Biomedical Informatics

Association of canonical pathways with length of survival in PAH

Amy Webb PhD; Charles Fauvel MD; Priscilla Correa-Jaque MTDA, CCRC, ACRP-PM; Shili Lin PhD; Vladimir Valtchinov PhD; Zilu Liu; Simon Kennedy; Franz P Rischard DO; Aaron Waxman MD, PhD; Roham T Zamanian MD, FCCP ; Raymond Benza MD, FACC, FAHA, FACP

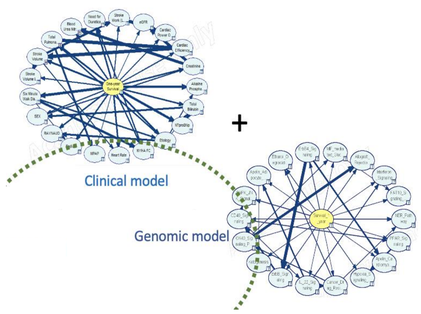
INTRODUCTION

Pulmonary arterial hypertension (PAH) is a chronic, progressive disease without cure. Treatment can improve outcome, but informed predictions with clinical and genomic measurements can guide treatment and therapy choices. Previous research has focused on identifying clinical variables that could predict future outcomes. The current research aims to find genomic variants that can predict survival time.

Risk prediction

Existing approaches for assessing risk in PAH patients include the use of equations and scores, developed from contemporary PAH registries.[1-4] However, these risk stratification tools vary in their precision, nature of their derivation, and utility for periodic use. They assume that the clinical variables that contribute to PAH risk are established, linear in robustness, and limited to established variables (i.e hemodynamics).
The existing tools for PAH survival stratification are so-called 'risk scores', which attempt to distill key clinical variables into a single numerical index (e.g., REVEAL, Pulmonary Hypertension Connection Score, French National registry score, Scottish registry score, and NIH risk survival score7b-10b). Included in these scoring systems are several variables that imply non-modifiable risk such as gender and subtype of PAH, yet these factors fail to fully account an individual's unique definable risk that can only be illustrated by their genetic fingerprints. PHORA 2.0 is a Bayesian network model built from select clinical measurements to predict survival for Pulmonary arterial hypertension. We are building on the PHORA 2.0 model by adding the ability to incorporate genomic data.

Risk prediction based on Clinical + Genomic variables



AIM

Whole genome sequencing was generated on a PAH cohort providing a broad opportunity to measure genomic variants in coding regions as well as less understood and often overlooked non-coding regions that may play regulatory roles. Additionally, WGS allows for the consideration of rare variants, structural variants, and copy number variants. Our initial look at the WGS data focuses on collapsing rare, function-changing variants into pathways with the intention to identify biological processes involved in survival of PAH.

METHODS

- 325 Patients from 2 clinical trial cohorts
- Survival time defined as the time from the date of enrollment to the date of death or censoring was the outcome of interest

- Sequencing whole genome with Illumina, alignment and variant calling using DRAGEN
- Explored different ways to define long/short
  - Strict=death before 5 yrs / survive past 7 years
  - Median=same, but split at median
  - Mean=same, but split at mean

- Data filtering
  - Remove variants as follows:
    - IMPACT=LOW exclude synonymous
    - IMPACT=MODIFIER exclude non-coding
    - acmg_classification=Benign
    - sift_pred=T / polyphen2_hdiv_prediction=B
    - gnomad_genome_af > 5% exclude common
  - Select genes with variant in >=3 samples
  - Select Ingenuity Pathway Analysis canonical pathways with >1 gene with variant

- Test each pathway for association with a long or short survival with a 2x2 Fisher exact test classifying each sample for each pathway as having a variant or not having a variant

- Independent dataset curated by TriAxia Health (TAH) used for replication included 263 samples with a physician provided diagnosis of PAH and time to death survival data
- Limited TAH variants to those already found in initial dataset

- Included associated pathways and select clinical measures in an Adaptive Elastic Net (AEN) to pick a panel most predictive of survival
- AEN models refined with additional machine learning methods

RESULTS

Patient demographics

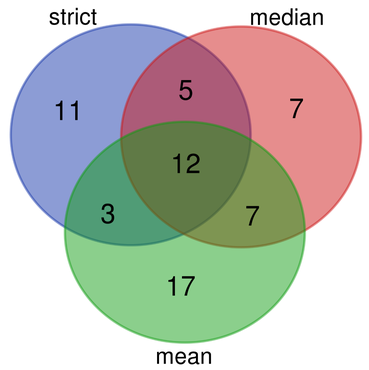
Samples were gathered from 2 studies: 60% from Encysive and 40% from Prospective. The long survival group has an average time to death of 2629 days and the short survival group has an average time to death of 1280 days. Average age was 50.9 years. 96% of the cohort were European, self-reported and confirmed by PCA. 49.3% were primary PAH. 79.3% were female. 23% CHD, 23% CTD, and 4% drug induced.

Associated pathways

Significance set at $p<0.05$ and no pathway met multiple testing correction.

Of pathways containing more than one gene mutated in 3 or more samples, 31 pathways were associated with strict division of survival length using a fisher exact test. 31 pathways associated using a median split of survival and 39 associated using a mean split. 12 of these pathways were associated regardless of split.

Venn diagram showing associated pathways using different definitions of long and short



Four pathways found in independent TriAxia Health dataset showing an association from the same definitions and the same direction

p value	In house	Triaxia	
B Cell Development	0.0142	0.0028	mean split
Cdc42 Signaling	0.0267	0.0198	strict split
Graft-versus-Host Disease Signaling	0.0273	0.0316	mean split
T Helper Cell Differentiation	0.0496	0.0089	mean split
	0.0185	0.0319	mean split

RESULTS, continued

Panel building and machine learning

AEN modeling using any identified associated pathway plus some clinical variables identified one panel using 37 pathways for a mean split with a testing AUC of 0.75 and a second panel of 31 pathways + Sex + Age for a strict split with a testing AUC of 0.78.
These two panels were used as input for further refining with a number of machine learning methods

Model_name	AEN Strict panel Test AUC	AEN Mean panel Test AUC
glmnet	0.79	0.75
neural_network	0.79	0.72
random_forest	0.75	0.74
Gradient_Boosting_machine	0.69	0.77
KNN	0.57	0.66
Support_vector_Machine	0.81	0.75
Lasso	0.82	0.79
Ridge	0.77	0.74

The 4 pathways replicated in the TriAxia Health dataset together with Sex + Age were fed into the AEN modeling using the TAH data giving a validation AUC of 0.66

CONCLUSIONS

Pulmonary arterial hypertension (PAH) is a proliferative vasculopathy, characterized by vasoconstriction, cell proliferation, fibrosis, and thrombosis. Underlying genetic predisposition to pulmonary vascular disease in patients with pulmonary hypertension is an emerging field, our findings now suggest that specific biological pathways may affect survival. Obviously, this novel role of these pathways will need further confirmation, replication and validation in future translational and in vitro studies. These future studies should importantly explore and seek epidemiologic evidence of environmental interactions in pathogenesis of PAH, which may open a window to further understand the unique genetic networks may shed new light on the complex interplay between the environment, female gender and this disease

BIBLIOGRAPHY

1. Galie N, Humbert M, Vachiery JL, Gibbs S, Lang I, Torbicki A, et al. 2015 ESC/ERS Guidelines for the diagnosis and treatment of pulmonary hypertension: The Joint Task Force for the Diagnosis and Treatment of Pulmonary Hypertension of the European Society of Cardiology (ESC) and the European Respiratory Society (ERS): Endorsed by: Association for European Paediatric and Congenital Cardiology (AEPC), International Society for Heart and Lung Transplantation (ISHLT). Eur Respir J. 2015;46(4):903-75.
2. Hoeper MM, Kramer T, Pan Z, Eichstaedt CA, Spiesshoefer J, Benjamin N, et al. Mortality in pulmonary arterial hypertension: prediction by the 2015 European pulmonary hypertension guidelines risk stratification model. Eur Respir J. 2017;50(2).
3. D'Alonzo GE, Barst RJ, Ayres SM, Bergofsky EH, Brundage BH, Detre KM, et al. Survival in patients with primary pulmonary hypertension. Results from a national prospective registry. Ann Intern Med. 1991;115(5):343-9.
4. Boucly A, Weatherald J, Savale L, Jais X, Cottin V, Prevot G, et al. Risk assessment, prognosis and guideline implementation in pulmonary arterial hypertension. Eur Respir J. 2017;50(2).

ACKNOWLEDGEMENTS

TriAxia Health, Inc. data developed in collaboration with Mass General Brigham, Stanford Medicine, Weill Cornell Medicine, UPMC, and the University of Arizona

PHORA is the result of a global endeavor with participation from Actelion, Bayer, FDA, United Therapeutics, NHLBI and PHA.
FUNDED BY NIH/NHLBI HL134673

THE OHIO STATE UNIVERSITY
COLLEGE OF MEDICINE

TriAxia HEALTH

PHORA