, chylothorax, pneumothorax, TSC and therapy used at the time of sample extraction. All patients provided written informed consent and the study was approved by the ethics committees of IDIBELL and the Instituto de Investigación Sanitaria La Princesa, Hospital de Henares. Control samples were obtained from healthy premenopausal women from a similar age distribution to that of the LAM patients. Lung disease-related patients corresponded to cases diagnosed with emphysema (excluding COPD; University Hospital of Bellvitge, IDIBELL), Langerhans cell histiocytosis, Sjögren syndrome and systemic lupus erythematosus. Blood samples of the former conditions were collected by the ILD Centre of Excellence, St. Antonius Hospital Biobank (Nieuwegein, The Netherlands). This study was approved by the institutional ethics committee (reference R05-08A), and all participants provided written informed consent. CNTN2 was quantified in blood plasma using the Human Contactin-2/TAG1 DuoSet enzyme-linked immunosorbent assay (ELISA; DY1714-05, R&D Systems, Minneapolis, MN, USA) following the manufacturer's instructions. Vascular endothelial growth factor-D (VEGF-D) levels were measured using a commercially available ELISA kit according to the manufacturer's protocol (DVED00, R&D Systems).

Results

Identification of shared genetics of LAM and cancers

As a low-grade neoplasm, LAM has a biological similarity to cancer [6, 17–19]. This connection might be extended to disease susceptibility [48, 49] and could indicate the existence of a shared genetic basis of the two conditions. To assess this hypothesis, we analysed summary statistic results from the original S-LAM GWAS [20] and from studies of 17 cancer types, including those of breast, colon, glioma, kidney, lung, neuroblastoma, ovary, pancreas, prostate, stomach and skin (supplementary Table S1). After SNP imputation, data preprocessing and quality control analyses, the summary statistics of 5684891 SNPs were evaluated between LAM and cancers. The overall pairwise genetic correlation between LAM and cancer risk did not reach significance for any cancer; the strongest correlation was negative and for cervical cancer risk ($r=-0.32$, $p=0.083$; supplementary Table S2). In addition, when conditioning LAM on any cancer type, there were no substantial signs of shared genetic architecture in the stratified Q-Q plots (supplementary Figure S1). However, for some cancer analyses, there was not enough statistical power to estimate genetic correlations accurately (supplementary Table S2).

Despite the non-significant overall LAM-cancer genetic correlation, seven shared genetic associations were identified with $\text{conjFDR}<0.05$, including variants between LAM and gastric, kidney or prostate cancer risk (table 1). These variants were mapped to six genomic loci, of which three were linked to gastric cancer, two to kidney cancer and one to prostate cancer risk. Four of the associated loci (including gastric, kidney and prostate cancer) were predicted to function as agonists; that is, the corresponding minor alleles showed the same direction of effect for LAM and cancer risk; in turn, two were predicted to act as antagonists (including gastric and kidney cancer; table 1).

The chromosome 6q24 rs3861451 variant shared between LAM and gastric cancer risk was relatively linked (Caucasian $D'=0.74$ and $r^2=0.52$) with a previously noted SNP association for this cancer type, rs618688 [23]. Next, we inspected the seven identified LAM-cancer shared variants (table 1) relative to the GWAS catalogue for human traits and diseases [50, 51]. This examination identified the chromosome 2q31

TABLE 1 Shared genetic variants between lymphangioleiomyomatosis (LAM) and cancer risk (ordered by cancer type)

Cancer type	SNP	Chr	Position (bp, hg19)	Reference allele	Alternative allele (MAF)	Cancer p-value	LAM p-value	Cancer Z score	LAM Z score	Pleiotropy	conjFDR	Locus
Gastric	rs3861451	6	148320047	T	C (0.33)	1.6×10^{-4}	1.2×10^{-5}	3.78	-4.38	Antagonist	0.022	Intergenic
	rs9528802	13	65235556	T	C (0.29)	1.8×10^{-4}	2.4×10^{-4}	3.74	3.68	Agonist	0.035	Intergenic
	rs10901587	10	128001759	C	T (0.23)	2.5×10^{-4}	2.5×10^{-4}	3.67	3.66	Agonist	0.037	<i>ADAM12</i> intronic
Kidney	rs4512050	4	158597491	A	G (0.13)	2.9×10^{-4}	1.9×10^{-4}	3.62	3.74	Agonist	0.044	Intergenic
	rs17036640		158605184	A	G (0.13)	2.5×10^{-4}	1.7×10^{-4}	3.67	3.76	Agonist	0.037	Intergenic
	rs2146084	9	112505617	G	A (0.12)	2.6×10^{-5}	2.7×10^{-4}	-4.21	3.65	Antagonist	0.047	<i>PALM2-AKAP2</i> intronic
Prostate	rs4668267	2	171360896	T	C (0.30)	7.9×10^{-5}	8.7×10^{-5}	3.95	3.93	Agonist	0.030	<i>MYO3B</i> intronic

SNP: single nucleotide polymorphism; Chr: chromosome; MAF: minor allele frequency; conjFDR: conjunctive false discover rate.

variant rs4668267 that connects LAM and prostate cancer risk as a genome-wide association signal of earlobe attachment [52].

Identification of shared genetics of LAM and pulmonary function

In parallel with cancer, we analysed shared genetic factors between LAM and pulmonary function, determined by spirometry measures of FEV₁, FVC, FEV₁/FVC ratio and PEF in the general population using summary statistic GWAS results from the UK Biobank and SpiroMeta consortium [24]. The overall pairwise genetic correlations between LAM and FVC, and PEF, were nominally significant and negative: $r = -0.074$, 95% CI -0.04 – -0.10 , $p = 0.013$; and $r = -0.10$, 95% CI -0.05 – -0.15 , $p = 0.029$, respectively (supplementary Table S2). The LAM correlation with FEV₁ was also negative but did not reach significance: $r = -0.061$, $p = 0.097$. By conditioning LAM on spirometry measures, the stratified Q-Q plots showed a leftwards deflection with increasingly strong associations with the depicted measures, such as PEF (figure 1; Q-Q plots for the rest of measures are shown in supplementary Figure S2).

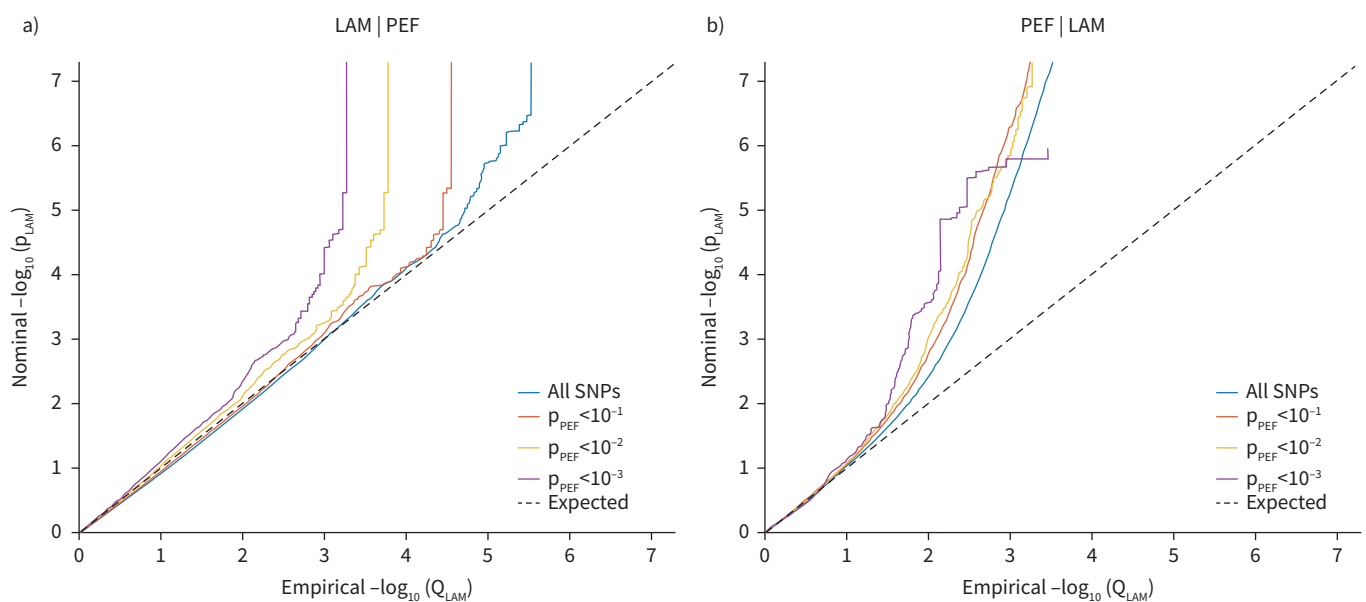


FIGURE 1 Stratified Q-Q plots including lymphangioleiomyomatosis (LAM) and peak expiratory flow (PEF) genome-wide association study (GWAS) results. a) Q-Q plot (nominal versus empirical $-\log_{10}$ p-values, corrected for inflation) conditioning LAM on PEF; and b) Q-Q plot conditioning PEF on LAM. Leftwards deflection from the null distribution of the observed p-value as the thresholds become more stringent indicates genetic overlap between the two traits. SNP: single nucleotide polymorphism.

TABLE 2 Shared genetic variants between lymphangioleiomyomatosis (LAM) and pulmonary function (ordered by chromosome)

Trait	SNP	Chr	Position (bp, hg19)	Reference allele	Alternative allele (MAF)	Trait <i>p</i> -value	LAM <i>p</i> -value	Trait Z score	LAM Z score	Pleiotropy	conjFDR	Gene locus
FEV ₁ /FVC	rs11240341	1	205 015 284	C	T (0.34)	7.7×10 ⁻⁶	7.5×10 ⁻⁶	-4.47	4.48	Antagonist	0.007	<i>CNTN2</i> intronic
FEV ₁ /FVC	rs16937		205 035 455	G	A (0.32)	1.8×10 ⁻⁵	1.9×10 ⁻⁵	-4.29	4.28	Antagonist	0.017	
FEV ₁	rs13410076	2	46 815 961	T	C (0.09)	3.4×10 ⁻⁴	3.9×10 ⁻⁵	-3.58	-4.11	Agonist	0.022	<i>PIGF</i> intronic
PEF	rs835069	5	14 984 116	C	T (0.15)	3.9×10 ⁻⁴	1.1×10 ⁻⁴	3.54	-3.86	Antagonist	0.044	Intergenic
FEV ₁	rs7701443		142 792 650	A	G (0.39)	3.6×10 ⁻⁴	1.5×10 ⁻⁵	3.57	4.32	Agonist	0.023	<i>NR3C1</i> intronic
FVC						2.1×10 ⁻⁷	1.5×10 ⁻⁵	5.19	4.32	Agonist	0.011	
FVC	rs2537572	7	17 771 780	A	G (0.41)	9.5×10 ⁻⁷	7.5×10 ⁻⁵	-4.90	3.96	Antagonist	0.046	Intergenic
FEV ₁						4.4×10 ⁻⁴	2.8×10 ⁻⁷	3.51	-5.14	Antagonist	0.026	
FEV ₁	rs9648555		46 656 869	A	G (0.47)	7.7×10 ⁻⁷	7.5×10 ⁻⁵	-4.94	3.96	Antagonist	0.033	Intergenic
FVC	rs2511654	8	117 623 003	T	C (0.30)	1.7×10 ⁻⁴	2.7×10 ⁻⁵	-3.76	4.19	Antagonist	0.018	Intergenic
FEV ₁	rs12542425		117 642 345	C	T (0.37)	2.2×10 ⁻⁴	1.4×10 ⁻⁶	-3.69	4.82	Antagonist	0.015	Intergenic
FVC						1.7×10 ⁻⁴	1.4×10 ⁻⁶	-3.77	4.82	Antagonist	0.014	
FVC	rs7839361		117 645 687	T	C (0.17)	6.4×10 ⁻⁵	7.4×10 ⁻⁶	-3.99	4.48	Antagonist	0.006	Intergenic
FEV ₁	rs17663673		117 650 006	T	C (0.31)	2.1×10 ⁻⁴	5.0×10 ⁻⁵	-3.71	4.05	Antagonist	0.023	Intergenic
FVC						3.1×10 ⁻⁴	5.0×10 ⁻⁵	-3.61	4.05	Antagonist	0.032	
FEV ₁	rs34672734		117 805 281	A	G (0.21)	3.6×10 ⁻⁴	3.7×10 ⁻⁵	-3.57	4.13	Antagonist	0.023	Intergenic
FVC						1.4×10 ⁻⁴	3.7×10 ⁻⁵	-3.82	4.13	Antagonist	0.025	
FEV ₁ /FVC	rs4961722	9	16 529 174	T	C (0.37)	1.8×10 ⁻⁴	3.4×10 ⁻⁵	3.75	4.15	Agonist	0.030	<i>BNC2</i> intronic
FEV ₁ /FVC	rs7959413	12	96 147 836	T	C (0.47)	7.3×10 ⁻¹²	5.2×10 ⁻⁶	6.85	4.56	Agonist	0.005	<i>NTN4</i> intronic
PEF	rs10859942		96 161 207	T	C (0.49)	1.7×10 ⁻⁴	2.0×10 ⁻⁵	3.76	4.27	Agonist	0.011	<i>NTN4</i> intronic
FEV ₁	rs59125351	15	96 144 157	T	G (0.24)	6.6×10 ⁻⁴	6.9×10 ⁻¹¹	3.40	-6.52	Antagonist	0.035	Intergenic
PEF	rs8036214		96 145 329	T	C (0.49)	3.6×10 ⁻⁴	5.0×10 ⁻⁵	3.57	-4.06	Antagonist	0.021	Intergenic
PEF	rs16975396		96 158 705	T	G (0.24)	2.6×10 ⁻⁴	3.9×10 ⁻¹⁰	3.65	-6.26	Antagonist	0.015	Intergenic
FEV ₁	rs8025061		96 176 965	A	G (0.43)	3.6×10 ⁻⁴	1.2×10 ⁻⁵	3.57	-4.38	Antagonist	0.023	Intergenic
PEF						2.4×10 ⁻⁵	1.2×10 ⁻⁵	4.22	-4.38	Antagonist	0.006	
PEF	rs16975446		96 177 806	G	A (0.33)	8.8×10 ⁻⁶	2.3×10 ⁻⁵	4.45	-4.23	Antagonist	0.011	Intergenic
FEV ₁	rs3996842		96 178 859	C	T (0.33)	4.5×10 ⁻⁴	2.9×10 ⁻⁵	3.51	-4.18	Antagonist	0.023	Intergenic
FEV ₁	rs4815366	20	25 049 909	T	G (0.34)	3.7×10 ⁻⁴	5.6×10 ⁻⁵	3.56	4.03	Agonist	0.026	Intergenic

SNP: single nucleotide polymorphism; Chr: chromosome; MAF: minor allele frequency; conjFDR: conjunctive false discovery rate; FEV₁: forced expiratory volume in 1 s; FVC: forced vital capacity; PEF: peak expiratory flow.

22 variants in 11 loci were common to LAM and pulmonary function: 11 shared with FEV₁, seven with FVC, four with FEV₁/FVC ratio and six with PEF (table 2). Six of the identified variants mapped to chromosome 15q26.2, relatively close (≤ 35 kb) to the primary S-LAM associations targeting *NR2F2* [20]; the six identified variants were in relative linkage disequilibrium (Caucasian $D' > 0.90$, $r^2 > 0.25$) with the original genome-wide significant associations detected by rs2006950 and rs4544201 [20]. Inspection of the UK Biobank and SpiroMeta results showed nominal associations between these two original SNPs and pulmonary function measures, with *p*-values between 0.05 and 0.0002. Interestingly, all six variants identified in 15q26.2 had opposite effects in pulmonary function compared with LAM risk: that is, their minor alleles were associated with relatively superior pulmonary function, but with lower LAM risk (table 2).

There was no overlap between the shared variants influencing LAM and cancer, and LAM and pulmonary function. We further inspected the 22 LAM-pulmonary function shared variants relative to cancer risk associations [50, 51] in the corresponding genomic regions. The chromosome 5q31.3 rs7701443 variant is located at 37 kb (Caucasian $D' = 0.90$, $r^2 = 0.15$) of a breast cancer association identified by the Breast Cancer Association Consortium, rs2963155 [53], and the chromosome 9p22.3 variant rs4961722 is located at <1 kb (Caucasian $D' = 1.0$, $r^2 = 0.03$) from a skin cancer association, rs10962474 [54]. In addition, the chromosome 8q23.3 region comprising five LAM-pulmonary function shared variants (table 2) was relatively strongly linked (Caucasian linkage disequilibrium $D' = 0.25-1$; maximum distance between variants ≤ 45 kb) to association signals of colorectal, gastric and rectal cancer risk: rs6469654, rs6469656, rs16892766, rs76316943, rs117079142 and rs200235517 [55].

Next, we inspected the identified LAM-pulmonary function loci for associations with lung-related traits. The LAM-FEV₁ chromosome 2p21 rs13410076 variant is located at 230 kb from a rare variant (Caucasian MAF <0.001) previously associated at the genome-wide level with oxygenated haemoglobin levels in Tibetan women living at high altitude, rs372272284, which may influence the expression of endothelial PAS domain-containing protein 1 (*EPAS1/HIF2A*) gene [56]. Evaluation of other traits identified the LAM-FEV₁/FVC chromosome 1q32.1 rs16937 and rs11240341 variants as genome-wide association signals for mean platelet volume [57] and schizophrenia [58].

Prioritisation of gene candidates

To evaluate potential pleiotropic genes, we examined eQTL data from non-diseased human tissue [38], chromatin interactions identified in lung tissue [39], and protein physical and complex membership interactions [40, 41]. Of the identified loci, the 5q31.3 LAM-pulmonary function shared variant rs7701443 maps in the first intron of the nuclear receptor subfamily 3 group C member 1 (*NR3C1*), and it is an eQTL for this gene (supplementary Table S3). *NR3C1*, also known as glucocorticoid receptor, physically interacts with *NR2F2*, and this relationship influences the transcriptional regulation of defined gene targets [59]. In addition, *NR3C1* positively regulates the expression of another potential pleiotropic factor, *EPAS1*, in breast cancer cells under hypoxic conditions [60]. Two other gene candidates, insulin-like growth factor-binding protein 3 (*IGFBP3*), defined by LAM-FEV₁ shared variant rs9648555, and disintegrin and metalloproteinase domain-containing protein 12 (*ADAM12*), defined by LAM-gastric cancer shared variant rs10901587, code for proteins that physically interact: *ADAM12* cleaves *IGFBP3*, and this biochemical modification regulates insulin-like growth factor (IGF) activity in regenerating and developing tissue, including cancer and pregnancy settings [61]. The variants shared by LAM and pulmonary function detected in chromosome 8q23.3 might target the eukaryotic translation initiation factor 3 subunit H (*EIF3H*) gene, as proposed for the investigated associations with colorectal cancer risk [62].

Expression of candidate genes in LAM cells

A recent landmark study has depicted LAM single-cell gene expression profiles from four patients [8]. This work defined a LAM^{core} expression signature that included *NR2F2* [8], as well as three of the candidates identified in our study: *ADAM12*, basonuclin 2 (*BNC2*) specificity protein 5 transcription factor (*SP5*) (supplementary Table S3). The identification of four LAM^{core} genes among 48 pleiotropic gene candidates (supplementary Table S3) was a higher proportion than expected by chance (hypergeometric test of overlap $p=0.039$). Examination of a second LAM single-cell study that analysed one tissue sample expanded the list of candidates to *IGFBP3* [7]. Then, differential expression analysis between LAM and non-LAM cells using the first dataset [8] identified several gene candidates frequently detected and overexpressed in LAM cells. In addition to *NR2F2*, this analysis identified *EIF3H*, *IGFBP3* and *NR3C1* as being linked to LAM cell profiles (supplementary Table S4). In turn, *EPAS1* was predicted to be underexpressed in LAM cells (supplementary Table S4).

Among the potential pleiotropic genes, we first evaluated *NR3C1* because its product has an established functional relationship with the primary LAM GWAS candidate, *NR2F2* [59], and because of the key role played by glucocorticoids in breast cancer development and metastasis [63–65]. Using RNA-seq data of 28 kidney AMLs [20, 43], a similar tumour entity to LAM, and of a large collection of 27 human cancer types [44], the expression level of *NR3C1* ranked relatively high in AMLs, just behind acute myeloid leukaemia and kidney renal clear cell carcinoma (figure 2a). In addition, we found a significant (Fishers' exact test $p=5.3\times 10^{-10}$) gene overlap between an experimentally defined *NR3C1*-activity signature [66] and genes differentially expressed in LAM single cells [8] (figure 2b). Then, immunohistochemistry assays in lung tissue from seven S-LAM patients confirmed substantial *NR3C1* expression in all of them, with nuclear positivity in epithelioid and spindle-like diseased cells (figure 2c). Co-staining with the LAM marker α -smooth muscle actin (α SMA) showed cellular colocalisation with *NR3C1* expression in lung nodules, although non- α SMA lung cells also presented nuclear expression of *NR3C1* (figure 2d).

Other candidate genes emerged from the exploration of eQTLs and genomic regulatory signals (supplementary Table S3), in combination with the suggestive evidence from other neoplasms. The netrin-4 (*NTN4*) gene—chromosome 12q22 LAM-pulmonary function shared associations (table 2)—was identified as being the target of a breast cancer risk association [67], and its product influences metastatic potential and angiogenesis [68, 69]. Immunohistochemistry assays did not reveal *NTN4* expression in LAM lung lesions, although the normal alveolar layer was positive (figure 3a). Of the other candidates (supplementary Table S3), the contactin-2 (*CNTN2*) gene was prioritised based on criteria similar to those for *NTN4* (supplementary Table S3), and its product is linked to inflammatory conditions [70] and interacts with oestrogen receptor α [71]. Since *CNTN2* can be detected in body fluids [72], we measured its levels in the plasma of LAM patients and compared the results with those of healthy women and patients with

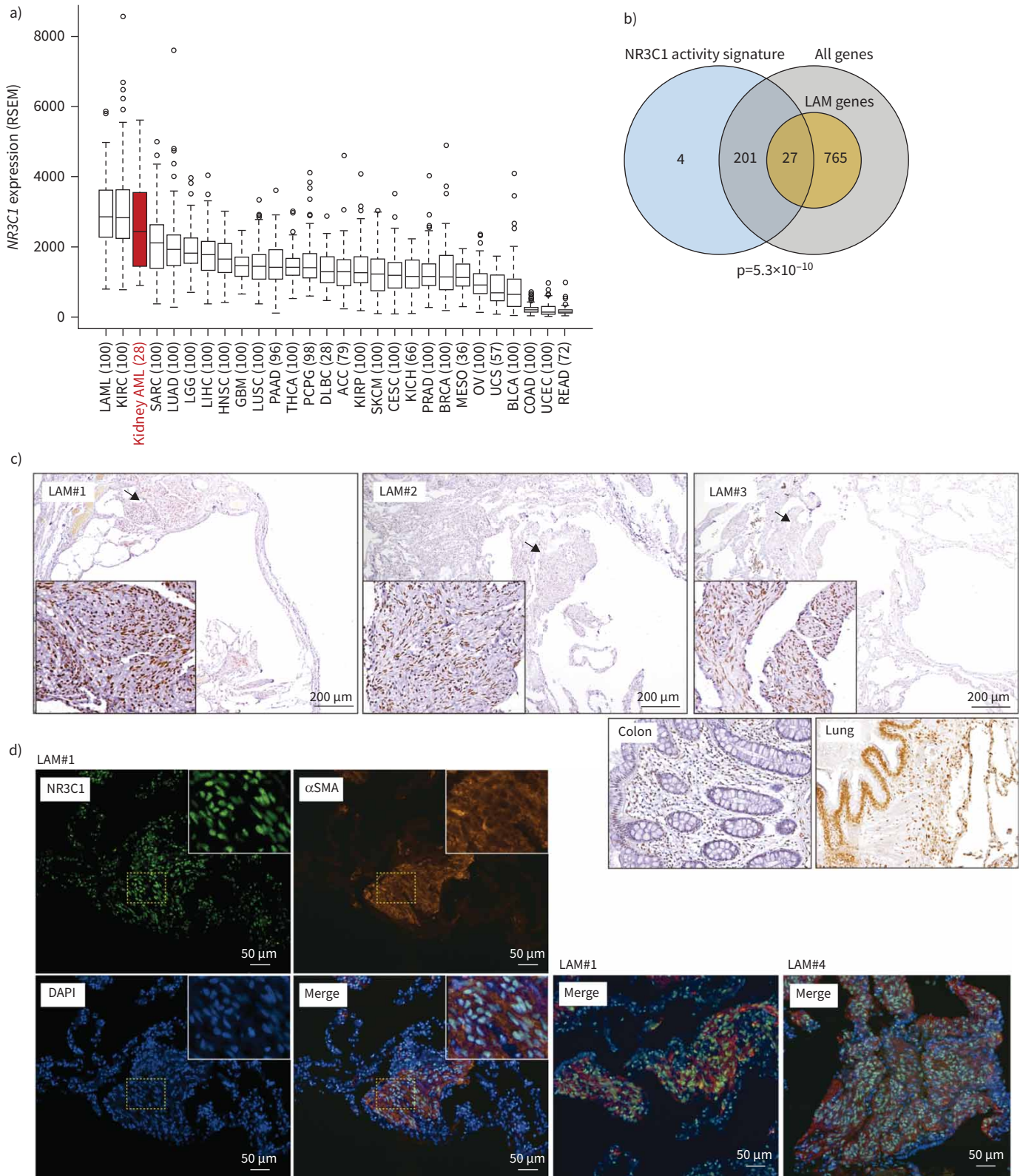


FIGURE 2 *NR3C1* gene expression in angiomyolipoma (AML) tumours and *NR3C1* protein expression in lymphangioleiomyomatosis (LAM) lung lesions. **a)** Comparison of *NR3C1* expression in kidney AMLs (red font) with other neoplasms (TCGA data: 2463 tumours of 27 histological types). Cancer abbreviations: LAML: acute myeloid leukaemia; KIRC: kidney renal clear cell carcinoma; SARC: sarcoma; LUAD: lung adenocarcinoma; LGG: low-grade glioma; LIHC: liver hepatocellular carcinoma; HNSC: head and neck squamous cell carcinoma; GBM: glioblastoma multiforme; LUSC: lung squamous cell carcinoma; PAAD: pancreatic adenocarcinoma; THCA: thyroid carcinoma; PCPG: pheochromocytoma and paraganglioma;

DLBC: lymphoid neoplasm diffuse large B-cell lymphoma; ACC: adrenocortical carcinoma; KIRP: kidney renal papillary cell carcinoma; SKCM: skin cutaneous melanoma; CESC: cervical squamous cell carcinoma and endocervical adenocarcinoma; KICH: kidney chromophobe; PRAD: prostate adenocarcinoma; BRCA: breast invasive carcinoma; MESO: mesothelioma; OV: ovarian serous cystadenocarcinoma; UCS: uterine carcinosarcoma; BLCA: bladder urothelial carcinoma; COAD: colon adenocarcinoma; UCEC: uterine corpus endometrial carcinoma; READ: rectum adenocarcinoma. The numbers in parentheses are the sample sizes of the indicated cancer types. The median, interquartile range and 95% range are shown for each setting, with outliers indicated by circles. **b)** Venn diagram showing the identity overlap (n=27) between genes identified in the NR3C1-activity signature and genes differentially expressed in LAM single cells. **c)** Representative images from immunohistochemistry assays for the detection of NR3C1 expression in LAM lung lesions of three patients (LAM #1-3). The arrows indicate the magnified lesion areas in the insets: in magnified lung nodules, epithelioid and spindle-like diseased cells are apparent from the observed nuclear shapes of positive NR3C1 staining. The positive control results from colon tissue showing positivity in the alveolar epithelium, and luminal and basal layers of the bronchioles, as well as those from normal lung tissue showing positivity in the alveolar epithelium, and luminal and basal layers of the bronchioles. **d)** Representative images of immunohistochemistry detection and colocalisation of NR3C1 and α SMA in LAM lung lesions; nuclei stained blue with DAPI (merged). Lung nodules of three LAM patients are shown.

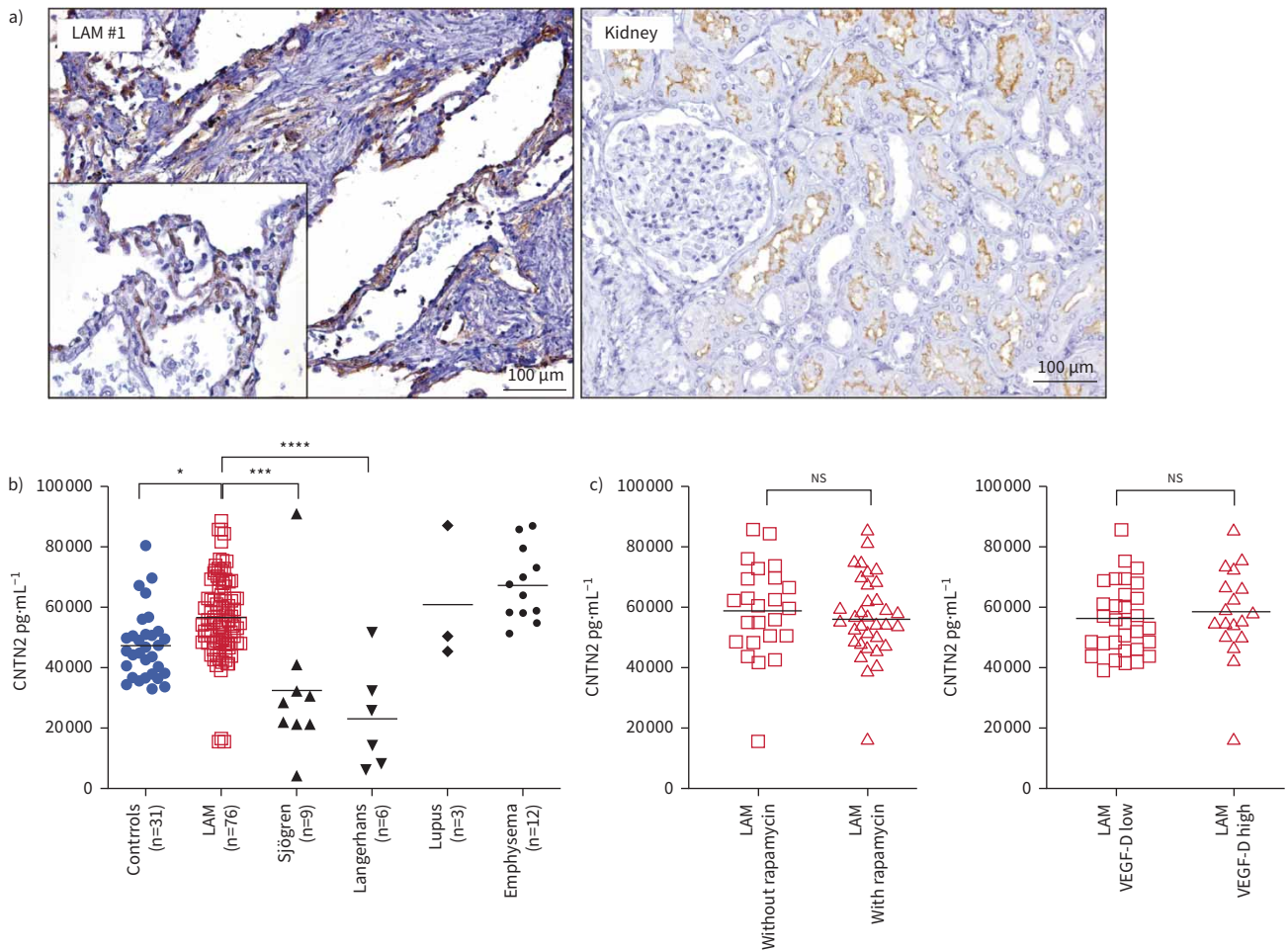


FIGURE 3 Evaluation of additional pleiotropic factors. **a)** Representative images from immunohistochemistry assays for detecting NTN4 expression in lymphangioleiomyomatosis (LAM) lung lesions. Lung nodules appear negative, while the alveolar layer is positive. The positive control of kidney tissue is shown. **b)** Overabundance of CNTN2 in LAM plasma relative to healthy women and two related pulmonary diseases, as indicated. The number of samples analysed in each setting (n) is indicated. Asterisks indicate significant differences based on two-sided Mann-Whitney tests (*p<0.05, ***p<0.001 and ****p<0.0001). Mean values are indicated with horizontal black lines. **c)** Absence of significant differences (n.s.: not significant) of CNTN2 plasma levels between LAM patients receiving and not receiving rapamycin treatment (left panel), and between LAM patients with high and low vascular endothelial growth factor-D (VEGF-D) plasma levels (right panel).

related pulmonary diseases. Using ELISA assays, we identified significant overabundance of CNTN2 in LAM plasma relative to healthy women, and to Langerhans cell histiocytosis and Sjögren syndrome patients; however, we found no differences with respect to emphysema and systemic lupus erythematosus patients (figure 3b). A comparison between LAM patients receiving and not receiving rapamycin treatment, and with low and high VEGF-D plasma levels (threshold of $800 \text{ pg}\cdot\text{mL}^{-1}$) revealed no significant differences (figure 3c). Subsequent examination of single-cell transcriptomic data of the human lung [46] identified CNTN2 expression exclusively in vascular smooth muscle cells (supplementary Figure S3). These cells also featured relatively high levels of expression of NR2F2 and NTN4; in contrast, NR3C1 was expressed in a wider range of lung cell types (supplementary Figure S3), consistent with the previous immunohistochemistry and immunohistofluorescence results (figure 2c,d).

Evaluation of LAM causality

Following on the identified shared genetics and pleiotropic factors, we used MR methods [73] to evaluate causality. Analysis of LAM as outcome did not show significant associations with cancer but suggested an association with FEV₁; this was supported by 245 genetic variants that depicted a LAM-FEV₁ negative correlation, as observed above, with no significant heterogeneity (supplementary Table S5, and Figures S4 and S5). However, a directionality test of causality showed inconsistent results with respect to the assumption that variance explained by the exposure should be greater than that of the outcome [36] (supplementary Table S5). Then, evaluation of LAM as exposure suggested negative correlations with bladder and endometrial cancer risk, and a positive correlation with FEV₁; in these comparisons, the directionality assumption was fulfilled, but it was based on eight to nine variants (supplementary Table S5).

Discussion

This study identifies genetic variants that may concurrently influence LAM and cancer risk, or LAM and pulmonary function. Interestingly, two loci on chromosomes 4 and 9 are shared between kidney cancer risk and LAM risk, and the kidney is one potential tissue of origin for LAM [6, 8, 19, 48, 74–77]. However, the gene candidates of these loci do not emerge as being linked to LAM single-cell transcriptome profiles [7, 8]. In contrast, the genetic connection between LAM and pulmonary function proved to be more relevant in several analyses. There were indications of substantial shared genetic architecture in LAM and FVC, and PEF measures, and a trend in the same direction was observed for FEV₁. In all comparisons, the LAM-pulmonary function correlation was negative, which is consistent with the expectation that a given genetic variant that associates with greater pulmonary function would have a corresponding lower risk of LAM, and vice versa. It is of note that the data of pulmonary function came from a population-based study [24]. In addition, the four pleiotropic gene candidates previously identified in a LAM single-cell signature (*ADAM12*, *BNC2*, *NR2F2* and *SP5*) [8] represented a higher proportion than would be expected by chance. However, the results of MR were not conclusive of a specific causal direction, likely because of the limited power and relative high variance of the LAM GWAS. These were also restrictions when examining genetic correlations, even if the HDL method [28] was used. There was an indication of LAM being causal of reduced FEV₁, which was expected, but a causal effect in the opposite direction remains unclear.

The biological function of some of the identified pleiotropic candidates appears consistent with LAM pathogenesis. NR3C1 (glucocorticoid receptor) can stimulate the transcriptional activity of NR2F2, which is the expected target of the primary LAM GWAS signal [20]. The activity of both NR3C1 and NR2F2 interacts with hormone signalling in health and disease [78, 79], and this might be connected to the role of oestrogens in LAM pathogenesis [80]. In addition, NR3C1 has been associated with breast cancer development and metastasis [63, 65], and may influence epithelial-to-mesenchymal transition, cell adhesion and tissue inflammation [53, 63, 81, 82]. The confirmation of NR3C1 expression in LAM lung lesions and diseased cells is further evidence of a role for this factor in disease development, as well as in pulmonary function. The indication of increased NR3C1 activity in single-cell LAM transcriptome profiles might anticipate a therapeutic benefit from specific NR3C1 antagonists, as proposed for other hormonal cancers [83]. It is of particular note that mifepristone, an antagonist of progesterone and glucocorticoid receptors yielded a preclinical benefit by inhibiting LAM tumorigenesis [84].

Other gene candidates were evaluated based on eQTL and genomic regulatory evidence. In a similar way to NR3C1, the identification of another gene linked to breast cancer risk, NTN4 [67], is particularly intriguing. These observations might in turn be connected to our previous results of a potential higher incidence of breast cancer diagnoses among LAM patients [48, 49]. We did not detect NTN4 expression in LAM-diseased cells, although a role for NTN4 influencing the tissue micro-environment cannot be excluded. Examining the other candidates, we found an overabundance of CNTN2 in LAM plasma, which,

in addition to its potential for revealing information about disease risk, raises the possibility of establishing an independent biomarker of VEGF-D. However, further analyses using samples collected at different times during the disease history are required to confirm these indications. Other pleiotropic gene candidates identified in this study might also be linked to LAM pathogenesis: IGF signalling and IGFBP2 function have been associated with LAM progression [85], though the expression of IGFBP3 in lung lesions is unclear [86]; and LAM lung lesions show expression of matrix metalloproteinases [87], which also promote disease progression [88, 89]. The LAM-FEV₁/FVC rs16937 and rs11240341 shared variants – which may target the transmembrane protein 81 (*TMEM81*) and/or retinoblastoma binding protein 5 (*RBBP5*) genes [90] – were previously associated with platelet measures in the general population [57] and alteration of this blood component is associated with several inflammatory lung diseases [91].

Collectively, this study proposes a common aetiology between LAM and pulmonary function. This connection may be due to genes whose function is particularly relevant in the cell(s) of origin of LAM as well as lung tissue development and/or may indicate a cell origin of LAM that resides in the lung cell populations. The latter hypothesis appears to be consistent with the recent demonstration that *Tsc2* loss in the lung mesenchymal lineage causes LAM-like disease in mice [7]. Thus, LAM might be an extreme phenotype of reduced lung function due to abnormal mTORC1-driven proliferation of mesenchymal-like cells. Particularly, NR3C1-glucocorticoid signalling regulates differentiation of proliferative mesenchymal progenitors into matrix fibroblasts [92], and these cells endorse synthesis of extracellular matrix and collagen during early lung development. The concept of altered mesenchymal cell differentiation and subsequent accumulation of extracellular matrix components would be consistent with the evolution of LAM lung pathology [93] and with identified pathway expression correlations with LAM^{core} [94]. Intriguingly, pleural mesothelioma cell lines have transcriptome profiles relatively similar to those of LAM cells [94]. In addition, the regulatory function of NR3C1 and NR2F2 may be coordinated during lung development [92]. Interestingly, GWAS signals of lung function were strongly predictive of COPD [24], which also shows extracellular matrix alterations and shares lymphatic vascular remodelling with LAM [95]. Several of the lung function signals were also found to be associated with other respiratory traits, including asthma [24], which is frequently diagnosed in LAM women, and with an inflammatory multisystem disease predominantly affecting women, systemic lupus erythematosus, which can be presented with cystic lung disease [1, 2]. Moreover, of the original 279 GWAS lung function signals [24], there were predicted 16 gene candidates that code for interactors of NR3C1, including the co-repressor NCOR1 [96] (supplementary Table S6).

The results of this study identify genetic factors and their molecular targets that may influence LAM development. However, as noted above, our study was limited by the relatively small sample size of the LAM GWAS, whose imputed summary statistics could also add noise to the results. Genetic studies of larger LAM cohorts are necessary to corroborate the findings and conclusively determine pleiotropy or causality with respect to pulmonary function and/or cancers. Likewise, studies of pleiotropic gene candidates may be warranted to better comprehend LAM aetiology and origin.

Provenance: Submitted article, peer reviewed.

Acknowledgements: We wish to thank all the participants of the lymphangioleiomyomatosis (LAM) genome-wide association study (GWAS) for their contribution. We also thank all the investigators who were involved in the other GWASs analysed in this study for making their results available to us.

Conflict of interest: None declared.

Support statement: This research was supported by Asociación Española de LAM; The LAM Foundation Seed Grant 2019; Carlos III Institute of Health grants PI18/01029, PI21/01306 and ICI19/00047 (co-funded by European Regional Development Fund (ERDF), “A way to build Europe”); Ministry of Economy and Competitiveness grant SAF2017-88457-R; the Generalitat de Catalunya SGR 2017-449 and 2017-529; PERIS PFI-Salut SLT017-20-000076; and the CERCA Program to IDIBELL and Institut Germans Trias i Pujol. X. Farré is supported by the VEIS project (001-P-001647, ERDF Operational Programme of Catalonia 2014–2020; co-funded by ERDF, “A way to build Europe”). Funding information for this article has been deposited with the Crossref Funder Registry.

References

- 1 Johnson SR, Cordier JF, Lazor R, *et al.* Review panel of the ERS LAM task force. European Respiratory Society guidelines for the diagnosis and management of lymphangioleiomyomatosis. *Eur Respir J* 2010; 35: 14–26.

- 2 McCormack FX, Gupta N, Finlay GR, *et al.* Official American Thoracic Society/Japanese Respiratory Society clinical practice guidelines: lymphangioleiomyomatosis diagnosis and management. *Am J Respir Crit Care Med* 2016; 194: 748–761.
- 3 Martignoni G, Pea M, Reghellin D, *et al.* Molecular pathology of lymphangioleiomyomatosis and other perivascular epithelioid cell tumors. *Arch Pathol Lab Med* 2010; 134: 33–40.
- 4 Harari S, Torre O, Cassandro R, *et al.* The changing face of a rare disease: lymphangioleiomyomatosis. *Eur Respir J* 2015; 46: 1471–1485.
- 5 Dongre A, Clements D, Fisher AJ, *et al.* Cathepsin K in lymphangioleiomyomatosis: LAM cell-fibroblast interactions enhance protease activity by extracellular acidification. *Am J Pathol* 2017; 187: 1750–1762.
- 6 Henske EP, McCormack FX. Lymphangioleiomyomatosis: a wolf in sheep's clothing. *J Clin Invest* 2012; 122: 3807–3816.
- 7 Obratsova K, Basil MC, Rue R, *et al.* mTORC1 activation in lung mesenchyme drives sex- and age-dependent pulmonary structure and function decline. *Nat Commun* 2020; 11: 5640.
- 8 Guo M, Yu JJ, Perl AK, *et al.* Single cell transcriptomic analysis identifies a unique pulmonary lymphangioleiomyomatosis cell. *Am J Respir Crit Care Med* 2020; 202: 1373–1387.
- 9 Crino PB, Nathanson KL, Henske EP. The tuberous sclerosis complex. *N Engl J Med* 2006; 355: 1345–1356.
- 10 Giannikou K, Malinowska IA, Pugh TJ, *et al.* Whole exome sequencing identifies *TSC1/TSC2* biallelic loss as the primary and sufficient driver event for renal angiomyolipoma development. *PLoS Genet* 2016; 12: e1006242.
- 11 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 USA* 2000; 97: 6085–6090.
- 12 McCormack FX, Inoue Y, Moss J, *et al.* Efficacy and safety of sirolimus in lymphangioleiomyomatosis. *N Engl J Med* 2011; 364: 1595–1606.
- 13 Krymskaya VP. Smooth muscle-like cells in pulmonary lymphangioleiomyomatosis. *Proc Am Thorac Soc* 2008; 5: 119–126.
- 14 Julian LM, Delaney SP, Wang Y, *et al.* Human pluripotent stem cell-derived *TSC2*-haploinsufficient smooth muscle cells recapitulate features of lymphangioleiomyomatosis. *Cancer Res* 2017; 77: 5491–5502.
- 15 de Garibay GR, Herranz C, Llorente A, *et al.* Lymphangioleiomyomatosis biomarkers linked to lung metastatic potential and cell stemness. *PLoS ONE* 2015; 10: e0132546.
- 16 Pacheco-Rodríguez G, Steagall WK, Samsel L, *et al.* Circulating lymphangioleiomyomatosis tumor cells with loss of heterozygosity in the *TSC2* gene show increased aldehyde dehydrogenase activity. *Chest* 2019; 156: 298–307.
- 17 McCormack FX, Travis WD, Colby TV, *et al.* Lymphangioleiomyomatosis: calling it what it is: a low-grade, destructive, metastasizing neoplasm. *Am J Respir Crit Care Med* 2012; 186: 1210–1212.
- 18 Liu H-J, Krymskaya VP, Henske EP. Immunotherapy for lymphangioleiomyomatosis and tuberous sclerosis: progress and future directions. *Chest* 2019; 156: 1062–1067.
- 19 Krymskaya VP, McCormack FX. Lymphangioleiomyomatosis: a monogenic model of malignancy. *Annu Rev Med* 2017; 68: 69–83.
- 20 Kim W, Giannikou K, Dreier JR, *et al.* A genome-wide association study implicates *NR2F2* in lymphangioleiomyomatosis pathogenesis. *Eur Respir J* 2019; 53: 1900329.
- 21 Litchfield LM, Klinge CM. Multiple roles of COUP-TFII in cancer initiation and progression. *J Mol Endocrinol* 2012; 49: R135–R148.
- 22 Polvani S, Pepe S, Milani S, *et al.* COUP-TFII in health and disease. *Cells* 2019; 9: 101.
- 23 Rashkin SR, Graff RE, Kachuri L, *et al.* Pan-cancer study detects genetic risk variants and shared genetic basis in two large cohorts. *Nat Commun* 2020; 11: 4423.
- 24 Shrine N, Guyatt AL, Erzurumluoglu AM, *et al.* New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries. *Nat Genet* 2019; 51: 481–493.
- 25 Rüeger S, McDaid A, Kutalik Z. Evaluation and application of summary statistic imputation to discover new height-associated loci. *PLoS Genet* 2018; 14: e1007371.
- 26 Rüeger S, McDaid A, Kutalik Z. Improved imputation of summary statistics for admixed populations. *BioRxiv* 2018; preprint [<https://doi.org/10.1101/203927>].
- 27 Smeland OB, Bahrami S, Frei O, *et al.* Genome-wide analysis reveals extensive genetic overlap between schizophrenia, bipolar disorder, and intelligence. *Mol Psychiatry* 2020; 25: 844–853.
- 28 Ning Z, Pawitan Y, Shen X. High-definition likelihood inference of genetic correlations across human complex traits. *Nat Genet* 2020; 52: 859–864.
- 29 Andreassen OA, Thompson WK, Schork AJ, *et al.* Improved detection of common variants associated with schizophrenia and bipolar disorder using pleiotropy-informed conditional false discovery rate. *PLoS Genet* 2013; 9: e1003455.
- 30 Liu JZ, Hov JR, Folseraas T, *et al.* Dense genotyping of immune-related disease regions identifies nine new risk loci for primary sclerosing cholangitis. *Nat Genet* 2013; 45: 670–675.

- 31 Schork AJ, Wang Y, Thompson WK, *et al.* New statistical approaches exploit the polygenic architecture of schizophrenia: implications for the underlying neurobiology. *Curr Opin Neurobiol* 2016; 36: 89–98.
- 32 Burgess S, Davey Smith G, Davies NM, *et al.* Guidelines for performing Mendelian randomisation investigations. *Wellcome Open Res* 2019; 4: 186.
- 33 Bowden J, Davey Smith G, Burgess S. Mendelian randomisation with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol* 2015; 44: 512–525.
- 34 Zhao Q, Wang J, Hemani G, *et al.* Statistical inference in two-sample summary-data Mendelian randomisation using robust adjusted profile score. *Ann Statist* 2020; 48: 1742–1769.
- 35 Higgins JPT, Thompson SG, Deeks JJ, *et al.* Measuring inconsistency in meta-analyses. *BMJ* 2003; 327: 557–560.
- 36 Hemani G, Tilling K, Davey Smith G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLoS Genet* 2017; 13: e1007081.
- 37 Hemani G, Zheng J, Elsworth B, *et al.* The MR-Base platform supports systematic causal inference across the human phenome. *eLife* 2018; 7: e34408.
- 38 Melé M, Ferreira PG, Reverter F, *et al.* Human genomics. The human transcriptome across tissues and individuals. *Science* 2015; 348: 660–665.
- 39 Wang Y, Song F, Zhang B, *et al.* The 3D genome browser: a web-based browser for visualising 3D genome organisation and long-range chromatin interactions. *Genome Biol* 2018; 19: 151.
- 40 Luck K, Kim D-K, Lambourne L, *et al.* A reference map of the human binary protein interactome. *Nature* 2020; 580: 402–408.
- 41 Oughtred R, Rust J, Chang C, *et al.* The BioGRID database: a comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Sci* 2020; 30: 187–200.
- 42 Butler A, Hoffman P, Smibert P, *et al.* Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* 2018; 36: 411–420.
- 43 Martin KR, Zhou W, Bowman MJ, *et al.* The genomic landscape of tuberous sclerosis complex. *Nat Commun* 2017; 8: 15816.
- 44 Liu J, Lichtenberg T, Hoadley KA, *et al.* An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics. *Cell* 2018; 173: 400–416.e11.
- 45 Cerami E, Gao J, Dogrusoz U, *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* 2012; 2: 401–404.
- 46 Travaglini KJ, Nabhan AN, Penland L, *et al.* A molecular cell atlas of the human lung from single-cell RNA sequencing. *Nature* 2020; 587: 619–625.
- 47 Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinforma* 2016; 32: 2847–2849.
- 48 Nuñez O, Román A, Johnson SR, *et al.* Study of breast cancer incidence in patients of lymphangioleiomyomatosis. *Breast Cancer Res Treat* 2016; 156: 195–201.
- 49 Nuñez O, Baldi BG, Radzikowska E, *et al.* Risk of breast cancer in patients with lymphangioleiomyomatosis. *Cancer Epidemiol* 2019; 61: 154–156.
- 50 Buniello A, MacArthur JAL, Cerezo M, *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res* 2019; 47: D1005–D1012.
- 51 Beck T, Hastings RK, Gollapudi S, *et al.* GWAS Central: A comprehensive resource for the comparison and interrogation of genome-wide association studies. *Eur J Hum Genet* 2014; 22: 949–952.
- 52 Shaffer JR, Li J, Lee MK, *et al.* Multiethnic GWAS reveals polygenic architecture of earlobe attachment. *Am J Hum Genet* 2017; 101: 913–924.
- 53 Michailidou K, Beesley J, Lindstrom S, *et al.* Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. *Nat Genet* 2015; 47: 373–380.
- 54 Liyanage UE, Law MH, Han X, *et al.* Combined analysis of keratinocyte cancers identifies novel genome-wide loci. *Hum Mol Genet* 2019; 28: 3148–3160.
- 55 Study COGENT, Houlston RS, Webb E, Broderick P, *et al.* Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer *Nat Genet* 2008; 40: 1426–1435.
- 56 Jeong C, Witonsky DB, Basnyat B, *et al.* Detecting past and ongoing natural selection among ethnically Tibetan women at high altitude in Nepal. *PLoS Genet* 2018; 14: e1007650.
- 57 Astle WJ, Elding H, Jiang T, *et al.* The allelic landscape of human blood cell trait variation and links to common complex disease. *Cell* 2016; 167: 1415–1429.e19.
- 58 Stefansson H, Ophoff RA, Steinberg S, *et al.* Common variants conferring risk of schizophrenia. *Nature* 2009; 460: 744–747.
- 59 De Martino MU, Bhattacharyya N, Alesci S, *et al.* The glucocorticoid receptor and the orphan nuclear receptor chicken ovalbumin upstream promoter-transcription factor II interact with and mutually affect each other's transcriptional activities: implications for intermediary metabolism. *Mol Endocrinol* 2004; 18: 820–833.

- 60 Anderson TMR, Ma SH, Raj GV, *et al.* Breast tumor kinase (Brk/PTK6) is induced by HIF, glucocorticoid receptor, and PELP1-mediated stress signaling in triple-negative breast cancer. *Cancer Res* 2016; 76: 1653–1663.
- 61 Biadasiewicz K, Fock V, Dekan S, *et al.* Extravillous trophoblast-associated ADAM12 exerts pro-invasive properties, including induction of integrin beta 1-mediated cellular spreading. *Biol Reprod* 2014; 90: 101.
- 62 Pittman AM, Naranjo S, Jalava SE, *et al.* Allelic variation at the 8q23.3 colorectal cancer risk locus functions as a cis-acting regulator of *EIF3H*. *PLoS Genet* 2010; 6: e1001126.
- 63 Chen Z, Lan X, Wu D, *et al.* Ligand-dependent genomic function of glucocorticoid receptor in triple-negative breast cancer. *Nat Commun* 2015; 6: 8323.
- 64 Sorrentino G, Ruggeri N, Zannini A, *et al.* Glucocorticoid receptor signalling activates YAP in breast cancer. *Nat Commun* 2017; 8: 14073.
- 65 Obradović MMS, Hamelin B, Manevski N, *et al.* Glucocorticoids promote breast cancer metastasis. *Nature* 2019; 567: 540–544.
- 66 West DC, Kocherginsky M, Tonsing-Carter EY, *et al.* Discovery of a glucocorticoid receptor (GR) activity signature using selective GR antagonism in ER-negative breast cancer. *Clin Cancer Res* 24: 3433–3446.
- 67 Beesley J, Sivakumaran H, Moradi Marjaneh M, *et al.* eQTL colocalisation analyses identify *NTN4* as a candidate breast cancer risk gene. *Am J Hum Genet* 2020; 107: 778–787.
- 68 Kociok N, Crespo-Garcia S, Liang Y, *et al.* Lack of netrin-4 modulates pathologic neovascularisation in the eye. *Sci Rep* 2016; 6: 18828.
- 69 Xu X, Yan Q, Wang Y, *et al.* *NTN4* is associated with breast cancer metastasis via regulation of EMT-related biomarkers. *Oncol Rep* 2017; 37: 449–457.
- 70 Derfuss T, Parikh K, Velhin S, *et al.* Contactin-2/TAG-1-directed autoimmunity is identified in multiple sclerosis patients and mediates gray matter pathology in animals. *Proc Natl Acad Sci U S A* 2009; 106: 8302–8307.
- 71 Abou-Kandil A, Eisa N, Jabareen A, *et al.* Differential effects of HTLV-1 Tax oncoprotein on the different estrogen-induced-ER α -mediated transcriptional activities. *Cell Cycle* 2016; 15: 2626–2635.
- 72 Yin GN, Lee HW, Cho J-Y, *et al.* Neuronal pentraxin receptor in cerebrospinal fluid as a potential biomarker for neurodegenerative diseases. *Brain Res* 2009; 1265: 158–170.
- 73 Smith GD, Ebrahim S. “Mendelian randomisation”: can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 2003; 32: 1–22.
- 74 Delaney SP, Julian LM, Stanford WL. The neural crest lineage as a driver of disease heterogeneity in tuberous sclerosis complex and lymphangioleiomyomatosis. *Front Cell Dev Biol* 2014; 2: 69.
- 75 Yue M, Pacheco G, Cheng T, *et al.* Evidence supporting a lymphatic endothelium origin for angiomyolipoma, a TSC2-tumor related to lymphangioleiomyomatosis. *Am J Pathol* 2016; 186: 1825–1836.
- 76 Hayashi T, Kumasaka T, Mitani K, *et al.* Prevalence of uterine and adnexal involvement in pulmonary lymphangioleiomyomatosis: a clinicopathologic study of 10 patients. *Am J Surg Pathol* 2011; 35: 1776–1785.
- 77 Prizant H, Sen A, Light A, *et al.* Uterine-specific loss of *Tsc2* leads to myometrial tumors in both the uterus and lungs. *Mol Endocrinol* 2013; 27: 1403–1414.
- 78 Zhao F, Franco HL, Rodriguez KF, *et al.* Elimination of the male reproductive tract in the female embryo is promoted by COUP-TFII in mice. *Science* 2017; 357: 717–720.
- 79 Dwyer AR, Truong TH, Ostrander JH, *et al.* 90 years of progesterone: steroid receptors as MAPK signaling sensors in breast cancer: let the fates decide. *J Mol Endocrinol* 2020; 65: T35–T48.
- 80 Prizant H, Hammes SR. Lymphangioleiomyomatosis (LAM): the “other” steroid-sensitive cancer. *Endocrinology* 2016; 157: 3374–3383.
- 81 Pan D, Kocherginsky M, Conzen SD. Activation of the glucocorticoid receptor is associated with poor prognosis in estrogen receptor-negative breast cancer. *Cancer Res* 2011; 71: 6360–6370.
- 82 Wu F, Chen W, Kang X, *et al.* A seven-nuclear receptor-based prognostic signature in breast cancer. *Clin Transl Oncol* 2020; 23: 1292–1303.
- 83 Kach J, Conzen SD, Szmulewitz RZ. Glucocorticoid receptor signaling in breast and prostate cancers: emergence as a therapeutic target. *Sci Transl Med* 2015; 7: 305ps19.
- 84 Sun Y, Zhang E, Lao T, *et al.* Progesterone and estradiol synergistically promote the lung metastasis of tuberlin-deficient cells in a preclinical model of lymphangioleiomyomatosis. *Horm Cancer* 2014; 5: 284–298.
- 85 Li X, Liu X, Zhang L, *et al.* Insulin growth factor binding protein 2 mediates the progression of lymphangioleiomyomatosis. *Oncotarget* 2017; 8: 36628–36638.
- 86 Valencia JC, Matsui K, Bondy C, *et al.* Distribution and mRNA expression of insulin-like growth factor system in pulmonary lymphangioleiomyomatosis. *J Investig Med* 2001; 49: 421–433.
- 87 Matsui K, Takeda K, Yu ZX, *et al.* Role for activation of matrix metalloproteinases in the pathogenesis of pulmonary lymphangioleiomyomatosis. *Arch Pathol Lab Med* 2000; 124: 267–275.
- 88 Glassberg MK, Elliot SJ, Fritz J, *et al.* Activation of the estrogen receptor contributes to the progression of pulmonary lymphangioleiomyomatosis via matrix metalloproteinase-induced cell invasiveness. *J Clin Endocrinol Metab* 2008; 93: 1625–1633.

- 89 Goncharova EA, Goncharov DA, Fehrenbach M, *et al.* Prevention of alveolar destruction and airspace enlargement in a mouse model of pulmonary lymphangiomyomatosis (LAM). *Sci Transl Med* 2012; 4: 154ra134.
- 90 Huo Y, Li S, Liu J, *et al.* Functional genomics reveal gene regulatory mechanisms underlying schizophrenia risk. *Nat Commun* 2019; 10: 670.
- 91 Middleton EA, Weyrich AS, Zimmerman GA. Platelets in pulmonary immune responses and inflammatory lung diseases. *Physiol Rev* 2016; 96: 1211–1259.
- 92 Bridges JP, Sudha P, Lipps D, *et al.* Glucocorticoid regulates mesenchymal cell differentiation required for perinatal lung morphogenesis and function. *Am J Physiol Lung Cell Mol Physiol* 2020; 319: L239–L255.
- 93 Miller S, Stewart ID, Clements D, *et al.* Evolution of lung pathology in lymphangiomyomatosis: associations with disease course and treatment response. *J Pathol Clin Res* 2020; 6: 215–226.
- 94 Espín R, Baiges A, Blommaert E, *et al.* Heterogeneity and cancer-related features in lymphangiomyomatosis cells and tissue. *Mol Cancer Res* 2021; in press [<https://doi.org/10.1158/1541-7786.MCR-21-0220>].
- 95 Hardavella G, Tzortzaki EG, Siozopoulou V, *et al.* Lymphangiogenesis in COPD: another link in the pathogenesis of the disease. *Respir Med* 2012; 106: 687–693.
- 96 Schulz M, Eggert M, Baniahmad A, *et al.* RU486-induced glucocorticoid receptor agonism is controlled by the receptor N terminus and by corepressor binding. *J Biol Chem* 2002; 277: 26238–26243.