COL18A1 genotypic associations with endostatin levels and clinical features in pulmonary arterial hypertension: a quantitative trait association study

Received: 28 Dec 2021
Accepted: 29 April 2022

To the Editor:

Endostatin (ES) is a circulating peptide derived from collagen XVIII alpha 1 (COL18A1) known to inhibit angiogenesis [1, 2]. Decreased angiogenesis is a feature of pulmonary arterial hypertension (PAH) in animal models [3] and human subjects [4]. Our group has reported strong associations between circulating ES levels and haemodynamics and survival in PAH [5–7]. We have also reported that a missense variant in *COL18A1*, which encodes ES, confers lower ES and longer survival, suggesting that variation within the gene contributes to circulating levels [5]. In the current study, we assessed *COL18A1* variant associations with clinical phenotypes and outcomes, including *COL18A1* associations with circulating ES levels, in a large, multicentre PAH cohort in which we previously investigated ES as a prognostic biomarker [6].

This study was approved by the Johns Hopkins University Institutional Review Board. Serum samples contributed to the National Heart, Lung, and Blood Institute-sponsored PAH Biobank underwent single nucleotide polymorphism (SNP) genotyping using the Omni5-4 BeadChip (Illumina) and whole exome sequencing (WES) through the Regeneron Genetics Center [8]. An electrochemiluminescence assay was developed to quantitate ES. Sample collection and processing methods have been previously published [6, 9, 10].

ES measurements were regressed on genotypes of *COL18A1* variants to perform a multivariable protein quantitative trait locus (pQTL) analysis. Linear regression models adjusted for age and sex were restricted to subjects of European ancestry (EA) or African ancestry (AA). To determine whether ES-associated SNPs also affected regulation of *COL18A1* gene expression, pQTLs were queried in a publicly available expression quantitative trait locus (eQTL) database of whole-blood RNA samples [11]. Associations with clinical phenotypes and survival were modelled using multivariable linear and Cox regressions. Minor allele frequencies (MAFs) for *COL18A1* variants were compared with the Genome Aggregation Database (https://gnomad.broadinstitute.org). Linkage disequilibrium (LD) across the *COL18A1* region was assessed using D′ [12]. A p-value of <0.05 was considered nominally significant. An LD-adjusted correction for multiple testing was applied for QTL analyses equalling 0.0016 in EA and 0.0013 in AA subjects. Statistical tests were performed using Stata version 15.1 (StataCorp., College Station, TX, USA), SAS version 9.4 (SAS Institute, Cary, NC, USA) and PLINK version 1.9 (http://pngu.mgh.harvard.edu/purcell/plink) [13].

The cohort consisted of 2017 subjects with median age 53 years, of whom 80% were female and 82% were EA subjects. Full clinical characteristics of this cohort have been published previously [6]. Briefly, subjects had prevalent disease (median duration at sample collection 48 months, interquartile range 14–92 months) and moderately severe PAH at enrolment, with mean±SD pulmonary artery pressure 50±15 mmHg, pulmonary vascular resistance 10±6 Wood units, and 45% with New York Heart Association functional class III or IV symptoms. Most subjects had idiopathic PAH (n=870) or connective tissue disease-associated PAH (n=623). From the Omni5 SNP array, 100 *COL18A1* variants in 1400 EA subjects and 126 *COL18A1* variants in 209 AA subjects passed quality control (Hardy–Weinberg equilibrium (HWE) >0.001, MAF >0.05 and genotype missing rate <5%), with 91 variants present in both EA and AA subjects. In multivariable pQTL analysis, 26 cis-acting SNPs were associated with ES levels in EA

Shareable abstract (@ERSpublications)
Variation around the *COL18A1* gene, which encodes the angiostatic peptide endostatin, may influence disease heterogeneity in pulmonary arterial hypertension https://bit.ly/3shXrNR

Cite this article as: Simpson CE, Griffiths M, Yang J, *et al*. *COL18A1* genotypic associations with endostatin levels and clinical features in pulmonary arterial hypertension: a quantitative trait association study. *ERJ Open Res* 2022; 8: 00725-2021 [DOI: 10.1183/23120541.00725-2021].

TABLE 1 *COL18A1* variants associated with serum endostatin (ES) levels, gene expression and clinical measures

QTL data	ES β[#]	ES p-value	eQTL β[¶]	eQTL p-value	FDR
EA subjects					
Omni5 array					
rs9976834	−3510	0.0001983	−0.66	2.22×10^{-197}	$<1.34\times10^{-5}$
rs9976784	−3205	0.0003702	−0.666	2.27×10^{-200}	$<1.34\times10^{-5}$
rs2838932	−3137	0.0004473	−0.636	1.17×10^{-184}	$<1.34\times10^{-5}$
rs2236461	−3089	0.0004758	−0.643	4.78×10^{-191}	$<1.34\times10^{-5}$
rs9977482	−2903	0.0006399	−0.705	9.81×10^{-254}	$<1.34\times10^{-5}$
rs2236470	−2864	0.0007104	−0.707	2.57×10^{-255}	$<1.34\times10^{-5}$
rs8133622	−2988	0.0007197	−0.632	1.27×10^{-182}	$<1.34\times10^{-5}$
rs9980531	−2947	0.0008603	−0.625	5.94×10^{-177}	$<1.34\times10^{-5}$
rs11702782	−3139	0.001041	−0.584	6.75×10^{-37}	$<1.34\times10^{-5}$
rs12482088[+]	−2758	0.001153	−0.695	1.02×10^{-248}	$<1.34\times10^{-5}$
rs201577993	−2552	0.001393	NA	NA	NA
rs2330180	−2898	0.001514	−0.632	5.80×10^{-180}	$<1.34\times10^{-5}$
rs2236459	−2475	0.001694	−0.513	1.35×10^{-143}	$<1.34\times10^{-5}$
rs2236464	−2677	0.002145	−0.654	1.06×10^{-196}	$<1.34\times10^{-5}$
rs11702494	−2567	0.00237	−0.674	1.76×10^{-231}	$<1.34\times10^{-5}$
rs2236451[+]	−2289	0.00332	−0.55	1.20×10^{-176}	$<1.34\times10^{-5}$
rs55684533	−2479	0.004203	NA	NA	NA
rs10854470	−1915	0.01018	−0.445	1.32×10^{-122}	$<1.34\times10^{-5}$
rs2236454	−1902	0.01065	−0.465	9.84×10^{-134}	$<1.34\times10^{-5}$
rs4819124	−1802	0.01436	−0.367	1.74×10^{-86}	$<1.34\times10^{-5}$
rs61633029	−1864	0.02498	−0.473	1.06×10^{-115}	$<1.34\times10^{-5}$
rs17338076	−2409	0.02935	−0.674	1.38×10^{-34}	$<1.34\times10^{-5}$
rs2236479	−1641	0.03106	−0.455	5.77×10^{-127}	$<1.34\times10^{-5}$
rs2150443	−1544	0.03625	−0.367	2.40×10^{-86}	$<1.34\times10^{-5}$
rs7409857	−1531	0.03885	−0.32	6.64×10^{-65}	$<1.34\times10^{-5}$
rs7281138	1752	0.04298	NA	NA	NA
WES					
rs9979845	−2675	0.0013	−0.705	9.81×10^{-254}	$<1.34\times10^{-5}$
rs11702425	−2107	0.0058	−0.532	4.67×10^{-157}	$<1.34\times10^{-5}$
rs749627	−1682	0.022	−0.304	6.79×10^{-58}	$<1.34\times10^{-5}$
AA subjects					
Omni5 array					
rs4819124	−5598	0.01829	−0.367	1.74×10^{-86}	$<1.34\times10^{-5}$
rs2150443	−5415	0.0219	−0.367	2.40×10^{-86}	$<1.34\times10^{-5}$
rs73370840	6899	0.02508	0.31	1.30×10^{-23}	$<1.34\times10^{-5}$
rs2838917	−4931	0.02828	0.153	7.35×10^{-13}	$<1.34\times10^{-5}$
rs114255716	10 260	0.0306	NA	NA	NA
rs78620106	10 810	0.0308	NA	NA	NA
rs61633029	5548	0.03098	−0.473	1.06×10^{-115}	$<1.34\times10^{-5}$
rs56327327	−4398	0.04565	NA	NA	NA
WES					
rs749627	5172	0.025	−0.304	6.79×10^{-58}	$<1.34\times10^{-5}$

Phenotypic data[§]	Clinical measure	Effect estimate[ƒ]	p-value
EA subjects			
rs7499	Survival	0.77 (0.62–0.96)[¶¶]	0.018
rs1050351	Survival	0.76 (0.61–0.95)[¶¶]	0.015
rs1131100	6MWD (m)	30.61 (2.09–59.13)[++]	0.035
rs1131101	6MWD (m)	30.61 (2.09–59.13)[++]	0.035
rs2236467	6MWD (m)	30.16 (1.80–58.53)[++]	0.037
rs1131102	6MWD (m)	30.11 (1.72–58.49)[++]	0.038
rs2236466	6MWD (m)	28.49 (−0.02–57.01)[++]	0.050
rs7281138[##]	Cardiac index ($L\cdot min^{-1}\cdot m^{-2}$)	−0.11 (−0.23– −0.004)[§§]	0.043
rs2838917[##]	Cardiac index ($L\cdot min^{-1}\cdot m^{-2}$)	−0.12 (−0.23– −0.01)[§§]	0.028
rs2230688	Cardiac index ($L\cdot min^{-1}\cdot m^{-2}$)	0.19 (0.04–0.33)[§§]	0.010
rs2230687	Cardiac index ($L\cdot min^{-1}\cdot m^{-2}$)	0.19 (0.05–0.33)[§§]	0.009
rs2236456	Cardiac index ($L\cdot min^{-1}\cdot m^{-2}$)	0.14 (0.002–0.28)[§§]	0.047

Continued

TABLE 1 Continued

QTL: quantitative trait locus; eQTL: expression QTL; FDR: false discovery rate; EA: European ancestry; WES: whole exome sequencing; AA: African ancestry; 6MWD: 6-min walk distance; NA: no association between variant and gene expression. [#]: ES β-coefficients (protein QTL analysis) reflect differences in ES in pg·mL^{-1} for each copy of the minor allele; linear regressions on ES levels are adjusted for age and sex. [¶]: eQTL β-coefficients reflect differences in robust multi-array analysis, a measure of intensity derived from Affymetrix gene expression data; methods for eQTL models have been previously published [11]. [+]: single nucleotide polymorphism (SNP) that appears on both Omni5 and WES arrays. [§]: phenotypic data are reported for EA only. [f]: effect estimates are hazard ratios for associations with survival and β-coefficients for associations with all other clinical measures; coefficients reflect differences in clinical measures for subjects with the presence *versus* the absence of the minor allele. [##]: SNP from Omni5 array; all others in the phenotypic data section are WES SNPs. [¶¶]: associations with survival are adjusted for age at enrolment, sex, pulmonary arterial hypertension (PAH) subtype and PAH therapies, and additionally for difference in time from PAH diagnosis to cohort enrolment. [++]: associations with 6MWD are adjusted for body mass index and the following comorbid conditions: hypertension, diabetes, obstructive lung disease, cardiomyopathy and chronic kidney disease. [§§]: associations with cardiac index are adjusted for age at enrolment, sex, PAH subtype and PAH therapies.

individuals, and eight were associated with ES levels in AA individuals. There were no pQTLs in common for EA and AA subjects. 23 of 26 pQTLs in EA and five of eight pQTLs in AA individuals were associated with differences in *COL18A1* gene expression. In EA subjects, two Omni5 SNPs demonstrated associations with cardiac index: the T allele was associated with a 0.11 L·min^{-1}·m^{-2} lower cardiac index for rs7281138 (p=0.043), and a 0.12 L·min^{-1}·m^{-2} lower cardiac index for rs2838917 (p=0.028). QTL data and genotype–phenotype associations are shown in table 1.

Of 102 *COL18A1* WES variants, 22 had a frequency of ⩾5%; none deviated from HWE. Six SNPs overlapped between exonic SNPs and Omni5 SNPs. Three of the 16 unique exonic variants in EA and one in AA subjects were associated with differences in serum ES. All four exonic pQTLs identified were also associated with significant differences in *COL18A1* gene expression in eQTL analysis. In EA subjects, two exome variants demonstrated associations with survival: the A allele was associated with 23% lower mortality for rs7499 (hazard ratio 0.77, 95% CI 0.62–0.96; p=0.018), and the A allele was associated with 24% lower mortality for rs1050351 (hazard ratio 0.76, 95% CI 0.61–0.95; p=0.015). Five exonic variants with chromosomal positions in close proximity were associated with longer 6-min walk distance, and an additional three exonic variants, also in close proximity, were associated with higher cardiac index (table 1). There were no observed differences in MAFs of *COL18A1* variants compared to available controls.

Our QTL results suggest circulating ES levels are partially genetically influenced by variants in and around *COL18A1*. The eQTL results suggest some variation in ES abundance may be due to variations in mRNA expression. Most known QTLs are associated with changes in mRNA expression, with downstream effects on ribosome occupancy and protein abundance [14]. Thus, eQTLs often have smaller effect sizes on protein levels than on gene expression [14, 15], consistent with our results. We found some signal for genetically influenced phenotypic variation, although none of the phenotypically associated variants were ES-associated pQTLs, and all but one (rs7499 in the 3′ untranslated region) were synonymous variants. Interestingly, rs7499 has been associated with significantly reduced risk of hepatocellular carcinomas in patients with hepatitis B infection [16], suggesting some biological significance of this variant in humans.

In contrast to our 2015 report [5], we did not find an association between rs12483377 and ES levels or outcomes. This discrepancy may be due to the smaller sample size in the first study. Genotype at rs12483377 was not associated with survival in two large PAH genome-wide association studies (GWAS) later published [17], although these GWAS excluded patients with connective tissue disease and may have investigated genetically different cohorts.

This study has several limitations. We are limited by the cohort size available for a rare disease; consequently, some of our results are of nominal significance, with a higher likelihood of observation due to chance alone. The QTL results are based on associations in whole blood, as mRNA or protein expression data from human tissues most relevant to disease are not available. The genetic and clinical associations with ES are based on a single time point for each subject. Furthermore, Omni5 genotyping and WES leaves many genetic variants uncharacterised. Therefore, the identified genotype–phenotype associations may not be causal themselves, but rather in LD with true, unidentified QTLs.

Aside from reports on BMPR2 (bone morphogenetic protein receptor type 2), our study is one of only a few [17, 18] that offer insights into genetic influences on disease severity and heterogeneity in PAH, a strength of our work. Heritable modifiers of phenotype have not been well-established in PAH, and most genetic studies have focused on identifying loci contributing to disease susceptibility, rather than disease severity or prognosis.

In conclusion, these results suggest that PAH disease heterogeneity is influenced in part by genetic variation around the *COL18A1* gene. ES levels have been linked to variation in PAH severity and outcomes, and our results suggest that ES levels may be genetically influenced. Understanding influences on transcription and translation of genes implicated in disease can clarify therapeutic targeting strategies. Future work on ES/COL18A1 is needed to better understand genetic and cellular mechanisms underlying PAH pathobiology.

Catherine E. Simpson [1], Megan Griffiths[2], Jun Yang[2], Melanie K. Nies[2], Dhananjay Vaidya[3], Stephanie Brandal[2], Lisa J. Martin[4], Michael W. Pauciulo[4], Katie A. Lutz[4], Anna W. Coleman[4], Eric D. Austin [5], D. Dunbar Ivy[6], William C. Nichols[4], Allen D. Everett[2], Paul M. Hassoun[1] and Rachel L. Damico[1]

[1]Johns Hopkins University, Dept of Medicine, Division of Pulmonary and Critical Care Medicine, Baltimore, MD, USA. [2]Johns Hopkins University, Dept of Pediatrics, Division of Pediatric Cardiology, Baltimore, MD, USA. [3]Johns Hopkins University, Dept of Medicine, Division of General Internal Medicine, Baltimore, MD, USA. [4]Cincinnati Children's Hospital Medical Center and University of Cincinnati College of Medicine, Dept of Pediatrics, Division of Human Genetics, Cincinnati, OH, USA. [5]Vanderbilt University, Dept of Pediatrics, Division of Allergy, Immunology, and Pulmonary Medicine, Nashville, TN, USA. [6]Children's Hospital Colorado, Dept of Pediatric Cardiology, Aurora, CO, USA.

Corresponding author: Rachel L. Damico (rdamico1@jhmi.edu)

Provenance: Submitted article, peer reviewed.

Acknowledgements: Exome sequencing and genotyping data were generated by Regeneron Genetics Center, Regeneron Pharmaceuticals, Tarrytown, NY, USA.

Author contributions: C.E. Simpson, R.L. Damico, P.M. Hassoun and A.D. Everett designed the study; S. Brandal and J. Yang performed the experiments and interpreted the results; C.E. Simpson, M. Griffiths, J. Yang, M.K. Nies, M.W. Pauciulo, E.D. Austin, D.D. Ivy and W.C. Nichols performed data collection, maintenance and analysis; C.E. Simpson, L.J. Martin and D. Vaidya performed statistical analyses; C.E. Simpson drafted the manuscript; all authors critically revised the manuscript for important intellectual content and approved the final version; P.M. Hassoun and R.L. Damico had access to all the data in the study and take full responsibility for the integrity and accuracy of the work.

Conflict of interest: C.E. Simpson reports support for the present manuscript received from NIH/NHLBI grants. M. Griffiths reports support for the present manuscript received from NIH/NHLBI grants. J. Yang has nothing to disclose. M.K. Nies has nothing to disclose. D. Vaidya reports receiving grants or contacts outside the submitted work from National Institutes of Health. S. Brandal has nothing to disclose. L.J. Martin reports support for the present manuscript received from NIH/NHLBI grants. M.W. Pauciulo has nothing to disclose. K.A. Lutz has nothing to disclose. A.W. Coleman has nothing to disclose. E.D. Austin reports support for the present manuscript received from NIH/NHLBI grants; and grants or contracts received from CMREF, outside the submitted work. D.D. Ivy reports support for the present manuscript received from NIH/NHLBI grants; and consulting fees received from The University of Colorado contracts with Actelion, Altavant, Bayer, and Gossamer Bio, outside the submitted work. W.C. Nichols reports support for the present manuscript received from NIH/NHLBI grants. A.D. Everett reports support for the present manuscript received from NIH/NHLBI grants. P.M. Hassoun reports receiving grants or contacts outside the submitted work from NIH/NHLBI; and he serves on a scientific Advisory Board for MSD, unrelated to the current work. R.L. Damico reports support for the present manuscript received from NIH/NHLBI grants.

Support statement: This study was supported by National Institutes of Health/National Heart, Lung, and Blood Institute awards R01HL135114 and R01HL150070 (A.D. Everett, M.K. Nies, J. Yang, R.L. Damico, D. Vaidya,

W.C. Nichols, D.D. Ivy and E.D. Austin), R01HL132153 (R.L. Damico and P.M. Hassoun), R24HL105333 (W.C. Nichols and M.W. Pauciulo), K12-HD000850 (M. Griffiths) and K23HL153781 (C.E. Simpson). Funding information for this article has been deposited with the Crossref Funder Registry.

References

1 Taddei L, Chiarugi P, Brogelli L, *et al.* Inhibitory effect of full-length human endostatin on *in vitro* angiogenesis. *Biochem Biophys Res Commun* 1999; 263: 340–345.
2 O'Reilly MS, Boehm T, Shing Y, *et al.* Endostatin: an endogenous inhibitor of angiogenesis and tumor growth. *Cell* 1997; 88: 277–285.
3 Sutendra G, Dromparis P, Paulin R, *et al.* A metabolic remodeling in right ventricular hypertrophy is associated with decreased angiogenesis and a transition from a compensated to a decompensated state in pulmonary hypertension. *J Mol Med* 2013; 91: 1315–1327.
4 Vogel-Claussen J, Shehata ML, Lossnitzer D, *et al.* Increased right ventricular septomarginal trabeculation mass is a novel marker for pulmonary hypertension: comparison with ventricular mass index and right ventricular mass. *Invest Radiol* 2011; 46: 567–575.
5 Damico R, Kolb TM, Valera L, *et al.* Serum endostatin is a genetically determined predictor of survival in pulmonary arterial hypertension. *Am J Respir Crit Care Med* 2015; 191: 208–218.
6 Simpson CE, Griffiths M, Yang J, *et al.* The angiostatic peptide endostatin enhances mortality risk prediction in pulmonary arterial hypertension. *ERJ Open Res* 2021; 7: 00378-2021.
7 Daly CM, Griffiths M, Simpson CE, *et al.* Angiostatic peptide, endostatin, predicts severity in pediatric congenital heart disease-associated pulmonary hypertension. *J Am Heart Assoc* 2021; 10: e021409.
8 Zhu N, Pauciulo MW, Welch CL, *et al.* Novel risk genes and mechanisms implicated by exome sequencing of 2572 individuals with pulmonary arterial hypertension. *Genome Med* 2019; 11: 69.
9 Simpson CE, Chen JY, Damico RL, *et al.* Cellular sources of interleukin-6 and associations with clinical phenotypes and outcomes in pulmonary arterial hypertension. *Eur Respir J* 2020; 55: 1901761.
10 Simpson CE, Damico RL, Hassoun PM, *et al.* Noninvasive prognostic biomarkers for left-sided heart failure as predictors of survival in pulmonary arterial hypertension. *Chest* 2020; 157: 1606–1616.
11 Jansen R, Hottenga JJ, Nivard MG, *et al.* Conditional eQTL analysis reveals allelic heterogeneity of gene expression. *Hum Mol Genet* 2017; 26: 1444–1451.
12 Slatkin M. Linkage disequilibrium – understanding the evolutionary past and mapping the medical future. *Nat Rev Genet* 2008; 9: 477–485.
13 Purcell S, Neale B, Todd-Brown K, *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; 81: 559–575.
14 Battle A, Khan Z, Wang SH, *et al.* Genomic variation. Impact of regulatory variation from RNA to protein. *Science* 2015; 347: 664–667.
15 Li JJ, Biggin MD. Gene expression. Statistics requantitates the central dogma. *Science* 2015; 347: 1066–1067.
16 Wu X, Wu J, Xin Z, *et al.* A 3′ UTR SNP in COL18A1 is associated with susceptibility to HBV related hepatocellular carcinoma in Chinese: three independent case-control studies. *PLoS One* 2012; 7: e33855.
17 Rhodes CJ, Batai K, Bleda M, *et al.* Genetic determinants of risk in pulmonary arterial hypertension: international genome-wide association studies and meta-analysis. *Lancet Respir Med* 2019; 7: 227–238.
18 Rhodes CJ, Otero-Núñez P, Wharton J, *et al.* Whole-blood RNA profiles associated with pulmonary arterial hypertension and clinical outcome. *Am J Respir Crit Care Med* 2020; 202: 586–594.