

The NIH Undiagnosed Diseases Program: bonding scientists and clinicians

William A. Gahl^{1,2,*}, Cornelius F. Boerkoel² and Manfred Boehm³

Summary

The majority of the biomedical research workforce and funds are focused on studying common diseases and the development of drugs to treat them. However, some of the most remarkable discoveries in physiology and medicine are uncovered by studying rare conditions, because the importance of certain molecular mechanisms is revealed only when their dysfunction results in disease. In 2008, the National Institutes of Health (NIH) launched the NIH Undiagnosed Diseases Program (UDP), which recruits and selects patients who suffer from diseases of unknown etiology, and studies their causes at the clinical, genetic and cellular levels. In this Editorial, we discuss how the UDP has enabled the discovery of several new diseases and disease mechanisms through

collaborations between clinical and basic science teams, using the power of both clinical medicine and biological models. Establishing programs with similar infrastructure at other centers around the world could help to benefit patients, their families and the entire medical community, by enhancing research productivity for rare and novel diseases.

Background

Natural selection solidifies the most favorable alterations in our replicative processes to create brilliantly functional biochemical and cellular pathways. By contrast, less favorable genetic changes or 'errors' disrupt nature's beautifully honed systems. Hence, it is axiomatic that rare diseases, in addition to being coruscating curiosities, also serve as windows to the world's normal order. Among the women and men who address human mutations, clinicians provide care for those affected with rare maladies, and a special cadre of scientists investigates the causes of those disorders. Together, these two groups employ their expertise to advance treatment and understanding.

The NIH Undiagnosed Diseases Program (UDP), announced May 19, 2008, serves as a focal point for the synergistic interplay between clinical and basic science investigators (Gahl and Tiff, 2011; Gahl et al., 2011). The UDP filters thousands of enquiries

down to hundreds of NIH Clinical Center inpatient admissions, and then chooses illustrative cases to pursue experimentally. Three years into its operation, the UDP had reviewed more than 1800 thick medical records, accepted approximately 430 patients and admitted 350 of them. The program had performed close to 600 single nucleotide polymorphism (SNP) arrays and analyzed approximately 230 whole exome sequences (WESs) using DNA from patients and their family members. Many specific cases warranted intense investigation, based on their compelling clinical presentations, the likelihood that the basic defects would be uncovered and the impact predicted to emanate from the 'new disease discovery'. In the sections below, we recount examples that illustrate the process of disease discovery and how it can, in turn,

provide new information about the role of genes and biological pathways in normal and impaired physiology.

Stepwise investigation of an undiagnosed disease

The paradigm for studying such medical mysteries begins with detailed and accurate phenotyping. For example, UDP clinical investigators ascertained a family of five adults from middle America with stunning arterial calcification and claudication confined to the lower extremities (St. Hilaire et al., 2011).

Intense clinical examinations verified that this was not atherosclerosis, that the coronaries were spared, and that the only additional sites of ectopic calcification involved the joint capsules of the hands and feet. Endocrine abnormalities of calcium and phosphate were ruled out.

The next step involved applying genetic methodologies to identify candidate genes causing this disease. In this case, the project benefited enormously from the fact that the parents of the five siblings were third cousins, unaffected, and readily available for study. The most appropriate genetic platform to employ was the million SNP array, which detects copy number variants (deletions and duplications) and also identifies regions of homozygosity. In a product of a consanguineous mating, most homozygous areas derive from a common ancestor, and a single pathogenic mutation in a shared ancestral gene could cause a recessive disease. Hence, we were looking for a region of the human genome in which all five affected siblings were homozygous, but the parents were heterozygous. Only one area satisfied these requirements – a 22-Mb region on chromosome 6. That area contained 92 genes; any one of them could have contained a homozygous mutation that caused the arterial calcifications. We needed an expert in vascular biology to identify the best candidates to pursue.

This was the third, and most crucial, step in discovering a new human disease, i.e. identifying and connecting with a basic scientist

“As our knowledge and data in genetics burgeon, we must engage the cognoscenti in a wide variety of fields to discover new diseases, new mechanisms of action and, ultimately, new therapeutic interventions”

¹Office of the Clinical Director, NHGRI, NIH, Building 10, Room 10C103, 10 Center Drive, MSC-1851, Bethesda, MD 20892-1851, USA

²NIH Undiagnosed Diseases Program, NIH, 9 Memorial Drive, Room 1N113, Bethesda, MD 20892, USA

³National Heart, Lung and Blood Institute, NIH, Building 31, Room 5A52, 31 Center Drive MSC 2486, Bethesda, MD 20892, USA

*Author for correspondence (bgahl@helix.nih.gov)

© 2012. Published by The Company of Biologists Ltd
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial Share Alike License (<http://creativecommons.org/licenses/by-nc-sa/3.0>), which permits unrestricted non-commercial use, distribution and reproduction in any medium provided that the original work is properly cited and all further distributions of the work or adaptation are subject to the same Creative Commons License terms.

whose lab possessed expertise in the field. In this case, Manfred Boehm and Cynthia St. Hilaire of the National Heart, Lung and Blood Institute were authorities on vascular cell metabolism, and cell-based assays were extant in the Boehm lab. In concert with the UDP, whose clinicians provided a skin biopsy, St. Hilaire cultured an affected patient's fibroblasts and performed expression analysis of candidate genes in the inherited homozygous region. Only one gene, *NT5E*, which encodes the CD73 protein, was differentially expressed. Direct sequencing of *NT5E* identified null mutations that resulted in complete inactivation of CD73 protein in all affected siblings. This newly identified vascular disease was referred to as 'arterial calcification due to deficiency of CD73' (ACDC). CD73 is a membrane-bound enzyme that converts extracellular AMP to adenosine and inorganic phosphate; this process was found to be severely impaired in fibroblasts from individuals with ACDC. The failure to generate extracellular adenosine leads to an increase in tissue non-specific alkaline phosphatase (TNAP), a key enzyme in tissue calcification in vitro and in vivo. Cultured fibroblasts of individuals with ACDC not only contained increased TNAP activity, but they also calcified. Finally, TNAP activity and in vitro calcification could be reversed by genetic rescue with *NT5E* cDNA or treatment with exogenous adenosine.

The role of adenosine in inhibiting default calcification in vascular endothelial cells is a new concept that could have an enormous influence on our understanding of ectopic calcification in general. The pathway might also have implications for other specific disorders, including Monckeberg's medial sclerosis and pseudoxanthoma elasticum (Markello et al., 2011). In addition, knowledge of the basic defect in ACDC allows for consideration of treatment for affected individuals; besides the five siblings described above, four other ACDC patients have been ascertained worldwide, and still more have come to our attention recently. Treatment with bisphosphonates (inhibitors of alkaline phosphatase) is a potential therapy; therefore, we have initiated a clinical protocol to test this.

Novel mutation, unique phenotype

A second example of disease discovery emanating from the NIH UDP focused on a novel spastic ataxia-neuropathy syndrome in two brothers (Pierson et al., 2011). Phenotyping revealed spasticity, peripheral neuropathy, cerebellar ataxia, oculomotor apraxia, dystonia and myoclonic epilepsy. One of the brothers died at age 13; the other was deteriorating at age 14.

Again, consanguinity was involved, but this time in the form of a first cousin marriage. Whereas third cousins share 1/128 of their genes, first cousins share 1/8 of their genes, meaning that more than 10% of the human genome would be homozygous on SNP array analysis. As a consequence, SNP arrays would yield too many candidate genes to guess which ones would warrant further pursuit. Rather, WES was employed to identify genes containing homozygous variants. WES involves massively parallel sequencing of ~40 million bases that constitute the <2% of the human genome (3.2 billion bases) encoding genes. Using a whole exome platform, not every gene is perfectly sequenced (i.e. 'covered'), but the breadth is so great that there is a good chance of detecting a causative variant. The problem is that non-pathogenic variants abound, so the exome sequences of other family members are required to reduce ('filter') the total number of variants. In the case of the two brothers, a total of 120,469 variants were identified among the two parents and two affected brothers; this was reduced

to 59,482 variants not found in the 1000 Genomes database. Of these, 11 were homozygous and, of the 11, three were not present in dbSNP, a database of known polymorphisms. One homozygous missense mutation (c.1847G>A; p.Y616C) in the *AFG3L2* gene was identified as the best candidate. *AFG3L2* encodes a subunit of a mitochondrial AAA protease that removes damaged or misfolded proteins (Koppen et al., 2007).

As in the case of ACDC, we had a phenotype and a gene, but needed scientific expertise to demonstrate a functional deficit and a mechanism of disease. In this case, Thomas Langer and Elena Rugarli of the University of Cologne in Germany provided the basic research knowledge by using a yeast expression system to study *AFG3L2*. It was known that AFG3L2 protein forms either a homooligomeric isoenzyme or a hetero-oligomeric complex with paraplegin. Langer's group used yeast complementation assays to show that the mutant (Y616C) AFG3L2 protein is a hypomorphic variant with an impaired ability to form complexes with either itself or with paraplegin (Pierson et al., 2011). This finding was well aligned with the clinical phenotype, because paraplegin was mutated in hereditary spastic paraplegia type 7 (SPG7), and heterozygous *AFG3L2* mutations caused autosomal dominant spinocerebellar ataxia type 28 (SCA28) (Atorino et al., 2003; Koppen et al., 2007). The specific homozygous Y616C mutation in our patients produced a unique combination of the phenotypes of SPG7 and SCA28, along with other mitochondrial disease features such as oculomotor apraxia, extrapyramidal dysfunction and myoclonic epilepsy. This new disease discovery significantly broadened the phenotypes associated with both spinocerebellar ataxia and hereditary spastic paraplegia for the neurological community.

Conclusion

The above examples employed a wide range of research modalities – including cutting-edge medical procedures, state-of-the-art genetics, appropriate cell culture systems and model organisms – to come to a final diagnosis. The NIH UDP has many other cases similar to those mentioned above, at different stages of maturation, as well as examples of extremely rare, known diseases. Representative disorders include congenital disorder of glycosylation IIb (De Praeter et al., 2000) being investigated by immunologists, follicular hyperkeratosis being pursued by cell cycle experts and an autosomal dominant myopathy being studied by cell biologists. In every case, genetics ties the clinical investigator to the basic scientist, but some investigations lie fallow for want of bench researchers or experts in the disease. As our knowledge and data in genetics burgeon, we must engage the cognoscenti in a wide variety of fields to discover new diseases, new mechanisms of action and, ultimately, new therapeutic interventions. Only a union of clinical and basic researchers – and an awareness of the rich resource that lies in the population of rare disease patients – can achieve this. We hope that the UDP can serve as a model for similar programs at major medical centers throughout the world.

REFERENCES

- Atorino, L., Silvestri, L., Koppen, M., Cassina, L., Ballabio, A., Marconi, R., Langer, T. and Casari, G. (2003). Loss of m-AAA protease in mitochondria causes complex I deficiency and increased sensitivity to oxidative stress in hereditary spastic paraplegia. *J. Cell Biol.* **163**, 777-787.
- De Praeter, C. M., Gerwig, G. J., Bause, E., Nuytinck, L. K., Vliegthart, J. F. G., Breuer, W., Kamerling, J. P., Espeel, M. F., Martin, J. J., De Paepe, A. M. et al. (2000). A novel disorder caused by defective biosynthesis of N-linked

- oligosaccharides due to glucosidase I deficiency. *Am. J. Hum. Genet.* **66**, 1744-1756.
- Gahl, W. A. and Tift, C. J.** (2011). The NIH Undiagnosed Diseases Program: lessons learned. *JAMA* **305**, 1904-1905.
- Gahl, W. A., Markello, T. C., Toro, C., Fuentes Fajardo, K., Sincan, M., Gill, F., Carlson-Donohoe, H., Gropman, A., Pierson, T. M., Golas, G. et al.** (2011). The NIH Undiagnosed Diseases Program: insights into rare diseases. *Genet. Medicine* (in press).
- Koppen, M., Metodiev, M. D., Casari, G., Rugarli, E. I. and Langer, T.** (2007). Variable and tissue-specific subunit composition of mitochondrial m-AAA protease complexes linked to hereditary spastic paraplegia. *Mol. Cell. Biol.* **27**, 758-767.
- Markello, T. C., Pak, L. K., St. Hilaire, C., Dorward, H., Ziegler, S. G., Chen, M. Y., Chaganti, K., Nussbaum, R. L., Boehm, M. and Gahl, W. A.** (2011). Vascular pathology of medial arterial calcifications in *NT5E* deficiency: implications for the role of adenosine in pseudoxanthoma elasticum. *Mol. Genet. Metab.* **103**, 44-50.
- Pierson, T. M., Adams, D., Bonn, F., Martinelli, P., Cherukuri, P. F., Teer, J. K., Hansen, N. F., Cruz, P., Mullikin, J. C., Blakesley, R. W. et al.** (2011). Whole-exome sequencing identifies homozygous *AFG3L2* mutations in a spastic ataxia-neuropathy syndrome linked to mitochondrial m-AAA proteases. *PLoS Genetics* **7**, e1002325.
- St. Hilaire, C., Ziegler, S. G., Markello, T., Brusco, A., Groden, C., Gill, F., Carlson-Donohoe, H., Lederman, R. J., Chen, M. Y., Yang, D. et al.** (2011). *NT5E* mutations and arterial calcifications. *N. Engl. J. Med.* **364**, 432-442.

Forthcoming DMM articles

For articles published online ahead of print see <http://dmm.biologists.org/content/early/recent>
Sign up for email alerts (eToC) at <http://dmm.biologists.org/cgi/alerts/etoc>

Research Articles

- **Neuronal regeneration in a zebrafish model of adult brain injury.** Norihito Kishimoto, Kohei Shimizu and Kazunobu Sawamoto
- **Mutation of zebrafish dihydrolipoamide branched-chain transacylase E2 results in motor dysfunction and models maple syrup urine disease.** Timo Friedrich, Aaron M. Lambert, Mark A. Masino and Gerald B. Downes
- **Conserved valproic-acid-induced lipid droplet formation in *Dictyostelium* and human hepatocytes identifies structurally active compounds.** Lucy M. Elphick, Nadine Pawolleck, Irina A. Guschina, Leila Chaieb, Daniel Eikel, Heinz Nau, John L. Harwood, Nick J. Plant and Robin S. B. Williams
- **Metformin differentially activates ER stress signaling pathways without inducing apoptosis.** Thomas Quentin, Michael Steinmetz, Andrea Poppe and Sven Thoms
- **Model for long QT syndrome type 2 using human iPS cells demonstrates arrhythmogenic characteristics in cell culture.** Anna L. Lahti, Ville J. Kujala, Hugh Chapman, Ari-Pekka Koivisto, Mari Pekkanen-Mattila, Erja Kerelä, Jari Hyttinen, Kimmo Kontula, Heikki Swan, Bruce R. Conklin, Shinya Yamanaka, Olli Silvennoinen and Katriina Aalto-Setälä
- **Excessive activity of cathepsin K is associated with cartilage defects in a zebrafish model of mucopolysaccharidosis II.** Aaron C. Petrey, Heather Flanagan-Steet, Steven Johnson, Xiang Fan, Mitche De la Rosa, Mark E. Haskins, Alison V. Nairn, Kelley W. Moremen and Richard Steet
- ***BTN1*, the *Saccharomyces cerevisiae* homolog to the human Batten disease gene, is involved in phospholipid distribution.** Sergio Padilla-López, Deanna Langager, Chun-Hung Chan and David A. Pearce

Resource Article

- **Unraveling the ischemic brain transcriptome in a permanent middle cerebral artery occlusion mouse model by DNA microarray analysis.** Motohide Hori, Tomoya Nakamachi, Randeep Rakwal, Junko Shibato, Keisuke Nakamura, Yoshihiro Wada, Daisuke Tsuchikawa, Akira Yoshikawa, Keiji Tamaki and Seiji Shioda