Integrating molecular genetics and systems approaches to pulmonary vascular diseases

Mark W. Geraci
University of Colorado School of Medicine, Denver, Colorado, USA

ABSTRACT

There is incredible potential to advance our understanding of disease pathogenesis, enhance our diagnostic capability, and revolutionize our treatment modalities with the advent of advanced systems approaches to genetic, genomic, and epigenetic discoveries. Investigation using these technologies is beginning to impact our understanding of pulmonary arterial hypertension (PAH). The following review details work to date on single gene mutations in PAH, and expression array analysis in the disease. The wider use of DNA-based arrays for genome wide association studies (GWAS) and copy number alterations is examined. The impact of epigenomic modulation in the pathobiology of PAH and its therapeutic implications is investigated. Finally, a summary of the capabilities and promises for next-generation sequencing is discussed. A framework for studies of the future is proposed.

Key Words: genomics, microarray, genome-wide association studies (GWAS), epigenetics, micro-RNA (miRNA)

In 2000, researchers from both Columbia University and Vanderbilt University reported that a mutation in the transforming growth factor β receptor (TGF-β) superfamily member was strongly associated with familial pulmonary arterial hypertension (PAH).

Since then it has been clearly demonstrated that heterozygous germline mutations of bone morphogenetic protein receptor 2 (BMPR2), a component of the TGF-β-related family, underlies up to 60% of cases of familial IPAH. More recently, the investigation of other signaling molecules within the BMPR2 pathway have led investigators to other mutations associated with PAH. Patients with hereditary hemorrhagic telangiectasia (HHT, previously known as Osler–Weber–Rendu) can exhibit pulmonary hypertension (PH). In these cases, coding changes in activin-receptor-like kinase 1 (ALK1) are associated with the disease. Further functional analysis of ALK1 in these kindreds demonstrated these mutations to be the cause of PH in patients with HHT.

Endoglin, an accessory TGF-β receptor highly expressed during angiogenesis, is essential for ALK1 signaling. Mutations in endoglin have also been reported in patients with HHT-associated PAH as well in patients with anorexigen-associated PAH. Gene rearrangements of BMPR have also been identified and account for a substantial proportion of the mutations in familial and idiopathic PAH.

A recent summary of the genetic and genomic alterations demonstrate that disruptions of the BMPR2 pathway may be present in a substantial number of families. Clearly, the study of pulmonary vascular disease has been advanced by recent genomic discoveries, but, in reality, the genetics of PH discovery is in its infancy. There is incredible potential to advance our understanding of disease pathogenesis, enhance our diagnostic capability, and revolutionize our treatment modalities with the advent of advanced systems approaches to genetic, genomic, and epigenetic discoveries.

miRNA EXPRESSION STUDIES

Early expression studies of PH lung tissue, though informative, were limited by small sample size. Alternative strategies to use surrogate tissue (peripheral blood) demonstrate the utility of transcriptional profiling in PAH. The utility of expanding cohort size and using highly defined phenotypes for array-based classification was demonstrated using a surrogate tissue (blood) and examining markers that differentiate scleroderma (SSc) without PAH from scleroderma-associated PAH.
patients. There is a clear benefit to utilize larger and well-characterized cohorts when examining lung tissue gene expression profiles. Several newer efforts have focused on this approach. A larger sampling of lung tissue array analysis demonstrates similar pathway disruption between PH and pulmonary fibrosis. Perhaps the largest study until date using lung tissue microarray profiling demonstrates that, in patients with pulmonary fibrosis, the presence of associated PAH is characterized by a specific gene expression signal in both a training and testing algorithm.

Cell-based expression studies have been useful in characterizing selected pathways as well as differences in selected cell populations. For scleroderma-associated PAH, pulmonary fibroblasts and lung tissue from patients with PAH and those SSc patients without PAH demonstrate characteristic gene expression signatures. Several studies have used global gene expression signatures to determine a more robust pathway analysis, including the effects of BMPR2 deficiency. The novel role of IL-13 in PAH pathobiology has been investigated based on array-generated data. Potential new therapeutic targets, such as Apelin and PPARγ, have been extensively studied using array-based platforms.

One significant challenge to all genomic approaches is the leveraging of large data sets into novel systems-based analysis approaches. Putting all the relevant information into a systems model of pulmonary vascular disease may provide unique insights.

**SINGLE NUCLEOTIDE POLYMORPHISMS ARRAYS FOR GENOME-WIDE ASSOCIATION STUDIES AND COPY NUMBER VARIATIONS ANALYSIS**

Microarrays can be used to interrogate SNPs across the genome at unprecedented densities, enabling the completion of GWAS. Until date, more than 1,500 GWAS studies have been published defining a statistical locus for 237 traits. Recently, Dr. William Nichols (1R24HL105333, NATIONAL BIOLOGICAL SAMPLE AND DATA REPOSITORY FOR PAH) received funding to meet the following objectives: (1) work with PAH academic centers in the United States to collect and maintain biological material from no less than 2,500 WHO Group 1 PAH patients. Samples to be banked on each patient include total genomic lymphocyte DNA, EBV transformed peripheral blood lymphocytes, and plasma; (2) generate both genome wide SNP genotype data and BMPR2/ALK1 sequence/MLPA data on 2,500 patients enrolled in the bank. The SNP data will be “cleaned” prior to its deposition in the database of Genotypes and Phenotypes (dbGaP); and (3) widely promote and distribute samples and genotype/sequence data maintained at the National Biological Sample and Data Repository for PAH. All biological samples, clinical data, as well as the SNP genotype and sequencing data for the patients will be made available to the scientific community. These studies should provide extremely valuable information regarding genotype/phenotype interactions and the discovery of modifier gene loci.

Utilization of SNP array data to assess CNVs can provide important information in PAH. Analysis of hyperproliferative endothelial cells and smooth muscles cells from patients with PAH uncovered large-scale genomic alterations in the endothelial cells, which were confirmed in patient lung tissue by fluorescent in situ hybridization (FISH). Using a similar strategy of array-based comparative genomic hybridization (aCGH), investigators were able to demonstrate a specific microdeletion of 17q22q23.2, encompassing T-box transcription factor-2 and 4 (TBX2 and TBX4) in a patient with a syndromic disorder, including PH. A more extensive analysis of SNP data in larger, well-defined cohorts, including analysis for novel copy number variations, may provide specific genomic information regarding the pathobiology of PH.

**ROLE OF microRNAs IN PULMONARY VASCULAR DISEASES**

miRNAs are small noncoding sequences of RNA that have the capacity to regulate many genes, pathways, and complex biological networks within cells. In diseases such as cancer and cardiac disease, their pathogenic role has been documented in detail. There is also, however, great therapeutic potential for miRNAs. Their role in the pathogenesis of PAH has recently come under scrutiny for probably the most extensive global investigation, leading to mechanistic studies and potential therapeutic implications for miRNAs in PAH centers on mir-204. Bonnett et al. provided a comprehensive model linking abnormal miRNA expression to recognized pathophysiological processes in PAH, including nucleated factor of activated T-cells (NFAT) activation, BMPR2 downregulation, IL-6 production, the Rho pathway, pulmonary artery smooth muscle cell (PASMC) proliferation, and apoptosis resistance. In effect, this study not only demonstrates the importance of miRNAs in PAH, but also suggests that re-establishing the miR-204 level might represent a novel therapeutic approach for human PAH.

There are several methods to assess global miRNA expression. Both array-based and polymerase chain reaction (PCR)-based methods represent biased approaches, relying on “known” miRNA sequences. Since miRNA processing can
result in changes of miRNA sequences, the most unbiased approach, and one that is increasingly adopted, is the utilization of massively parallel sequencing strategies targeting small RNA species.

**EPigenetic modifications and Pulmonary vascular diseases**

An epigenetic trait is defined as a stably heritable phenotype resulting from changes in a chromosome without alterations in DNA sequence. Epigenetic changes are thought to be at the root of cellular reprogramming, the process by which a differentiated cell type can be induced to adopt an alternate cell fate. This concept is consistent with observations in PH, where endothelial cells, SMCs, and adventitial fibroblasts all demonstrated significantly altered characteristics including stable increases in proliferation, resistance to apoptosis, metabolic switching, and proinflammatory gene expression. Recent studies have documented changes in the methylation status of important genes, such as superoxide dismutase in cells from the pulmonary hypertensive vasculature. A key finding was lung specific upregulation of DNA methyltransferase (DNAMT) 3b, which allows de novo methylation of CpG islands and inhibition of transcription of a nonmutated gene.

Histone deacetylases (HDACs) catalyze removal of acetyl groups from lysine residues in a variety of proteins. HDACs have mainly been studied in the context of chromatin, where they regulate gene transcription by deacetylating nucleosomal histones. The 18 mammalian HDACs are grouped into four classes. Dysregulation of HDACs is associated with a variety of pathophysiological processes, including cancer and inflammatory signaling in rheumatoid arthritis. Suberoylanilide hydroxamic acid (SAHA) is a pan-HDAC inhibitor that contains a zinc-binding hydroxamic acid warhead. Interestingly, SAHA and other pan-HDAC inhibitors have been shown to be efficacious in rodent models of left ventricular (LV) dysfunction, reducing pathological hypertrophy and fibrosis and improving contractile function, suggesting a novel application for HDAC inhibitors for the treatment of human heart failure. Medicinal chemistry efforts have led to the discovery of compounds that selectively inhibit specific HDAC isoforms. It is believed that these compounds will be safer and more efficacious than pan-HDAC inhibitors in the setting of heart failure.

**HDACs and the right ventricle**

Two reports have addressed effects of HDAC inhibitors in models of right ventricular (RV) remodeling. Valproic acid was shown to block RV cardiac hypertrophy in response to pulmonary artery banding (PAB) as well as in the setting of PH caused by monocrotaline-induced lung injury. In contrast, trichostatin A (TSA) failed to block hypertrophy in response to PAB and actually appeared to worsen RV function. Valproic acid is a weak HDAC inhibitor with many additional pharmacological activities and, thus, it is unclear whether its effects on the RV were a direct consequence of HDAC inhibition. TSA is a potent pan-HDAC inhibitor. The deleterious effects of this compound on the RV could be a reflection of a protective role for an HDAC(s) in RV function. Additional investigation is needed to elucidate the roles of specific HDACs in RV remodeling.

**HDACs and the pulmonary vasculature**

Expression of Class I HDACs, particularly HDAC1, is dramatically elevated in pulmonary arteries of humans with PH and in lungs and vessels from pulmonary hypertensive models. Based on these findings, recent studies have begun to address the role of Class I HDACs in the pathogenesis of PH. In a three-week rat model of hypobaric hypoxia, the Class I HDAC-selective inhibitor, MGCD0103, reduced pulmonary arterial pressure through a mechanism involving suppression of pulmonary artery smooth muscle cell proliferation. The antiproliferative effect of MGCD0103 was due, in part, to upregulation of the FoxO3a transcription factor and induction of a downstream target gene encoding the p27 cyclin-dependent kinase inhibitor. Recent studies have demonstrated that Class I specific HDAC inhibitors can prevent hypoxia reduced remodeling of the pulmonary arteries and preserve right ventricular function.

**Potential impact of next generation sequencing**

Next-generation sequencing provides an incredibly powerful and comprehensive tool to examine genomic data including RNA expression, splice variation, allele-specific expression, miRNA expression, transcription factor binding, DNA sequence variation, copy number variation, and rare genomic variants. While the technology has rapidly advanced to enable the accumulation of massive amounts of data, the protocols for data analysis, especially the development of “best practices” has unfortunately lagged behind. There continues to be debate regarding the best strategies for sequence alignment as well as RNA sequence evaluation.

Until date, few studies directly related to pulmonary vascular disease and next-generation sequencing have been published. Therefore, the utility of these techniques for discovery in PH pathobiology remains an unrealized opportunity. However, there are several examples of rapid translational science in the realm of oncology that provide a guiding framework for future PH studies.
Several recent sentinel studies exemplify the utility of “big science” in oncology. In one example, investigators, with the collaboration of the Novartis Institutes for Biomedical Research, were able to generate the Cancer Cell Line Encyclopedia. In this work, a compilation of gene expression, chromosomal copy number and massively parallel sequencing data from 947 human cancer cell lines was created. When coupled with pharmacological profiles for 24 anticancer drugs across 479 of the cell lines, this collection allowed identification of genetic, lineage, and gene-expression-based predictors of drug sensitivity. This same group performed massively-parallel deep sequencing of 137 “actionable” cancer genes on paraffin-embedded tumor samples. Four-hundred-fold coverage was achieved and mutations, gene amplifications, and gene fusions were identified for each sample. This data can now be used to guide specific therapy. Such examples illustrate the powerful potential of next-generation sequencing performed on clinical (even archival) specimens in advancing the reality of “personalized medicine.”

SUMMARY AND OPPORTUNITIES FOR THE FUTURE

Significant advances have been made in our understanding of the genetics, genomics, and epigenetics of PH until date. However, this area of research remains in its infancy. Several systematic approaches and high priority areas of investigation are likely to lead to further insight into pathogenesis and treatment of PAH.

Six broad research strategic opportunities have been identified:

1. **Cohorts.** Prior to the investigation using any of the high-throughput “omics” technologies, there was a consensus that exquisite phenotyping of cohorts is needed. In effect, the topic of investigating “phenomes” would be important for any future investigations. The definition of selected phenotypes and endophenotypes is necessary for future work. Moreover, a comprehensive and integrated approach to patient enrollment and databasing of large cohorts is viewed as highly valuable. Because the disease has protean manifestations, yet remains relatively rare, a consortium approach to acquisition of cohorts was viewed as the most likely to succeed.

2. **Examination of DNA variation.** With the advent of high-throughput technologies, including array-based technologies and DNA sequencing, the application of the examination of DNA sequence variation related to disease states and defined phenotypes may enable a highly accurate determination of the importance of rare variants.

3. **Gene expression and control of transcription.** A broad approach to analysis of gene expression, including involved tissues, laser capture of defined elements, cell line-based examinations, and surrogate tissues such as blood may be required. Technologies have evolved to examine the control of transcription including epigenomic modifications and the role of miRNA in disease processes. In particular, the potential for defining therapeutic targets through these broad approaches should be a reasonable goal.

4. **Examination of proteomic and metabolomic signatures.** Augmented capacities in proteomic technologies enable broader examination of proteomic profiles, post-translational modification, and metabolomic signatures. The applications of these technologies for discovering disease pathogenesis and biomarker discovery and validation are promising future areas of investigation.

5. **Systems biology approaches.** Integration of the previously mentioned broad-based approaches is essential for better definition of the disease, pathogenesis of the lung vascular disease, and therapeutic intervention. Network analysis can be derived for simple canonical system motifs or more complex, scale-free, systems may be envisioned to examine the potential for disease similarities by common hubs and nodes. miRNAs may in effect work as network super hubs. Through an extensive understanding of the science of computational biology applied to this disorder, diagnostic and therapeutic implications are expected.

6. **Biobanks.** The systematic approach to well-annotated biobanks remains strongly endorsed and at the center of translational research. In addition to tissue from explanted lungs, the capacity for the prospective collection of cells and tissues from patients may present model systems for further examination. Public and private partnership between industry sponsored trials and independent investigation by way of ancillary studies to clinical trials is highly desirable. Moreover, the development of novel model systems, including but not limited to cell-based systems, rodent-based animal model systems, and other genetically modulated animal systems should be optimized to most accurately reflect the human condition and should be utilized for experimentation and potential therapeutic intervention.

REFERENCES

1. Deng Z, Morse JH, Slager SL, Cuervo N, Moore KJ, Venetos G, et al. Familial primary pulmonary hypertension (gene PPH 1) is caused by mutations in the bone morphogenetic protein receptor-II gene. Am J Hum Genet 2000; 67:737-44.

2. International PPH Consortium, Lane KB, Machado RD, Pauolo MW, Thomson JR, Phillips JA 3rd, et al. Heterozygous germline mutations in BMPR2, encoding a TGF-beta receptor, cause familial primary pulmonary hypertension. The International PPH Consortium. Nat Genet 2000; 26:81-4.

3. Deng Z, Morse JH, Slager SL, Cuervo N, Moore KJ, Venetos G, et al. Familial
primary pulmonary hypertension (gene PPH1) is caused by mutations in the bone morphogenetic protein receptor-II gene. Am J Hum Genet 2000;67:237-44.
4. Trembath RC, Thomson JR, Machado RD, Morgan NV, Atkinson C, Winship I, et al. Clinical and molecular genetic features of pulmonary hypertension. Eur Respir J 2013;41:628-37.
5. Harrison R, Flanagan J, Sankelo M, Abdalla S, Rowell J, Machado R, et al. Molecular and functional analysis identifies ALK-1 as the predominant cause of pulmonary hypertension related to hereditary haemorrhagic telangiectasia. J Med Genet 2003;40:865-71.
6. Chaouat A, Coulet F, Favre C, Simononneau G, Weitzenblum E, Soubrier F, et al. Endoglin germline mutation in a patient with hereditary haemorrhagic telangiectasia and dexamethasone-associated primary arterial hypertension. Thorax 2004;59:446-8.
7. Aldred MA, Vijayakrishnan J, James V, Soubrier F, Gomez-Sanchez MA, Martenson G, et al. BMPR2 gene rearrangements account for a significant proportion of mutations in familial and idiopathic pulmonary arterial hypertension. Hum Mutat 2006;27:212-3.
8. Machado RD, Eickelberg O, Elliott CG, Geraci MW, Hanaoka M, Loyd JE, et al. Genetics and genomics of pulmonary arterial hypertension. J Am Coll Cardiol 2009;54:532-42.
9. Geraci MW, Moore M, Gewell T, Yeager ME, Alger L, Golpon H, et al. Gene expression patterns in the lungs of patients with primary pulmonary hypertension: A gene microarray analysis. Circ Res 2001;88:555-62.
10. Bull TM, Coldren CD, Moore M, Soubrier F, Edwards MG, Collier D, et al. Altered D immune phenotype in peripheral blood cells of patients with scleroderma-associated pulmonary hypertension. Clin Transl Sci 2010;3:210-8.
11. Rajkumar R, Konishi K, Richards TJ, Ishizawar DC, Wichert AC, Kaminski N, et al. Genome-wide RNA expression profiling in lung identifies distinct signatures in idiopathic pulmonary arterial hypertension and secondary pulmonary hypertension. Am J Physiol Heart Circ Physiol 2010;298:H1235-48.
12. Mura M, Amraku M, Yun Z, McAke R, Liu M, Waddell TK, et al. Gene expression profiling in the lungs of patients with pulmonary hypertension associated with pulmonary fibrosis. Chest 2012;141:661-73.
13. Hsu E, Shi H, Jordan RM, Lyons-Weiler J, Pilewski JM, Feghali-Bostwick CA. Lung tissues in patients with systemic sclerosis have gene expression patterns unique to pulmonary fibrosis and pulmonary hypertension. Arthritis Rheum 2011;63:783-94.
14. Davies RJ, Holmes AM, Deighton J, Long L, Yang X, Barker L, et al. BMP type II receptor deficiency confers resistance to growth inhibition by TGFBeta in pulmonary artery smooth muscle cells: Role of proinflammatory cytokines. Am J Physiol Lung Cell Mol Physiol 2012;302:L604-15.
15. Hecker M, Zasłona Z, Kwapiszewska G, Niess G, Zakrzewicz A, Davies RJ, Holmes AM, Deighton J, Long L, Yang X, Barker L, et al. Microdeletion of 17q22q23.2 encompassing TBX2 and TBX4 in a patient with congenital microcephaly, thyroid duct cyst, sensorineural hearing loss, and pulmonary hypertension. Am J Med Genet A 2011;155A:418-23.
16. Hsu E, Shi H, Jordan RM, Lyons-Weiler J, Pilewski JM, Feghali-Bostwick CA. MicroRNA and vascular remodeling in acute vascular injury and pulmonary vascular remodeling. Cardiovasc Res 2012;93:594-604.
17. Small EM, Olson EN. Pervasive roles of microRNAs in cardiovascular biology. Nature 2011;469:336-42.
18. Courbon E, Azain P, Giguere NJ, Sakoun N, Perreau T, Meloche J, et al. Role for miR-204 in human pulmonary arterial hypertension. J Exp Med 2011;208:535-48.
19. Bruneau BG. Epigenetic regulation of the cardiovascular system: introduction to a review series. Circ Res 2010;107:324-6.
20. Archer SL, Weir EK, Willkins MR. Basic science of pulmonary arterial hypertension for clinicians: New concepts and experimental therapies. Circulation 2010;121:2045-66.
21. Gregoretti IV, Lee YM, Goodson HV. Molecular evolution of the histone deacetylase family: functional implications of phylogenetic analysis. J Mol Biol 2004;338:17-31.
22. Bush EW, McKinsey TA. Protein acetylation in the cardiorenal axis: The promise of histone deacetylase inhibitors. Circ Res 2010;106:272-84.
23. McKinsey TA. Isoform-selective HDAC inhibitors: Closing in on translational medicine for the heart. J Mol Cell Cardiol 2011;51:491-6.
24. Cho YK, Eom GH, Kee HJ, Kim HS, Choi WY, Nam KI, et al. Sodium valproate, a histone deacetylase inhibitor, but not captopril, prevents right ventricular hypertrophy in rats. Circ J 2010;74:760-70.
25. Bogaard HJ, Mizzuto S, Hussaini AA, Toldo S, Abbate A, Kraskauskas D, et al. Suppression of histone deacetylases worsens right ventricular dysfunction after pulmonary artery banding in rats. Am J Respir Crit Care Med 2010;181:1402-10.
26. Terbach N, Williams RS. Structure-function studies for the panacea, valproic acid. Biochem Soc Trans 2009;37:1126-32.
27. Cavasin MA, Demos-Davies K, Horn TR, Walker LA, Lemon DD, Birdsey N, et al. Selective class I histone deacetylase inhibition suppresses hypoxia-induced cardiopulmonary remodeling through an antiproliferative mechanism. Circ Res 2011;109:739-48.
28. Treangen TJ, Saltzberg SL. Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nat Rev Genet 2012;13:36-46.
29. Martin JA, Wang Z. Next-generation transcriptome assembly. Nat Rev Genet 2011;12:671-82.
30. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature 2012;483:603-7.
31. Wagle N, Berger MF, Davis MJ, Blumenstiel B, Defelice M, Pochanard P, et al. High-throughput detection of actionable genomic alterations in clinical tumor samples by targeted, massively parallel sequencing. Cancer Discov 2012;2:82-93.
32. Erzurum S, Rounds SI, Stevens T, Aldred M, Aliotta J, Archer SL, et al. Strategic plan for lung vascular research: An NHLBI-ORDR Workshop Report. Am J Respir Crit Care Med 2010;182:1554-62.

Source of Support: Nil, Conflict of Interest: None declared.