The undiagnosed diseases program: Approach to diagnosis

Ellen F. Macnamara[a,*], Precilla D'Souza[a], Undiagnosed Diseases Network
and Cynthia J. Tifft[a,b]
[a]*National Institutes of Health, Undiagnosed Diseases Program, Common Fund,
Office of the Director, Bethesda, MD, USA*
[b]*Office of the Clinical Director, National Human Genome Research Institute,
National Institutes of Health, Bethesda, MD, USA*

Abstract. Undiagnosed and rare conditions are collectively common and affect millions of people worldwide. The NIH Undiagnosed Diseases Program (UDP) strives to achieve both a comprehensive diagnosis and a better understanding of the mechanisms of disease for many of these individuals. Through the careful review of records, a well-orchestrated inpatient evaluation, genomic sequencing and testing, and with the use of emerging strategies such as matchmaking programs, the UDP succeeds nearly 30 percent of the time for these highly selective cases. Although the UDP process is built on a unique set of resources, case examples demonstrate steps genetic professionals can take, in both clinical and research settings, to arrive at a diagnosis for their most challenging cases.

Keywords: Undiagnosed, rare disease, genome sequencing, model organisms, RNA sequencing

1. Introduction

The "diagnostic odyssey" is well known in medical genetics. Undiagnosed conditions affect three million Americans and include those with rare, difficult to identify conditions, atypical presentations of known conditions, and diseases yet to be discovered [1]. Patients and their families can spend years on this diagnostic odyssey and never arrive at a diagnosis. The average length of a diagnostic odyssey is eight years [1].

The Undiagnosed Diseases Program (UDP) at the National Institutes of Health (NIH) was established in May 2008. The goals of the UDP were two-fold: 1) to achieve a comprehensive diagnosis for patients who have already undergone an exhaustive evaluation and remain on the odyssey; and 2) to identify new biochemical, physiological, and cell biological pathways to better understand the mechanisms of disease. The many meanings of "diagnosis" include a simple histological description, a specific collection of clinical symptoms, or a genetic mutation. The UDP, however, believes that the "most satisfying definition of diagnosis includes an understanding of the disease pathogenesis, linking genetic and clinical findings and informing prognosis and therapy" [2].

---

*Corresponding author: Ellen F. Macnamara, National Institutes of Health, Undiagnosed Diseases Program, Common Fund, Office of the Director, Bethesda, MD, USA. E-mail: ellen.macnamara@nih.gov.

To reach these goals, UDP team members review all available past medical records of applicants. Accepted individuals receive careful and extensive clinical evaluations at the NIH Clinical Center. The patients and their families then undergo integrated genomic analyses. Following identification of candidate genes, collaborations are established to identify additional patients through matchmaking services, prove causality through *in vitro* cell studies, and perform functional assays in model systems.

In 2014, the UDP expanded to the Undiagnosed Diseases Network (UDN), and included six additional clinical sites, a coordinating center, two sequencing cores, a metabolomics core, a model organisms core, and a central biorepository. The UDN, modeled after the UDP, shares the mission of providing diagnoses and discovering new diseases through wider access to cross-disciplinary expertise and a collaborative network of experts [3]. This paper outlines the strategies employed by the UDP and UDN to achieve diagnoses, the lessons learned, and the application of these methods to clinical genetics.

2. Application and acceptance

Applicants to the UDN apply through an online portal with a referral letter from a clinician [4]. Applicants share medical and surgical records (for pediatric patients, birth records and growth curves are also requested), genetic testing results, imaging studies, and biopsy slides. These records are reviewed by experts in the disease category appropriate for the patient (e.g., neurology, rheumatology, immunology, etc.) and the UDN Board accepts cases with "objective findings, novel phenotypic manifestations, and a high likelihood of obtaining a diagnosis" [5]. Historically, the most informative notes for pediatric applicants have been consultant notes that speak to the patients' symptomatology, developmental notes that address the patients' trajectory, newborn records to assist in acquired versus congenital onset, and growth curves. Occasionally, diagnoses are made by careful review of these records; this is most commonly accomplished through receipt of original genetic testing reports that may have been previously misinterpreted.

Since the inception of the UDN, approximately 39% of applicants have been accepted network wide. Applicants have a wide-range of objective and subjective symptoms. Walley, et. al., reviewed 151 "Not Accepted" patients and 50 "Accepted" patients in the UDN to determine what distinguished these groups from one another [6]. On average, accepted patients were younger, had a longer percentage of time with illness and an earlier onset of disease, and were referred by specialists. Neurology was the primary category for symptoms in both the "Accepted" and "Not Accepted" patients, followed by allergy & immunology and musculoskeletal disease.

Making a distinction between undiagnosed individuals with and without objective findings increases the likelihood of successful evaluation in the UDN. Those with objective findings often have a genetic etiology and therefore are more likely to benefit from the resources of the UDN [6].

3. Admission and evaluation

The UDN functions under a single protocol with a common Institutional Review Board and enrolls all patients and consenting family members [7]. Patients accepted to the NIH UDP, one of the UDN clinical sites, are consented and admitted to the NIH Clinical Center, generally for a five-day comprehensive inpatient evaluation. The Clinical Center is supported by intramural NIH funding and patients are enrolled regardless of their ability to pay. This assures that patients are accepted into the UDP on the uniqueness of their clinical case. Alternative approaches across other NIH-funded UDN sites include both inpatient and outpatient models. The cost is borne by a combination of third party payors for clinically-indicated testing and extramural NIH grant funding.

Patients at NIH are seen by a wide range of specialists engaged in a highly collaborative diagnostic approach. The standard pediatric UDP consultations include: genetics, nutrition, neurology, audiology, ophthalmology, neuropsychiatry, rehabilitative medicine, and physical, speech, and occupational therapy. More than half of the pediatric patients, i.e., 56% between 2008 and 2015 (unpublished observations), received a single 3-4-hour long sedation allowing for neuroimaging and multiple evaluations that were potentially painful or difficult to accomplish in an awake child. Patients had an average of three procedures following imaging; they include lumbar puncture, skin biopsy, and dilated eye exam. Additional procedures often involved auditory evoked response testing, dental evaluation and cleaning, electromyogram and nerve conduction studies, dysmorphology evaluation in an uncooperative patient, catheterization for urine collection, and large blood draws. More than 300 children, many with multisystem disease American Society of Anesthesiology (ASA) Score III and above, have been sedated without complications. The NIH diagnostic evaluations have led to clinical diagnoses, management recommendations, targeted testing, and occasionally treatment.

At the conclusion of the week-long evaluation the proband's phenotype is recorded in a secure database using standardized terms taken from the Human Phenotype Ontology (HPO) [8]. PhenoTips® software, embedded within the uniquely designed UDPICS database, facilitates entry of HPO terms [9–11]. Since accurately identifying family members as "affected" or "unaffected" with respect to the proband's disorder is critical for genetic analysis, family members are often evaluated for subtle manifestations of the proband's phenotype; the use of HPO terms allows for harmonization of phenotypic information across family members.

HPO terms are publicly available through https://hpo.jax.org/app/ and the goal is to enable the integration of phenotypic information across scientific fields and databases and even species [12]. Since its main application involves rare disorders, HPO terms have been informed by OMIM, Orphanet, and Decipher entries. There are many benefits to using HPO terms, including specificity and creation of an annotation profile that allows for identification of candidate genes, candidate pathways, and potential model organisms. Additionally, HPO terms allow for comparisons across groups and research endeavors by creating standardized descriptions of symptoms. Greater use of HPO terms across the genetics community, for example when submitting clinical sequencing requests, would facilitate case sharing and recognition of similarly affected individuals.

4. Genomic sequencing and testing

Since many rare disorders are genetically-based, the most effective tool at the disposal of the UDP has been sequencing [1]. Over the last 5 years the majority of children and a smaller number of adult applicants have already had non-diagnostic exome sequencing prior to their acceptance into the program. Hence, additional sequencing (including exome, genome, and/or RNA) performed on the proband and family members and analyzed through research pipelines have been important for diagnosis.

4.1. Sequencing analysis

The NIH UDP approach to exome and genome analysis is based on the implicit understandings that (1) there are limited pre-sequencing clues for diagnosis, (2) families have seemingly unique constellation of features, and (3) pedigrees are too small for conventional linkage analysis [13]. From 2008 to 2012, sequencing in the UDP was performed through the NIH Intramural Sequencing Center Comparative Sequencing Program. Since 2012, however, sequencing has been performed by one of two clinical laboratories funded by the UDN. An initial Clinical Laboratory Improvement Amendments

(CLIA) report is generated for probands and family members; if no known disorder is identified, then research analysis begins.

The UDP begins with an annotated variant candidate list from the exome or, now more commonly, the genome, sequencing data [14]. Each variant is given a genotype quality score based on a Bayesian statistic of the Most Probable Genotype (MPG) and a ratio of MPG to the coverage of any given variant [15]. Variants are then filtered for variant type, giving priority to coding sequence variants that result in missense or nonsense mutations, canonical slice site variants, or insertions/deletions. This list is further filtered by population frequency, assuming that disease-causing alleles in our population are likely rare, highly penetrant, and responsible for significant health problems [13].

The candidate list then requires manual curation. Variants within highly polymorphic genes (e.g., *TTN)* can be excluded; other individual base pairs are excluded due to consistent alignment problems, with the reference sequences having minor alleles [16]. This list is then filtered for segregation with disease among family members. SNP array data are also incorporated to reveal recombination mapping, mosaicism, regions of homozygosity, uniparental disomy, confirmation of parentage, and detection of copy number variants [17, 18]. The smaller candidate list is then reviewed for individual variants' goodness of fit with clinical data, pathogenicity, and previously reported information. Pierson, et al, describes one of the first cases diagnosed with this process in the UDP [19].

Over time, the NIH UDP processes have evolved to include strategies to solve the completeness problem including exome capture, inclusion of intronic variants, and evaluation of medium-sized structural variants [20]. Automated programs have also been developed for ethnicity matched genotyping, salvage pathways for Mendelian inconsistencies, exon deletion filtration, and pedigree aware BAM file noise evaluation [21]. Internally referred to as the "forwards-backwards" analysis, this most recent toolset demonstrates the successful implementation of our analytic pipeline, since affected individuals had significantly more seemingly pathogenic variants than their unaffected siblings. In other words, the pipeline is developed with enough constraint to keep the likely causal variant in the candidate list without creating an unmanageable number of variants.

4.2. Novel inheritance of known disease genes

The key to interpreting genetic analyses is determining which variants may be causing a proband's disease. It is well known that different variants in the same gene can lead to different diseases, such as beta-galactosidase mutations causing either GM1 gangliosidosis or mucopolysaccharidosis type IVB, or *BRCA2* variants causing either susceptibility to breast cancer or Fanconi anemia. Nevertheless, variants are frequently ruled out when a known disease is caused by variants in the gene but the phenotype of the proband does not match.

We have identified two novel human diseases by considering new inheritance patterns of known disease-causing genes. Monoallelic variants in *GARS* are known to cause Charcot-Marie-Tooth disease, type 2D (OMIM 601472). A seven-year-old female with extreme growth retardation and multiple organ involvement was found to have compound heterozygous variants in *GARS* [22]. Neither parent had symptoms of Charcot-Marie-Tooth disease and both had normal nerve conduction velocities. Functional studies demonstrated loss-of-function for both variants and, given the proband's overlap with other ARS-related recessive conditions, the phenotypic and inheritance pattern of *GARS*-associated diseases was expanded.

Saul-Wilson syndrome (OMIM 618150), is a rare skeletal dysplasia described in 1982, but the genetic etiology was unknown until 2018 [23]. A UDN participant was evaluated and met clinical criteria for Saul-Wilson. Genomic sequencing analysis identified a *de novo* variant in *COG4*. Biallelic variants in *COG4*, however, are the known cause of type IIj congenital disorder of glycosylation (OMIM 613489). The UDN recruited twelve new patients with Saul-Wilson syndrome and confirmed that this disease

is caused by the same recurrent monoallelic variant in *COG4* (p.Gly516Arg), likely producing a gain of function.

4.3. Genome sequencing – Structural variant calling

Since 2013, genome sequencing has been employed for UDP participants with prior non-diagnostic clinical exome sequencing and a compelling phenotype. Genome sequencing allows for better detection of structural variants as demonstrated by our recent discovery of a novel disease, Kilquist syndrome [24, 25]. The proband was previously known to have uniparental isodisomy for chromosome 5, so genome analysis targeted this region for candidate genes. Trio sequencing identified a 22kb homozygous deletion of *SLC12A2*, that had not been identified by prior exome sequencing. We compared the HPO phenotype terms for the proband with the findings in a previously published murine model of SLC12A2 (NKCC1) and found compelling overlap, suggesting that this deletion was causative [26, 27].

5. Emerging strategies

5.1. Matchmaking – Internally

Participation in a network, with multiple sites evaluating and sequencing patients, enables data-sharing and, therefore, matchmaking, across sites [28]. Clinical sequencing labs employ similar strategies, but the robust clinical information available through the UDN increases its ability to find similarly affected patients.

Exome sequencing identified seven individuals with significant phenotypic overlap and *de novo* variants in *TRAF7* (Tokita, 2018). When the UDP was alerted to this match, an internal database search was performed and an additional patient with a *de novo TRAF7* variant and substantial phenotypic overlap was identified. We frequently employ this searching strategy for potential matches as new disease genes are identified both within and outside of our cohort.

5.2. Matchmaking – Externally

5.2.1. PhenomeCentral

Matchmaking services, or "genetic dating sites", have been developed to allow researchers and clinicians to find similarly affected individuals using phenotypic or genetic data. The UDN utilizes PhenomeCentral, a restricted access network for clinicians, researchers, and scientists, to share patient phenotype and genotype data [29]. De-identified patient information is submitted to the database and the user is provided a list of cases that appear most similar to the data submitted. Cases can be submitted as private (not shared in matchmaking), match-able (seen only when matched), or public (visible to all users).

PhenomeCentral participates in the Matchmaker Exchange (MME), a federated network of databases with genotypic and phenotypic information through an application programming interface (API) [30]. MME enables searches of multiple databases with a single query and users can choose where to deposit their data depending on the type of data they are submitting. Other sites within MME include GeneMatcher, DECIPHER, MyGene2, and *matchbox* [31–34]. Increased use of MME by clinicians will improve the ability to identify new diseases and provide diagnoses for patients and their families across platforms.

5.2.2. Social media outreach

In addition to secure matchmaking sites, the UDN employs social media to share patient data from consented individuals. Participant Pages are created to summarize a proband's medical history, significant findings, and candidate genes. The goal of these pages is two-fold: to find similarly affected individuals who may help clarify the cause of the condition; and to find external researchers who may have expertise in the candidate gene. Participant Pages are modeled after the success of Matthew and Cristina Might in identifying additional cases of *NGLY1* deficiency. By harnessing the power of social media and a well-written blog post, the Might family identified additional patients with variants in *NGLY1* [35, 36]. Finding additional patients led to the identification of a new disease and a natural history study [37, 38].

The UDN has successfully utilized these pages, and new disease discoveries have benefited from the addition of probands found through Participant Pages. The newly described syndrome caused by variants in *NACC1* utilized these webpages to identify two additional patients for their cohort [39]. A fifth patient was identified with variants in *IRF2BPL* utilizing a gene page, designed in the same format as participant pages, but focusing on the gene itself [40]. These pages are subsequently shared on Facebook and Twitter to reach participants directly. Rare disease patients are already using social media to find one another; these pages increase our reach and allow us to "meet" participants where they are [41].

5.3. RNA sequencing

RNA sequencing (RNA-seq) allows for the direct probing of variation in RNA content and in RNA sequence. In combination with exome or genome sequencing, RNA-seq can help prioritize variants that would otherwise be uninterpretable or lost to filtration. RNA-seq is especially helpful in determining the effect of splice variants and can occasionally identify splicing defects not recognized in exome and genome sequence data.

A recent UDP case was solved using just this strategy. Agnostic analysis of RNA-seq data for splicing variants identified a novel splice junction in a patient with a non-diagnostic genome sequence (unpublished data). The splice junction was not observed in any control sample and review of the genome data showed a de novo, deep intronic (+1242bp) single nucleotide variant believed to be responsible for the aberrant splicing in *PHIP* (OMIM #617991). The proband's phenotype was consistent with published cases of individuals with truncating variants in *PHIP*, thus identifying the diagnosis for this proband.

5.4. Model organisms

The goal of the UDN Model Organism Screening Center (MOSC) is to provide compelling data from studies in c. elegans, drosophila, and zebrafish that either support or refute disease causality of specific variants [42]. To do this, the MOSC utilizes a standard pipeline for review of candidate genes. Variants are prioritized for potential study in a model organism when they: 1) are in novel or candidate disease-causing genes; or 2) are novel variants in a known disease-causing gene in a patient with a novel phenotype.

Variants that pass an initial quality control review (i.e., rare, no existing disease association, and identification of potential second case) are then reviewed by each model organism group. Cases are further prioritized based on whether there is an ortholog or paralog in the model system, whether the amino acid is conserved in the species, and whether the gene or variant has been previously studied in a model system. If a case meets criteria, it is assigned a model organism and the group begins their research.

The MOSC has been instrumental in solving cases in the UDN. One such case was that of a 7-year-old male with global developmental delay, hypotonia, expressive speech delay, intellectual disability, and dysmorphic features. Trio exome sequencing identified a *de novo* variant in *EBF3* [43]. Two additional probands were identified with *de novo* variants at the same amino acid and overlapping phenotypes. The phenotype was recapitulated in a fly model. Flies were selected for this project as the Knot gene has a 62% identify and a 70% similarity with human *EBF3*. Multiple experiments demonstrated that the two *EBF3* variants failed to rescue the lethality of the Knot knockout flies, while wildtype *EBF3* was able to rescue Knot knockouts, thus demonstrating the loss of function of the *de novo EBF3* variants. This functional validation of the *EBF3* variants identified a new neurodevelopmental disorder and provided three families with a diagnosis.

5.5. Metabolomics

The UDN Metabolomics Core provides "untargeted and targeted quantitative metabolomic approaches for bioinformatic, and clinical interpretation of specimens" [42]. The Core performs global, untargeted glycan, lipid, and mitochondrial metabolite profiling to identify priority targets for further study. The core continues to develop assays for identifying new metabolites and offers interpretations of identified metabolites.

Since inborn errors of metabolism have been associated with a wide range of symptoms that affect multiple organ systems, abnormal metabolites are a consideration for any UDN participant. Because metabolites in plasma or urine are subject to external factors, such as medications, diet, and supplements, metabolomic analysis in the UDN is performed on cultured fibroblasts grown and prepared under identical controlled conditions. Although metabolomics has not yet solved a case, we are hopeful that their agnostic and hypothesis driven analyses will improve our diagnostic capabilities, both in identifying and ruling out candidate genes.

6. Conclusions

The UDN evaluates a highly selected group of participants who have usually spent years on the diagnostic odyssey traveling to multiple academic centers in search of a diagnosis. Despite using many cutting-edge tools, the UDN fails more often than it succeeds; only 27.5% of all UDP participants (35% of pediatric participants) have received a diagnosis. However, for many of the 286 individuals who have participated in the UDP over the last 10 years, receiving a diagnosis has been life-changing. For some diagnoses, treatment is possible, for others a change in management or access to additional services has been facilitated. Diagnoses that have led to treatment include: diagnoses of *KMT2B* which have led to treatment with deep-brain stimulation and treatment of congenital serine biosynthesis defect due to biallelic *PSAT1* variants with serine [44, 45]. For other families, diagnoses of *ADNP*, *KAT6A*, *GLB1* mutations among many others have led to inclusion in both ongoing natural history studies and a gene therapy trial (NCT03952637). For some families, determining an inheritance pattern has provided family planning options and healthy children have been born to parents who would have otherwise avoided pregnancy. Some families with the same diagnosis have joined together to start advocacy groups for mutual support and to raise funds for therapeutic research. Yet for most participants who receive a diagnosis the odyssey is over, and they can move on to the next phase of their meaningful lives [3]. What's in a name? Everything!!

Members of the Undiagnosed Diseases Network

Maria T. Acosta	Sirisak Chanprasert	Rizwan Hamid	May Christine V. Malicdan	Genecee Renteria	Dave Viskochil
Margaret Adam	Hsiao-Tuan Chao	Neil A. Hanchard	Laura A. Mamounas	Chloe M. Reuter	Tiphanie P. Vogel
David R. Adams	Mei-Jan Chen	Nichole Hayes	Teri A. Manolio	Lynette Rives	Colleen E. Wahl
Pankaj B. Agrawal	Gary D. Clark	Frances High	Rong Mao	Amy K. Robertson	Stephanie Wallace
Mercedes E. Alejandro	Terra R. Coakley	Anne Hing	Kenneth Maravilla	Lance H. Rodan	Nicole M. Walley
Justin Alvey	Laurel A. Cobban	Fuki M. Hisama	Thomas C. Markello	Jill A. Rosenfeld	Chris A. Walsh
Laura Amendola	Joy D. Cogan	Ingrid A. Holm	Ronit Marom	Natalie Rosenwasser	Melissa Walker
Ashley Andrews	F. Sessions Cole	Jason Hom	Gabor Marth	Robb K. Rowley	Jennifer Wambach
Euan A. Ashley	Heather A. Colley	Martha Horike-Pyne	Beth A. Martin	Maura Ruzhnikov	Jijun Wan
Mahshid S. Azamian	Cynthia M. Cooper	Alden Huang	Martin G. Martin	Ralph Sacco	Lee-kai Wang
Carlos A. Bacino	Heidi Cope	Yong Huang	Julian A. Martínez-Agosto	Jacinda B. Sampson	Michael F. Wangler
Guney Bademci	William J. Craigen	Rosario Isasi	Shruti Marwaha	Susan L. Samson	Patricia A. Ward
Eva Baker	Michael Cunningham	Fariha Jamal	Jacob McCauley	Mario Saporta	Daniel Wegner
Ashok Balasubramanyam	Precilla D'Souza	Gail P. Jarvik	Allyn McConkie-Rosell	C. Ron Scott	Mark Wener
Dustin Baldridge	Hongzheng Dai	Jeffrey Jarvik	Colleen E. McCormack	Judy Schaechter	Tara Wenger
Jim Bale	Surendra Dasari	Suman Jayadev	Alexa T. McCray	Timothy Schedl	Katherine Wesseling Perry
Michael Bamshad	Mariska Davids	Jean M. Johnston	Elisabeth McGee	Kelly Schoch	Monte Westerfield
Jordan Barham	Jyoti G. Dayal	Lefkothea Karaviti	Heather Mefford	Daryl A. Scott	Matthew T. Wheeler
Deborah Barbouth	Esteban C. Dell'Angelica	Emily G. Kelley	J. Lawrence Merritt	Lisa Shakachite	Anastasia L. Wise
Gabriel F. Batzli	Shweta U. Dhar	Dana Kiley	Matthew Might	Prashant Sharma	Lynne A. Wolfe
Pinar Bayrak-Toydemir	Katrina Dipple	Isaac S. Kohane	Ghayda Mirzaa	Vandana Shashi	Jeremy D. Woods
Anita Beck	Daniel Doherty	Jennefer N. Kohler	Eva Morava-Kozicz	Jimann Shin	Shinya Yamamoto
Alan H. Beggs	Naghmeh Dorrani	Deborah Krakow	Paolo M. Moretti	Rebecca Signer	John Yang
Gill Bejerano	Emilie D. Douine	Donna M. Krasnewich	Marie Morimoto	Catherine H. Sillari	Guoyun Yu
Hugo J. Bellen	David D. Draper	Susan Korrick	John J. Mulvihill	Edwin K. Silverman	Diane B. Zastrow
Jimmy Bennet	Laura Duncan	Mary Koziura	David R. Murdock	Janet S. Sinsheimer	Chunli Zhao
Beverly Berg-Rood	Dawn Earl	Joel B. Krier	Mariko Nakano-Okuno	Kathy Sisco	Stephan Zuchner
Raphael Bernier	David J. Eckstein	Seema R. Lalani	Avi Nath	Edward C. Smith	
Jonathan A. Bernstein	Lisa T. Emrick	Byron Lam	Stan F. Nelson	Kevin S. Smith	
Gerard T. Berry	Christine M. Eng	Christina Lam	John H. Newman	Lilianna Solnica-Krezel	
Anna Bican	Cecilia Esteves	Brendan C. Lanpher	Sarah K. Nicholas	Rebecca C. Spillmann	
Stephanie Bivona	Tyra Estwick	Ian R. Lanza	Deborah Nickerson	Joan M. Stoler	
Elizabeth Blue	Liliana Fernandez	C. Christopher Lau	Donna Novacic	Nicholas Stong	
John Bohnsack	Carlos Ferreira	Kimberly LeBlanc	Devin Oglesbee	Jennifer A. Sullivan	
Carsten Bonnenmann	Elizabeth L. Fieg	Brendan H. Lee	James P. Orengo	Angela Sun	
Devon Bonner	Paul G. Fisher	Hane Lee	Laura Pace	Shirley Sutton	
Lorenzo Botto	Brent L. Fogel	Roy Levitt	Stephen Pak	David A. Sweetser	
Brenna Boyd	Irman Forghani	Richard A. Lewis	J. Carl Pallais	Virginia Sybert	
Lauren C. Briere	Laure Fresard	Sharyn A. Lincoln	Christina GS. Palmer	Holly K. Tabor	
Elly Brokamp	William A. Gahl	Pengfei Liu	Jeanette C. Papp	Cecelia P. Tamburro	
Gabrielle Brown	Ian Glass	Xue Zhong Liu	Neil H. Parker	Queenie K.-G. Tan	
Elizabeth A. Burke	Rena A. Godfrey	Nicola Longo	John A. Phillips III	Mustafa Tekin	
Lindsay C. Burrage	Katie Golden-Grant	Sandra K. Loo	Jennifer E. Posey	Fred Telischi	
Manish J. Butte	Alica M. Goldman	Joseph Loscalzo	John H. Postlethwait	Willa Thorson	
Peter Byers	David B. Goldstein	Richard L. Maas	Lorraine Potocki	Cynthia J. Tifft	
William E. Byrd	Alana Grajewski	Ellen F. Macnamara	Barbara N. Pusey	Camilo Toro	
John Carey	Catherine A. Groden	Calum A. MacRae	Aaron Quinlan	Alyssa A. Tran	
Olveen Carrasquillo	Andrea L. Gropman	Valerie V. Maduro	Wendy Raskind	Tiina K. Urv	
Ta Chen Peter Chang	Sihoun Hahn	Marta M. Majcherska	Archana N. Raja	Matt Velinder	

References

[1] G. Genes, Rare Facts. 2015 [cited 2019; Available from: https://globalgenes.org/rare-facts/.

[2] W.A. Gahl and C.J. Tifft, The NIH Undiagnosed Diseases Program: Lessons learned, *JAMA* **305**(18) (2011), 1904–1905.

[3] K. Splinter, et al., Effect of Genetic Diagnosis on Patients with Previously Undiagnosed Disease, *N Engl J Med* **379**(22) (2018), 2131–2139.

[4] W.A. Gahl, A.L. Wise and E.A. Ashley, The Undiagnosed Diseases Network of the National Institutes of Health: A National Extension, *JAMA* **314**(17) (2015), 1797–1798.

[5] W.A. Gahl, et al., The NIH Undiagnosed Diseases Program and Network: Applications to modern medicine, *Mol Genet Metab* **117**(4) (2016), 393–400.

[6] N.M. Walley, et al., Characteristics of undiagnosed diseases network applicants: Implications for referring providers, *BMC Health Serv Res* **18**(1) (2018), 652.

[7] K. Splinter, et al., Implementing the Single Institutional Review Board Model: Lessons from the Undiagnosed Diseases Network, *Clin Transl Sci* **11**(1) (2018), 28–31.

[8] T. Groza, et al., The Human Phenotype Ontology: Semantic Unification of Common and Rare Disease, *Am J Hum Genet* **97**(1) (2015), 111–124.

[9] M. Girdea, et al., PhenoTips: Patient phenotyping software for clinical and research use, *Hum Mutat* **34**(8) (2013), 1057–1065.

[10] W.P. Bone, et al., Computational evaluation of exome sequence data using human and model organism phenotypes improves diagnostic efficiency, *Genet Med* **18**(6) (2016), 608–617.

[11] J. Guzman, et al., The Undiagnosed Diseases Program Integrated Collaboration System (UDPICS): One Program's Experience Developing Custom Software to Support Research for Complex-Disease Families, *Children (Basel)* **2**(3) (2015), 342–357.

[12] S. Kohler, et al., The Human Phenotype Ontology in 2017, *Nucleic Acids Res* **45**(D1) (2017), D865–D876.

[13] D.R. Adams, et al., Analysis of DNA sequence variants detected by high-throughput sequencing, *Hum Mutat* **33**(4) (2012), 599–608.

[14] M. Sincan, et al., VAR-MD: A tool to analyze whole exome-genome variants in small human pedigrees with mendelian inheritance, *Hum Mutat* **33**(4) (2012), 593–598.

[15] J.K. Teer, et al., Systematic comparison of three genomic enrichment methods for massively parallel DNA sequencing, *Genome Res* **20**(10) (2010), 1420–1431.

[16] K.V. Fuentes Fajardo, et al., Detecting false-positive signals in exome sequencing, *Hum Mutat* **33**(4) (2012), 609–613.

[17] T.C. Markello, et al., Sensitive quantification of mosaicism using high density SNP arrays and the cumulative distribution function, *Mol Genet Metab* **105**(4) (2012), 665–671.

[18] T.C. Markello, et al., Recombination mapping using Boolean logic and high-density SNP genotyping for exome sequence filtering, *Mol Genet Metab* **105**(3) (2012), 382–389.

[19] T.M. Pierson, et al., Exome sequencing as a diagnostic tool in a case of undiagnosed juvenile-onset GM1-gangliosidosis, *Neurology* **79**(2) (2012), 123–126.

[20] C. Du, et al., Explorations to improve the completeness of exome sequencing, *BMC Med Genomics* **9**(1) (2016), 56.

[21] F. Gu, et al., A suite of automated sequence analyses reduces the number of candidate deleterious variants and reveals a difference between probands and unaffected siblings. Genet Med, 2019.

[22] S.N. Oprescu, et al., Compound heterozygosity for loss-of-function GARS variants results in a multisystem developmental syndrome that includes severe growth retardation, *Hum Mutat* **38**(10) (2017), 1412–1420.

[23] C.R. Ferreira, et al., A Recurrent De Novo Heterozygous COG4 Substitution Leads to Saul-Wilson Syndrome, Disrupted Vesicular Trafficking, and Altered Proteoglycan Glycosylation, *Am J Hum Genet* **103**(4) (2018), 553–567.

[24] E.F. Macnamara, et al., Kilquist Syndrome: A Novel Syndromic Hearing Loss Disorder Caused by Homozygous Deletion of SLC12A2. Hum Mutat, 2019.

[25] A.C. Lionel, et al., Improved diagnostic yield compared with targeted gene sequencing panels suggests a role for whole-genome sequencing as a first-tier genetic test, *Genet Med* **20**(4) (2018), 435–443.

[26] E. Delpire, et al., Deafness and imbalance associated with inactivation of the secretory Na-K-2Cl co-transporter, *Nat Genet* **22**(2) (1999), 192–195.

[27] M. Flagella, et al., Mice Lacking the Basolateral Na-K-2Cl Cotransporter Have Impaired Epithelial Chloride Secretion and Are Profoundly Deaf, *Journal of Biological Chemistry* **274**(38) (1999), 26946–26955.

[28] C.A. Brownstein, et al., Data sharing in the undiagnosed diseases network, *Hum Mutat* **36**(10) (2015), 985–988.

[29] O.J. Buske, et al., PhenomeCentral: A portal for phenotypic and genotypic matchmaking of patients with rare genetic diseases, *Hum Mutat* **36**(10) (2015), 931–940.

[30] A.A. Philippakis, et al., The Matchmaker Exchange: A platform for rare disease gene discovery, *Hum Mutat* **36**(10) (2015), 915–921.

[31] N. Sobreira, et al., GeneMatcher: A matching tool for connecting investigators with an interest in the same gene, *Hum Mutat* **36**(10) (2015), 928–930.

[32] E.A. Chatzimichali, et al., Facilitating collaboration in rare genetic disorders through effective matchmaking in DECIPHER, *Hum Mutat* **36**(10) (2015), 941–949.

[33] Website aims to accelerate gene discovery, diagnosis, treatment: MyGene2.org fosters open sharing among families, researchers, and clinicians, *Am J Med Genet A* **170**(6) (2016), 1388–1389.

[34] H. Arachchi, et al., matchbox: An open-source tool for patient matching via the Matchmaker Exchange, *Hum Mutat* **39**(12) (2018), 1827–1834.

[35] M. Might, Hunting down my son's killer. 2012: http://matt.might.net/articles/my-sons-killer/.

[36] M. Might and C.C. Might, What happens when $N = 1$ and you want plus 1? *Prenat Diagn* **37**(1) (2017), 70–72.

[37] G.M. Enns, et al., Mutations in Ngly1 Cause an Inherited Disorder of the Endoplasmic Reticulum-Associated Degradation (Erad) Pathway, *Molecular Genetics and Metabolism* **111**(3) (2014), 236–237.

[38] C. Lam, et al., Prospective phenotyping of NGLY1-CDDG, the first congenital disorder of deglycosylation, *Genet Med* **19**(2) (2017), 160–168.

[39] K. Schoch, et al., A Recurrent De Novo Variant in NACC1 Causes a Syndrome Characterized by Infantile Epilepsy, Cataracts, and Profound Developmental Delay, *Am J Hum Genet* **100**(2) (2017), 343–351.

[40] P.C. Marcogliese, et al., IRF2BPL Is Associated with Neurological Phenotypes, *Am J Hum Genet* **103**(2) (2018), 245–260.

[41] K.F. Lambertson, et al., Participant-driven matchmaking in the genomic era, *Hum Mutat* **36**(10) (2015), 965–973.

[42] About the Undiagnosed Diseases Network.

[43] H.T. Chao, et al., A Syndromic Neurodevelopmental Disorder Caused by De Novo Variants in EBF3, *Am J Hum Genet* **100**(1) (2017), 128–137.

[44] E. Meyer, et al., Mutations in the histone methyltransferase gene KMT2B cause complex early-onset dystonia, *Nat Genet* **49**(2) (2017), 223–237.

[45] S. Debs, et al., Adult diagnosis of congenital serine biosynthesis defect: A treatable cause of progressive neuropathy. In Progress, 2019.