Familial Pulmonary Fibrosis and Hermansky-Pudlak Syndrome Rare Missense Mutations in Context

To the Editor:

Hermansky-Pudlak syndrome (HPS) is an autosomal-recessive disease characterized by a constellation of phenotypes related to the proper functioning of lysosome-related organelles, including dense platelet granules. Currently, 10 *HPS* genes (*HPS1*–*HPS10*) have been identified in humans. Interestingly, HPS-associated pulmonary fibrosis (PF) is associated with complete penetrance in *HPS1*, *HPS2*, and *HPS4*, but lung disease has not been reported in other HPS subtypes (1). An initial study genetically mapped the *HPS1* locus in Puerto Rican families and discovered a founder-effect mutation (16 bp duplication) (2); however, other *HPS1* variants have been identified worldwide. The prevalence of PF in HPS carriers is an open question, as a detailed study examining this relationship, especially focused on milder phenotypes, has not been completed.

We reasoned that *HPS1*, *HPS2*, or *HPS4* mutations in heterozygote carriers may be important in PF presentation. We focused on *HPS1* and *HPS4* as target genes because they form the BLOC-3 complex, which is required for proper protein trafficking in cells (3). Defects in protein trafficking can be used as a model for the induction of PF in vulnerable epithelial cells (4). *HPS1* and *HPS4* genes are ~30–35 kb in length. We developed a long-range PCR strategy (6–7 kb products) to generate end-to-end gene sequencing for *HPS1* and *HPS4* (including ~500 bp of the promoter and flanking the 3′ untranslated region [UTR]), which allowed us to examine other imbedded alterations (splice sites, microRNA sites, enhancer/regulator sites, etc.). Genomic DNA from 87 proband individuals from a collection of patients with familial PF (FPF; obtained from Dr. James E. Loyd, Vanderbilt University) was sequenced. The R package haplo.stats (5) was used to assemble the predicted protein haplotypes for each individual with FPF.

We identified five patients with FPF and rare *HPS1* or *HPS4* missense mutations (three in *HPS1* and one in *HPS4*), and a novel *HPS4* mutation containing a frameshift mutation (Table 1). In addition, 10 patients with FPF had the *HPS1*:G283W mutation (Table 1), which has a 3% haplotype frequency from the 1000 Genomes Project (1KGP) database. This suggests a slight enrichment for *HPS1*:G238W in our patients with FPF (Fisher's exact test *P* value = 0.03). Table 1 shows different functional predictions of the consequence of the missense mutations, identifying *HPS1*-2 and *HPS1*-3 as probably damaging (Exome Variant Server, <http://evs.gs.washington.edu/evs/>). *HPS1*-1 and the novel *HPS4*-1 frameshift were not found in the Exome Variant Server database. In addition, the data for *HPS1*:V4A and *HPS1*:G283W are shown. Using amino-terminal tagged reference

sequence cDNAs (*HA*₃-*HPS1* wild-type [*HA*₃-*HPS1*wt] and *MYC*₃-*HPS4*wt [6]), we introduced each mutation by site-directed mutagenesis. Cycloheximide treatment of each cell line was completed after it was transfected either singly or cotransfected with its wild-type partner (e.g., *HPS1*-1/*HPS4*wt). We used Western blot quantitation to measure tagged protein levels over time, approximating protein stability compared with the wt. When they were singly transfected, we found that the *HPS1*-1 and *HPS1*-2 proteins were less stable than *HPS1*wt, although *HPS1*-3 appeared to be more stable. *HPS4*-1fs was less stable than *HPS4*wt, and the protein stability of *HPS4*-2 was similar to that of *HPS4*wt. When the mutant forms were cotransfected with their wt counterpart, a similar pattern emerged (preliminary data collected in several cell lines).

We built each patient's predicted *HPS1* and *HPS4* protein haplotypes using a subset of SNPs that we annotated as being within the promoter, mRNA (including the 5′UTR and 3′UTR), and within 50 bp of the exon/intron splice site. In addition, we wanted the average heterozygosity to be between 10% and 30% (from the dbSNP database; <https://www.ncbi.nlm.nih.gov/snp/>) for maximum utility in building haplotypes. In most cases, haplo.stats built one unique pair of protein haplotypes with high *post hoc* probability (0.9–1.0). In cases where more than two pairs were predicted, the changes were either in noncoding regions or were silent mutations.

Table 2 summarizes the predicted protein haplotype for each missense mutation found in our FPF cohort, along with the most common and/or reference genome sequence protein haplotype. The predicted protein haplotypes for *HPS1*-1, 1-2, and 1-3 mutations indicate that *HPS1*-1 and *HPS1*-3 were not tested in the appropriate protein haplotype (Table 2). Even though *HPS1*-2 was tested in the correct protein haplotype, that particular patient, FPF47, has the *HPS1*:G283W mutation in his/her other allele, suggesting a potential partial loss of function from both *HPS1* alleles. *HPS4* protein haplotypes presented a different problem. The reference genome sequence (tested in a wt cDNA clone) is actually a minor haplotype (13.5%) within the 1KGP, and the *HPS4* mutations identified in our FPF cohort are in fact on the major *HPS4* protein haplotype (43.7%; Table 2). This demonstrates that the correct protein context for the *HPS1* and *HPS4* mutations found in our cohort have not been adequately tested in their protein haplotype context.

With the increased use of next-generation-sequencing platforms, the identification of rare and novel missense mutations occurs frequently in large disease-specific studies. HPS is a recessive disorder with a high penetrance for PF in the affected HPS subtypes. However, in the context of PF disease, heterozygous carriers, as identified in this study, may confer an added risk in a two-hit hypothesis, akin to *BMPR2* (bone morphogenetic protein receptor 2) mutations in pulmonary arterial hypertension. The initial approach is to test these missense mutations by computational methods to suggest the potential degree of maladaptation of a given change (for a review of current methods, see Reference 7). The various approaches used to obtain the data shown in Table 1 tested the mutation either in isolation (how biochemically similar is glycine to tryptophan?) or in its relative evolutionary conservation (local sequence preservation over evolutionary time). None of these methods take into account that a particular missense mutation is in fact part of a specific protein haplotype.

Author Contributions: R.S.S., L.R.Y., and M.W.G. conceived the project and designed the experiments. R.S.S., A.R.C., L.R.Y., D.S.C., E.A.M., X.L., N.H., L.M.F., and T.L.P. conducted experiments or helped with analysis. R.S.S., L.R.Y., and M.W.G. wrote the manuscript. All authors read and approved the manuscript.

Originally Published in Press as DOI: 10.1164/rccm.201902-0457LE on April 15, 2019

Table 1. Identification and Characteristics of the Newly Identified HPS1 and HPS4 Missense Mutations

Gene	Mutation	rsID	Function	Protein Variant	Conservation Score (GERP)	Grantham Score	Polyphen2 (Class: Score)	1KPG Frequency	ExAC Project Frequency	ClinVar Predicted Effect
HPS1	V4A	rs58548334	Missense	p.(V4A)	4.62	64	Probably damaging: 0.966	2.47%	3.32%	Likely benign
HPS1	G238W	rs11592273	Missense	p.(G283W)	—	184	Probably damaging: 0.99	3.02%	5.13%	Likely benign
HPS1	HPS1-1	rs201728087	Missense	p.(R429H)	—	—	—	—	0.03%	Uncertain
HPS1	HPS1-2	rs139061260	Missense	p.(V630I)	3.73	29	Probably damaging: 0.999	<0.1%	0.13%	Benign
HPS1	HPS1-3	rs116698870	Missense	p.(G639S)	4.64	56	Probably damaging: 1.0	<0.1%	0.05%	Uncertain
HPS4	HPS4-1	—	Frameshift	p.(S655InsCA)	—	—	—	novel	novel	—
HPS4	HPS4-2	rs147435410	Missense	p.(R466C)	—1.63	180	Benign: 0.002	<0.1%	0.13%	Not reported

Definition of abbreviations: 1KPG = 1000 Genomes Project; ExAC = Exome Aggregation Consortium; GERP = genomic evolutionary rate profiling; HSP = Hermansky-Pudlak syndrome; rsID = reference SNP cluster ID.

The characteristics of the five rare or novel HPS1 and HPS4 mutations are shown, along with those of two HPS1 missense mutations found in the human population that are important for this work. Columns 3–8 are from the Exome Variant Server database (<http://evs.gs.washington.edu/evs/>). Also shown are three types of amino acid substitution effects: evolutionary rate (GERP), chemical dissimilarity (Grantham score), and potential function/structure comparison (PolyPhen2). Column 9 is from the 1KGP (<http://phase3browser.1000genomes.org/index.html>) predicted protein haplotype frequencies, column 10 is from the ExAC project database (<http://exac.broadinstitute.org/>), and column 11 is a functional predicted effect of the missense mutation from the ClinVar database (<https://www.ncbi.nlm.nih.gov/clinvar/>). HPS1-1 and HPS4-1 are not reported in the Exome Variant Server database.

Table 2. Predicted Protein Haplotypes of the Identified HPS1 and HPS4 Missense Mutations

Gene	1KGP Protein Haplotype	Haplotype Frequency	Source	Human Genome	V4	G283	R429	P491	Q603	V630	G639
HPS1*	1	46.30%	1KGP	Reference	V4	G283	R429	P491	Q603	V630	G639
HPS1	3	6.18%	1KGP	—	V4	G283	R429	R491	R603	V630	G639
HPS1	5	3.02%	1KGP	—	V4	W283	R429	P491	Q603	V630	G639
HPS1	6	2.47%	1KGP	—	A4	G283	R429	P491	Q603	V630	G639
HPS1	HPS1-1	—	p.(R429H)	—	A4	G283	R429H	R491	R603	V630	G639
HPS1	HPS1-2	—	p.(V630I)	—	V4	G283	R429	P491	Q603	V630I	G639
HPS1	HPS1-3	—	p.(G639S)	—	V4	G283	R429	R491	R603	V630	G639S
HPS4	1	43.70%	1KGP	—	G229	L443	R466	M552	Y606	H625	S655
HPS4	2	29.90%	1KGP	—	G229	V443	R466	M552	Y606	H625	S655
HPS4*	3	13.50%	1KGP	Reference	E229	L443	R466	V552	H606	Q625	S655
HPS4	HPS4-1	—	p.(S655InsCA)	—	G229	L443	R466	M552	Y606	H625	S655fs+CA
HPS4	HPS4-2	—	p.(R466C)	—	G229	L443	R466C	M552	Y606	H625	S655

Definition of abbreviations: 1KGP = 1000 Genomes Project; HPS = Hermansky-Pudlak syndrome.

The upper part of this table focuses on HPS1-relevant protein haplotypes (predicted by 1KGP) for comparison with the rare HPS1 mutations identified in this study. The lower part focuses on the HPS4 protein haplotypes. Codons that differ from the most common protein haplotype identified by 1KGP are in bold type. The rare/novel HPS1 and HPS4 mutations are in bold italic type. As described in the text, the HPS1 mutations are in a complex background of their predicted protein haplotype. The HPS4 mutations are on the most common protein haplotype rather than on the rarer reference genome protein haplotype.

*Protein haplotypes for HPS1 and HPS4 representing the human reference genomic sequence.

As in the HPS1 and HPS4 system, proteins do not generally function as monomers but are in fact part of large functional complexes. Additional consideration of the role a given protein has in a functional complex may change the interpretation of a “benign” missense to “detrimental” because of intermolecular interference (8), which currently cannot be tested computationally. However, in our system, we have the potential to test this experimentally. In future studies, we plan to incorporate the discovered rare missense *HPS1* or *HPS4* mutations into their predicted protein haplotype and test their protein stability, ability to form a molecular complex, and potential functional interference in protein-trafficking pathways. ■

Naoko Hara, M.S.
Lynsey M. Fettig, Ph.D.
Tzu L. Phang, Ph.D.
University of Colorado School of Medicine
Aurora, Colorado

Mark W. Geraci, M.D.
Indiana University School of Medicine
Indianapolis, Indiana

ORCID IDs: 0000-0002-7628-7066 (R.S.S.); 0000-0002-2280-3424 (M.W.G.).

*Corresponding author (e-mail: rosteam@iu.edu).

Author disclosures are available with the text of this letter at www.atsjournals.org.

Robert S. Stearman, Ph.D.*
Indiana University School of Medicine
Indianapolis, Indiana

Amber R. Cornelius, M.S.
New York University School of Medicine
New York, New York

Lisa R. Young, M.D.
David S. Conklin, M.D.
Vanderbilt University Medical Center
Nashville, Tennessee

Elizabeth A. Mickler, M.S.
Indiana University School of Medicine
Indianapolis, Indiana

Xiao Lu, D.O.
St. John Medical Center
Westlake, Ohio

References

- Vicary GW, Vergne Y, Santiago-Cornier A, Young LR, Roman J. Pulmonary fibrosis in Hermansky-Pudlak syndrome. *Ann Am Thorac Soc* 2016;13:1839–1846.
- Oh J, Bailin T, Fukai K, Feng GH, Ho L, Mao JI, et al. Positional cloning of a gene for Hermansky-Pudlak syndrome, a disorder of cytoplasmic organelles. *Nat Genet* 1996;14:300–306.
- Gerondopoulos A, Langemeyer L, Liang JR, Linford A, Barr FA. BLOC-3 mutated in Hermansky-Pudlak syndrome is a Rab32/38 guanine nucleotide exchange factor. *Curr Biol* 2012;22:2135–2139.
- Kropski JA, Lawson WE, Young LR, Blackwell TS. Genetic studies provide clues on the pathogenesis of idiopathic pulmonary fibrosis. *Dis Model Mech* 2013;6:9–17.
- Sinnwell JP, Schaid DJ. haplo.stats: statistical analysis of haplotypes with traits and covariates when linkage phase is ambiguous. 2018 [accessed 2019 Feb 23]. Available from: <http://www.mayo.edu/research/labs/statistical-genetics-genetic-epidemiology/software>.
- Martina JA, Moriyama K, Bonifacio JS. BLOC-3, a protein complex containing the Hermansky-Pudlak syndrome gene products HPS1 and HPS4. *J Biol Chem* 2003;278:29376–29384.
- Wang T, Bu CH, Hildebrand S, Jia G, Siggs OM, Lyon S, et al. Probability of phenotypically detectable protein damage by

ENU-induced mutations in the Mutagenetix database. *Nat Commun* 2018;9:441.

8. Carmona-Rivera C, Simeonov DR, Cardillo ND, Gahl WA, Cadilla CL. A divalent interaction between HPS1 and HPS4 is required for the formation of the biogenesis of lysosome-related organelle complex-3 (BLOC-3). *Biochim Biophys Acta* 2013;1833:468–478.

Copyright © 2019 by the American Thoracic Society

The Mortality Burden of Idiopathic Pulmonary Fibrosis in the United Kingdom

To the Editor:

We have previously demonstrated that mortality from idiopathic pulmonary fibrosis (IPF) is on the rise in the United Kingdom (1) and globally (2). Although numerous clinical trials and genetic and metabolomics studies have provided insight, large epidemiological studies on disease burden remain limited. This is particularly important now that therapeutic options for IPF are available. Despite limitations, registered deaths from IPF are a pragmatic and validated (3) way of assessing and providing reliable estimates of disease burden in terms of mortality. The aim of our study was to provide up-to-date IPF mortality rate estimates and changes to trends in the United Kingdom.

Methods

Data source. We used routine mortality data from England and Wales obtained from the Office of National Statistics (4), which are derived from registered death certificates and coded for the underlying cause of death (5). The accuracy of coding of underlying cause of deaths is high (6).

We used the term IPF clinical syndrome (IPF-CS) (1, 7) when using routinely collected clinical data to acknowledge the difficulty in demonstrating that our study population meets American Thoracic Society/European Respiratory Society diagnostic criteria (8).

Statistical analysis. We obtained the annual number of deaths from IPF-CS stratified by age and sex for the years 1979 to 2016. During our study period, deaths were coded using different International Classification of Diseases (ICD) codes; from 1979 to 2000 ICD Ninth Revision (ICD-9) codes 515 (post inflammatory fibrosis) and 516.3 (idiopathic fibrosing alveolitis) were used, and from 2001 onward we used ICD 10th Revision (ICD-10) code J84.1 (idiopathic pulmonary fibrosis). We extracted general population estimates for the same period. All data were grouped into 5- to 6-year periods, but without combining the different ICD codes, and 5-year age groups >55 years of age.

Author Contributions: V.N. and R.B.H. conceived and designed the study. V.N. prepared and analyzed the data. V.N. and R.B.H. were involved in the interpretation of the data and in writing or revising the manuscript before submission. V.N. takes responsibility for the integrity of the work in this manuscript and is the guarantor of the manuscript.

Originally Published in Press as DOI: 10.1164/rccm.201902-0467LE on April 11, 2019

Crude mortality rates stratified by age, sex, and calendar period were estimated. We used direct standardization to calculate age-standardized mortality rates, standardized to the 2016 population of England and Wales. We performed a subanalysis of the two different ICD-9 codes for IPF-CS (ICD-9 code 515 and ICD-9 code 516.3). Poisson regression was used to estimate mortality rate ratios, adjusting for age and sex. We explored effect modification between trends in annual mortality rates with age and sex. Likelihood ratio tests were used for hypothesis testing. Stata v15 (StataCorp) was used for all statistical analyses.

Results

A total of 82,702 deaths were attributed to IPF-CS in England and Wales from 1979 to 2016. The overall age-standardized mortality rate for this period was 4.68 per 100,000 person-years (95% confidence interval [CI], 4.64–4.71 per 100,000 person-years). Age-standardized mortality rates increased from 1.66 (95% CI, 1.54–1.79) in 1979 to 8.29 (95% CI, 8.06–8.53) per 100,000 person-years in 2016 (Figure 1). The annual increase in the number of IPF-CS deaths, after adjusting for age and sex, was approximately 5% (rate ratio [RR], 1.050; 95% CI, 1.049–1.051; $P < 0.001$). Age-standardized mortality rates were higher in men and the elderly (Table 1). There was also strong evidence of effect modification by age ($P < 0.001$) and sex ($P < 0.001$). The increase in annual mortality rates was highest among men (RR per year, 1.052; 95% CI, 1.051–1.053) and patients older than 85 years of age (RR per year, 1.094; 95% CI 1.092–1.096).

Subanalysis of ICD-9 codes. From 1979 to 2000, there were 12,989 deaths from postinflammatory fibrosis (ICD-9 code 515) and 16,989 deaths from idiopathic fibrosing alveolitis (ICD-9 code 516.3). The age-standardized mortality rates for both ICD-9 codes increased during this period. For postinflammatory fibrosis, the age-standardized mortality rate increased from 0.92 (95% CI, 0.88–0.97) per 100,000 person-years in the 1979 to 1983 calendar period to 1.77 (95% CI, 1.72–1.82) per 100,000 person-years in the 1995 to 2000 calendar period. Age-standardized mortality rates for idiopathic fibrosing alveolitis increased from 0.88 (95% CI, 0.84–0.92) to 2.41 (95% CI, 2.35–2.47) per 100,000 person-years for the same calendar periods. Yearly mortality increased by 4.4% (RR, 1.044; 95% CI, 1.041–1.047) for postinflammatory fibrosis and 6.3% (RR, 1.063; 95% CI, 1.060–1.066) for idiopathic fibrosing alveolitis, after controlling for age and sex.

Discussion

This study demonstrates that mortality from IPF-CS continues to increase by 5% per year, after accounting for an ageing population. We found a marked increase in IPF-CS deaths from 2010 onward, with IPF-CS mortality in 2016 being 51% higher than 2010. Subanalysis of the ICD-9 codes showed that age-standardized mortality rates increased for both codes. Registered deaths were highest in men and the elderly, which is consistent with previous studies (1, 2). On the basis of these findings, we estimate that approximately 5,500 people die from IPF-CS in the United Kingdom each year.

One strength of our study is the large number of registered deaths and long study period, which enabled us to estimate precise changes to mortality trends and stratify our results by sex, age, and calendar period. We were also able to explore if the