# UCLA
## UCLA Previously Published Works

**Title**

A comprehensive iterative approach is highly effective in diagnosing individuals who are exome negative.

**Permalink**

**Journal**

**ISSN**

**Authors**

Shashi, Vandana
Schoch, Kelly
Spillmann, Rebecca
et al.

**Publication Date**

2019

**DOI**

Peer reviewed

# A Comprehensive Iterative Approach is Highly Effective in Diagnosing Individuals who are Exome Negative

**Vandana Shashi, MD, MBBS**[1], **Kelly Schoch, MS**[1], **Rebecca Spillmann, MS**[1], **Heidi Cope, MS**[1], **Queenie K.-G. Tan, MD, PhD**[1], **Nicole Walley, MS**[1], **Loren Pena, MD, PhD**[1], **Allyn McConkie-Rosell**[1], **Yong-Hui Jiang, MD, PhD**[1], **Nicholas Stong, PhD**[2], **Anna C. Need**[3], **Undiagnosed Diseases Network**, and **David B. Goldstein, PhD**[2]

[1]Department of Pediatrics, Division of Medical Genetics, Duke University School of Medicine< Durham NC

[2]Institute for Genomic Medicine, Columbia University, New York, NY

[3]Division of Brain Sciences, Department of Medicine, Imperial College London, London, UK

## Abstract

**Purpose**—Sixty-75% of individuals with rare and undiagnosed phenotypes remain undiagnosed after whole exome sequencing (ES). With standard ES reanalysis resolving 10–15% of the ES negatives, further approaches are necessary to maximize diagnoses in these individuals.

**Methods**—In 38 ES negative patients an individualized genomic-phenotypic approach was employed utilizing: A) Phenotyping; B) Reanalyses of FASTQ files, with innovative bioinformatics; C) Targeted molecular testing; D) Whole genome sequencing (WGS) and E) Conferring of clinical diagnoses when pathognomonic clinical findings occurred.

**Results**—Certain and Highly Likely diagnoses were made in 18/38 (47%) individuals, including identifying two new developmental disorders. The majority of diagnoses (>70%) were due to our bioinformatics, phenotyping and targeted testing identifying variants that were undetected or not prioritized on prior ES. WGS diagnosed 3/18 individuals, with structural variants not amenable to ES. Additionally, Tentative diagnoses were made in three (8%) and in five individuals (13%) candidate genes were identified. Overall, diagnoses/potential leads were identified in 26/38 (68%).

**Conclusions**—Our comprehensive approach to ES negatives maximizes the ES and clinical data for both diagnoses and candidate gene identification, without WGS in the majority. This iterative approach is cost-effective and is pertinent to the current conundrum of ES negatives.

Corresponding Author: Vandana Shashi, MD, Professor of Pediatrics, Division of Medical Genetics, Duke University School of Medicine, Tel: (919) 681-2616, Fax: (919) 668-0414, vandana.shashi@duke.edu.

## INTRODUCTION

Whole exome sequencing (ES) has transformed the diagnostic approach to rare and undiagnosed Mendelian phenotypes, with diagnosis rates of 25–50%. [1–4] However, 50–75% of individuals remain undiagnosed after ES (ES negatives). The next steps after a negative ES are currently limited. Some commercial laboratories offer one free ES reanalysis and this can provide a diagnoses in 10%–15%, with the majority (~70%) occurring due to interim new gene-disease associations. [5,6] Other studies have reported diagnosis rates of 15–36% with ES reanalyses: although the raw data are reanalyzed, the diagnoses are mostly related to resequencing singletons as trios, looking for copy number variants (CNV), literature reports, and case matching through platforms such as Matchmaker exchange. [7,8] Resequencing is reported to result in ~15% new molecular diagnoses, due to addition of family members and new gene-disease associations. [9] We reported that reanalyses of ES can improve the diagnostic yield due to phenotypic characterization, improved exome coverage, realignment and variant calling in addition to new disease gene discovery [10] and careful phenotyping leading to targeted molecular testing can detects variants missed by ES [11]. Whole genome sequencing (WGS) can be an option for ES negative patients, with its ability to detect variants in noncoding regions, uniform coverage and better detection of structural variants; ~15% of variants missed by ES may be detected by WGS[12–14]. However, WGS is not widely available clinically and is not covered by third party payers and thus ES remains the standard diagnostic approach to rare Mendelian phenotypes.

There are many reasons why ES may miss variants of interest. Firstly, the underlying genetic etiology may be non-Mendelian and thus not amenable to ES (e.g., complex diseases). Secondly, the underlying disorder may be Mendelian, but due to variants undetectable with ES technology (e.g., trinucleotide repeats). Finally, variants that should be tractable to ES may still not be detected or reported due to variants not being detected or not recognized as disease-causing. [10] This may occur due to: 1) Analytical factors/differences (e.g. difficult regions of exome, different quality filters); [15,16] 2) Knowledge gaps since initial ES (e.g. evolving phenotypes, gene-disease relationships not well established); [17,18] 3) Interpretation/ reporting differences between labs (e.g. variant not reported due to poor phenotypic fit). [11,19]

The Undiagnosed Diseases Network (UDN) (https://undiagnosed.hms.harvard.edu) is a nationwide NIH-funded research study that accepts patients with intractable phenotypes for further diagnostic resolution. Genomic sequencing is a major component of the UDN, since most undiagnosed and rare diseases (~85%) are believed to be genetic. [20] At the Duke/ Columbia clinical site of the UDN, we observed that the majority (~60%) who enter the study have a negative ES result through prior commercial or research sequencing. In comparison to previous studies, these individuals are among the most challenging, with the majority having undergone trio ES prior to the UDN and in some instances, an ES

reanalyses as well. We thus devised a systematic approach to resolving these phenotypes, including reanalyses of the ES data with our innovative and agnostic approach in parallel to phenotyping and then utilizing the information from these iteratively. If the phenotype was specific enough to warrant targeted molecular tests, these were pursued and if still not resolved, WGS was utilized. Our study provides an integrated genomic-phenomic approach to resolving ES negative individuals that extends well beyond just ES reanalyses.

## MATERIALS AND METHODS

The study was performed under protocols approved by the Institutional Review Boards of Duke University Medical Center and the NHGRI.

### Demographics

Thirty-eight individuals with a pre-UDN negative ES evaluated at the Duke/Columbia UDN clinical site from September 2015-October 2017 were included. Nineteen patients (50%) were male, 29 (77%) were Caucasians, with two (5%) African-Americans, five (13%) Asians, two (5%) Others and six (16%) were Hispanic. The mean age was 7.07±5.82 years, ranging from 0–26 years. The mean age of onset of illness was 0.51±1.04 years and the mean duration of illness was 6.35±5.59 years. The organ system most often involved was the nervous system (58%) with the musculoskeletal and gastrointestinal systems being the next most frequent at 7% each (Table S1). The time to diagnosis was 5.76±5.22 months (0–23 months) in those that obtained a diagnosis (n=21), compared to 10.83±5.2 months, for declaration of no diagnosis in 12 individuals (t=2.21, p<0.05, Figure S1).

### Details of prior ES

Pre-UDN ES had been performed in 37 individuals and a pre-UDN WGS in one. A negative ES/WGS was operationalized as an ES/WGS report that was non-diagnostic and had either: (a) no variants of interest, (b) variants of uncertain significance (VUS) in a known disease-causing gene or (c) variants in candidate genes/genes of uncertain significance not associated with human disease. These pre-UDN variants are in Table S2.

Commercial ES had been performed in 22 individuals (59%), research ES in 13 (35%) and two (6%) had undergone both clinical and research ES (Individual 23 had clinical WGS). The majority were trios (33/38, 86%), two were quartets, two were duos (parent-child) and one was a singleton. The pre-UDN sequencing had occurred from 2012–2016, with the majority (36/38, 95%) occurring after 2012. One ES reanalysis prior to UDN entry had occurred in 18/38 (48%) individuals, 2.11±1.07 years (1–4 years) after the initial ES.

### Process for evaluation of ES negative individuals

We began with simultaneous ES data reanalyses and phenotyping. These data were iteratively used to derive variants of interest that could be pursued further for diagnoses. If the reanalyses found variants that were likely pathogenic, then the phenotyping was customized to capture clinical manifestations related to that particular disorder. If the phenotyping suggested specific conditions in the differential diagnosis, the ES data were reexamined for pertinent genes. Then, if no variants were detected and the clinical suspicion

for a particular disorder was high, direct Sanger sequencing/deletion-duplication testing and/or biochemical testing was pursued. WGS was utilized when these procedures did not result in resolution.

### UDN Phenotyping

Thirty-seven individuals underwent phenotyping (Individual 23 died after acceptance and prior to evaluation). This included customized clinical consultations, imaging, procedures and laboratory tests, during a one-week visit to the Duke campus. Clinical consultations were the most often obtained (median= 3), with radiological, laboratory tests and procedures being performed as needed (median=1).

### Review of other prior pertinent results

A chromosomal microarray that was at least at the level of an oligonucleotide array was available on 36/38 individuals and reviewed; no CNVs that could explain the individuals' features were evident. Regions of homozygosity on the array, if present, were utilized to identify autosomal recessive genes of interest. Other pre-UDN laboratory test results were reviewed, but details are beyond the scope of this publication.

### ES Reanalyses

FASTQ files were obtained directly, or generated with data from the pertinent laboratory in 35/38 individuals. In three individuals (24, 27 and 23) raw ES/WGS data could not be obtained. Primary alignment was performed with the DRAGEN platform. [21] Duplicate removal was performed using Picard tools and index realignment and variant calling conducted with GATK v3.6. Variants were annotated using Clin-Eff with Ensembl-GRCh37.73. Our bioinformatics is agnostic in its approach, utilizing the innovative tools developed by our group. The Residual Variation Intolerance Score (RVIS) assesses whether genes have accumulated common functional variation; subRVIS applies the RVIS approach to sub-regions of genes and captures regional changes due to isoform inclusion/exclusion of exons, and/or by gene domain. [22,23] Novel genotypes were filtered into tier one and tier two variants. Tier one variants were strictly filtered for quality and control observations in public databases (ExAC, gnomAD [24] and EVS [25]), and 13,000 internal controls. Tier one variants were further prioritized: hotzone variants (polyphen 2 score > 0.95 in an intolerant gene with an RVIS or a sub-RVIS score <25) were predicted damaging in an intolerant gene. We highlighted hotzone variants in known OMIM genes, or mouse essential genes. We also highlighted loss of function (LoF) variants that are in genes with known pathogenic LoF variants or reported as haploinsufficient by ClinGen [26], or LoF intolerant by high pLI score and estimated conservation/constraint of a variant site with the Genomic Evolutionary Rate Profiling (GERP) score. [27] We curated ClinVar, HGMD, and internal cases to annotate all variants previously reported pathogenic. Tier 2 variants had less strict filters for quality and control observations, but required that a variant is a known or expected pathogenic variant. This allowed pathogenic variants that might otherwise be filtered due to noise in the control data sets. *De novo*, newly homozygous, newly hemizygous, and compound heterozygous variants were identified. All coding and intron/exon boundary (up to 8bp) variants were also considered. An inheritance naïve filter was also applied to identify any variants which may be incompletely penetrant or mosaic in the parent. For genes known to be disease-

associated, we also used the ACMG criteria for variant classification (Table S1). In select cases, CNV analysis was performed with the target coverage and segmentation tools in GATK 4. These rely on normal samples sequenced on the same sequencing platform. With reanalysis these controls were not always available.

### Whole exome sequencing

Two individuals (15 and 22, Table S1) had a repeat ES since the prior trio ES had been performed in early 2012, when ES capture kits were more incomplete. These were performed at the Baylor Miraca sequencing core of the UDN, using methodology and analyses previously published [28,29].

### Whole genome sequencing

UDN WGS was performed by the HudsonAlpha UDN sequencing core on 27 individuals (26 trios, one quartet), with methodology and analyses as previously published. [30] The 27 individuals included 17 whose ES reanalyses through our study was negative and Individual 23 who had a pre-UDN negative WGS, as well as nine other individuals whose WGS was done in parallel with the ES reanalyses (Figure 1 and Table S1).

### Communication with laboratories regarding ES negative results

When new variants were detected, we corresponded with the pertinent laboratory to discuss the reasons for the variant not being detected or not prioritized previously. This information is in the relevant tables (Tables 1–4 and Table S1).

### Determination of Diagnoses

The genomic and clinical information was combined for diagnostic interpretation by consensus. The UDN has created categories of diagnoses, recognizing that it is difficult to determine the certainty of diagnosis in rare phenotypes and that the certainty may change over time. Of the four categories of Certain, Highly Likely, Tentative and Low, we used the first three to classify the diagnoses in the ES negatives in this study. Further considerations in this rubric are the method used to achieve the diagnosis (e.g. genomic sequencing, directed testing based on phenotype or clinical grounds), the mechanistic characterization of disease pathology, the degree to which the diagnosis explains the phenotypes of the patients and consequences of the diagnoses. Whenever pertinent, variants were confirmed by Sanger sequencing/MLPA/exon array, prior to communication to the individuals and their families. When bioinformatically compelling variants in novel genes were identified, these were categorized as candidate genes. If further avenues such as GeneMatcher and functional studies led to the determination that they were new disease genes, they were then classified as a diagnosis (Certain, Highly Likely or Tentative, depending on the strength of the supporting evidence).

## RESULTS

Overall, 18/38 (47%) individuals received Certain (n=12) or Highly Likely diagnoses (n=6) and three (8%) received Tentative diagnoses. Candidate genes were identified in five (13%) individuals. In total, we identified diagnoses/potential leads in 26/38 (68%) individuals

(Table 1). In the individuals with a Certain or Likely Diagnoses (excluding the two with clinical diagnoses only), eight had *de novo* autosomal dominant variants, six had biallelic autosomal recessive variants, one had an inherited autosomal dominant variant and one had a *de novo* X-linked dominant variant (Tables 2, 3 and S1).

### ES Reanalyses

In 8/35 (23%) individuals, a Certain or Highly Likely diagnosis was made and a ninth individual received a Tentative diagnosis after ES reanalyses. (Table 2, Figure 1). The reasons for a pre-UDN negative ES in these individuals are listed in Tables 2, 4 and S1. Other variants detected in our ES reanalyses are in Tables S2 and S3. Overall, in these nine individuals, in *one* instance there was an interim literature report of a new gene-disease association (*EFL1*, Table 2). Except for the homozygous CNV in *NPHP1* (Individual 3, Table 2) which may have been easier to detect by WGS, all the variants were of the type that are tractable by ES.

### Integration of Phenotype with Genomic Data from ES Reanalyses

Genomic findings directed the phenotyping and the phenotypic information led to the examination of specific genes. For example, for Individual 1 (Tables 2 and S1), the *EFL1* gene variant was detected just as phenotyping was beginning; further evaluations resulted in finding hematological, hepatic and pancreatic abnormalities consistent with the Shwachman-Diamond (SDS)-like syndrome associated with *EFL1*.[31] Conversely, phenotyping by the epileptologist led to a recommendation to examine the *CACNA1A* gene in Individual 2 (Tables 2 and S1) and a likely pathogenic variant was detected on manual inspection of the gene and also on reanalyses through the pipeline. There were no significant differences in whether the pre-UDN ES was clinical or research based, among the Certain and Highly Likely diagnoses ($\chi^2$=.46, p>0.05).

### Phenotype Guided Diagnoses

A strong clinical suspicion of specific disorders occurred in two individuals. Targeted molecular testing then led to pathogenic variants that had been missed on the pre-UDN ES, as published previously[11] (Table 3, Figure 1). A VUS in a known disease-causing gene in Individual 17 was reinterpreted since she and her mother (who also has this variant) have pathognomonic features of the *HEPACAM* related disorder on further phenotyping. A CNV was established as being diagnostic for some features in individual 18, based on an interim literature report of this CNV being associated with features that overlapped his.[32] In individuals 15 and 19 clinical diagnoses were conferred according to the UDN diagnostic rubric, due to their clinical features being so exactly consistent with a specific disorder, that the lack of molecular confirmation after all testing did not take away the diagnoses (Tables 3 and S1).

### Updated ES

Individual 14 (Tables 3 and S1) had a repeat trio ES and a candidate gene variant in *ASXL2* was proven to be associated with a new neurodevelopmental disorder.[17] Individual 22 had a negative repeat ES and was subsequently diagnosed on WGS due to a structural variant.

### Whole Genome Sequencing (WGS)

Three Certain/Highly Likely diagnoses were obtained in 3/18 (16%) individuals who underwent WGS, after all other modalities to achieve a diagnosis had failed. All three diagnoses were due to structural variants that had not been detected on pre-UDN ES, due to the difficulty in detecting indels larger than 15 bp with ES. [12,33] Two others obtained a Tentative diagnosis and one candidate gene was identified on WGS (Tables 1, 3 and S1). Interestingly, WGS was also pursued in nine individuals (Figure 1, Table S1) whose ES reanalyses were in progress. In all nine individuals, it was the ES reanalyses that led to either a diagnosis or a candidate gene, with the WGS not prioritizing these variants (Reasons in Tables 1–4).

### New Gene-Disease Associations

Two new gene-disease associations were established (*ASXL2* and *NACC1*) [17,18] after initial identification as candidate genes. For two other genes identified as candidates we have evidence through further functional studies, animal modeling and networking through GeneMatcher [34] to judge these as disease-associated (*AGTPBP1, IRF2BPL,* publications in progress) (Tables 2 and S1).

### Candidate Genes

Our ES reanalyses identified four new candidate genes. WGS did identify a fifth candidate gene, *TBX2*, and all are being studied currently (Figure 1, Tables 3 and S1).

### Secondary and Incidental Findings

Two individuals were found on WGS to have incidental findings. The father of Individual 24 was homozygous for the common pathogenic variant in the *HFE* gene for hemochromatosis and Individual 15 had a pathogenic variant in a long QT syndrome gene *KCNE1* (Table S1). These were communicated to the families with management recommendations and genetic counseling.

### Phenotypes of the ES negatives who remain Undiagnosed

Twelve of the 38 individuals remain without a diagnosis or candidate genes. There were no significant demographic differences between these individuals and the others. Their manifestations were less often within the nervous system (41%) compared to 65% in the 26 individuals wherein a diagnosis or a potential lead was available; although this difference was not significant (Fisher's exact test p>0.05), we also observed that many of the 12 individuals had phenotypes that were representative of complex disorders (Table S1).

## DISCUSSION

A systematic approach to resolving diagnoses in ES negative individuals is a critical need, as the genomics community is increasingly utilizing ES in routine clinical practice and yet 50–75% of individuals remain without a diagnosis. We demonstrate that careful consideration of the phenotypic features, combined with innovative agnostic bioinformatics ES reanalysis, targeted molecular testing and subsequent WGS results in a significant number of the ES

negatives being resolved (47%), with an additional ~20% obtaining tentative diagnoses or candidate genes. Our experience is that WGS is highly effective in detecting structural variants, making it an important adjunct approach to ES negatives. However, mining ES data to maximize its potential and utilizing phenotype directed targeted testing can detect/ prioritize variants not reported (due to analytical factors, knowledge gaps and variability in laboratory reporting), so that the more expensive option of WGS may be minimized (> 80% of the molecular diagnoses we made were made without WGS).

Prior studies on ES negatives have employed various approaches, including reanalyses of the raw ES data (sometimes with more relaxed filters), moving from singleton to trio sequencing to detect *de novos* and compound heterozygous variants, utilizing CNV analyses, considering the interim literature for new gene-disease associations and networking to identify additional patients; these procedures have yielded 10–36% additional diagnoses. [5–8,35] Our cohort was particularly challenging, since the majority had been sequenced as trios (88%), with almost all having a negative pre-UDN CNV analyses (94%) and a substantial number (48%) entering the study with one negative ES reanalysis. Thus, many logical next steps were not avenues that we could pursue. Despite this, our systematic and comprehensive approach resulted in ~ 70% of the individuals obtaining diagnoses or potential leads that could be pursued further. Our ES reanalyses alone were highly effective in providing diagnostic resolution in approximately 25% of the ES negative individuals in this study. Only two diagnoses were facilitated by new disease gene reports in the interim literature and all diagnoses were achieved without the relatively easy step of moving from a singleton to a trio. Utilization of networking platforms such as GeneMatcher or Matchmaker Exchange [34,35] did facilitate candidate gene follow-up. [17,18]

Our approach enabled us to identify variants that had been not been previously reported. Due to our innovative bioinformatics tools such as RVIS and our ranking of variants into tiers we were able to overcome analytical factors to select bioinformatically compelling variants. Capturing phenotypic changes allowed us to bridge knowledge gaps, resulting in identification of significant variants. Indeed analytical factors and knowledge gaps were the major reasons (70%) for a pre-UDN negative ES (Table 4). Variability in laboratory reporting resulted in non-reporting of significant variants when they did not fit the reported phenotype; this has implications for clinical practice as diagnoses can be missed and phenotypic expansion of a disorder may go unrecognized.

An important component of our systematic approach is to phenotype the ES negative individuals in parallel with the ES reanalyses. Phenotyping is also critical in solving ES negatives without automatically resorting to WGS. When the clinical phenotypes are specific enough to be suggestive of one or a few disorders, targeted molecular testing, such as Sanger sequencing can be effective in determining variants that can be difficult to detect on ES, and is also cost-effective [11]. Finally, we were able to confer clinical diagnoses in two individuals, even in the absence of molecular conformation, since unmistakable pathognomonic phenotypic features of a specific disorder were present. Such clinical diagnoses, when prudently made with irrefutable findings, provide a guide to the families and enable reasonable genetic counseling and estimates of reproductive risk, even as efforts to find a molecular basis continue.

The role of WGS in current diagnostics of rare and undiagnosed phenotypes is still being determined. In a cohort of individuals with intellectual disabilities and a negative ES, WGS led to diagnosis in ~40% due to detection of *de novo* and structural variants in the exome [12]; in such earlier studies, limitations of older ES capture kits may have led to coding variants not being detected. Other publications have estimated that 15–17% of additional diagnoses variants are made on WGS, mostly due to detection of variants not amenable to ES. [13,14] Interestingly, in our cohort, WGS led to a similar rate of Certain and Likely diagnoses (16%) in the ES negatives and all were due to structural variants that would be not be easily amenable to ES or chromosomal microarrays. We acknowledge that the majority of variants that we detected in this study would have been amenable to WGS, but several patients (n=9) who underwent WGS were ultimately resolved by our systematic approach and not by WGS. Varying reasons are operative for the negative WGS in these individuals, such as the UDN WGS laboratory not reporting variants that do not fit the describe phenotype well and not reporting variants in genes of uncertain significance. This further emphasizes the value of using different pipelines in reinterpreting raw data on ES negatives. Establishing collaborations with researchers at their institutions or outside may enable clinicians to utilize a different bioinformatics pipeline for the reanalyses of ES data.

Twelve individuals in our study have no diagnosis or candidate genes, despite all efforts. A few of these individuals have phenotypes wherein the etiology could be complex (e.g. inflammatory bowel disease, autism and recurrent fevers) and we propose that such disorders are difficult to solve by sequencing, unless larger cohorts with similar manifestations are accumulated and studied.

In conclusion, a systematic and comprehensive iterative approach to ES negatives that includes ES reanalysis, careful phenotyping, targeted testing and in select cases WGS, can result in a high rate of resolution. We recommend that with the high cost, relatively low incremental yield over ES, and complexity of analyses, that WGS be utilized only after ES data have been extensively mined and combined with the phenotypic data to maximize its yield. Many aspects of our approach can be implemented in practice. Commercial laboratories could adopt an agnostic approach (which could be easily automatable) to the raw data in ES negatives, so that variants that may be otherwise be filtered out due to stringent settings or phenotypic mismatch would be detected. Clinicians can also update laboratories about interim changes or atypical aspects of the phenotypes and ask about bioinformatically compelling variants that may have been initially unreported due to phenotypic mismatch. Utilizing targeted testing such as Sanger sequencing for disorders that are high in the differential diagnosis is useful; these variants may have been missed on ES, due to various analytical factors, as illustrated by Individuals 16 and 20 in this study. Additionally, considering disorders that are not amenable to ES (e.g. epigenetic disorders) and obtaining a chromosomal microarray (if not previously done), are useful approaches to ES negatives. Finding additional cases through networking such as GeneMatcher are also feasible in clinical practice. As the genomics community faces the challenge of the ES negatives, approaches such as ours provide viable avenues to maximize their resolution.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## References

1. Lee H, Deignan JL, Dorrani N, et al. Clinical exome sequencing for genetic identification of rare Mendelian disorders. JAMA. 2014; 312(18):1880–1887. [PubMed: 25326637]

2. Gilissen C, Hoischen A, Brunner HG, Veltman JA. Unlocking Mendelian disease using exome sequencing. Genome Biol. 2011; 12(9):228. [PubMed: 21920049]

3. Need AC, Shashi V, Hitomi Y, et al. Clinical application of exome sequencing in undiagnosed genetic conditions. J Med Genet. 2012; 49(6):353–361. [PubMed: 22581936]

4. Yang Y, Muzny DM, Xia F, et al. Molecular findings among patients referred for clinical whole-exome sequencing. JAMA. 2014; 312(18):1870–1879. [PubMed: 25326635]

5. Wenger AM, Guturu H, Bernstein JA, Bejerano G. Systematic reanalysis of clinical exome data yields additional diagnoses: implications for providers. Genet Med. 2017; 19(2):209–214. [PubMed: 27441994]

6. Smith ED, Radtke K, Rossi M, et al. Classification of Genes: Standardized Clinical Validity Assessment of Gene-Disease Associations Aids Diagnostic Exome Analysis and Reclassifications. Hum Mutat. 2017; 38(5):600–608. [PubMed: 28106320]

7. Eldomery MK, Coban-Akdemir Z, Harel T, et al. Lessons learned from additional research analyses of unsolved clinical exome cases. Genome medicine. 2017; 9(1):26. [PubMed: 28327206]

8. Nambot S, Thevenon J, Kuentz P, et al. Clinical whole-exome sequencing for the diagnosis of rare disorders with congenital anomalies and/or intellectual disability: substantial interest of prospective annual reanalysis. Genet Med. 2017

9. Gibson KM, Nesbitt A, Cao K, et al. Novel findings with reassessment of exome data: implications for validation testing and interpretation of genomic data. Genetics In Medicine. 2017

10. Need AC, Shashi V, Schoch K, Petrovski S, Goldstein DB. The importance of dynamic re-analysis in diagnostic whole exome sequencing. J Med Genet. 2017; 54(3):155–156. [PubMed: 27899421]

11. Pena LDM, Jiang YH, Schoch K, et al. Looking beyond the exome: a phenotype-first approach to molecular diagnostic resolution in rare and undiagnosed diseases. Genet Med. 2017

12. Gilissen C, Hehir-Kwa JY, Thung DT, et al. Genome sequencing identifies major causes of severe intellectual disability. Nature. 2014; 511(7509):344–347. [PubMed: 24896178]

13. Taylor JC, Martin HC, Lise S, et al. Factors influencing success of clinical genome sequencing across a broad spectrum of disorders. Nat Genet. 2015; 47(7):717–726. [PubMed: 25985138]

14. Lionel AC, Costain G, Monfared N, 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. 2017

15. Sims D, Sudbery I, Ilott NE, Heger A, Ponting CP. Sequencing depth and coverage: key considerations in genomic analyses. Nature reviews Genetics. 2014; 15(2):121–132.

16. Fang H, Wu Y, Narzisi G, et al. Reducing INDEL calling errors in whole genome and exome sequencing data. Genome medicine. 2014; 6(10):89. [PubMed: 25426171]

17. Shashi V, Pena LD, Kim K, et al. De Novo Truncating Variants in ASXL2 Are Associated with a Unique and Recognizable Clinical Phenotype. Am J Hum Genet. 2016; 99(4):991–999. [PubMed: 27693232]

18. Schoch K, Meng L, Szelinger S, 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. 2017; 100(2):343–351. [PubMed: 28132692]

19. Pabinger S, Dander A, Fischer M, et al. A survey of tools for variant analysis of next-generation genome sequencing data. Briefings in bioinformatics. 2014; 15(2):256–278. [PubMed: 23341494]

20. Rare Disease Statistics. 2015. https://globalgenes.org

21. genome E. http://edicogenome.com/dragen-bioit-platform/

22. Petrovski S, Wang Q, Heinzen EL, Allen AS, Goldstein DB. Genic intolerance to functional variation and the interpretation of personal genomes. PLoS Genet. 2013; 9(8):e1003709. [PubMed: 23990802]

23. Gussow AB, Petrovski S, Wang Q, Allen AS, Goldstein DB. The intolerance to functional genetic variation of protein domains predicts the localization of pathogenic mutations within genes. Genome Biol. 2016; 17:9. [PubMed: 26781712]

24. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016; 536:285–291. [PubMed: 27535533]

25. (ESP) NGESP. Exome Variant Server. http://evs.gs.washington.edu/EVS/

26. Resource CG. https://www.clinicalgenome.org/

27. Cooper GM, Stone EA, Asimenos G, Green ED, Batzoglou S, Sidow A. Distribution and intensity of constraint in mammalian genomic sequence. Genome Res. 2005; 15(7):901–913. [PubMed: 15965027]

28. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015; 17(5):405–424. [PubMed: 25741868]

29. Yang Y, Muzny DM, Reid JG, et al. Clinical whole-exome sequencing for the diagnosis of mendelian disorders. N Engl J Med. 2013; 369(16):1502–1511. [PubMed: 24088041]

30. Bick D, Fraser PC, Gutzeit MF, et al. Successful Application of Whole Genome Sequencing in a Medical Genetics Clinic. Journal of pediatric genetics. 2017; 6(2):61–76. [PubMed: 28496993]

31. Stepensky P, Chacon-Flores M, Kim KH, et al. Mutations in EFL1, an SBDS partner, are associated with infantile pancytopenia, exocrine pancreatic insufficiency and skeletal anomalies in aShwachman-Diamond like syndrome. J Med Genet. 2017; 54(8):558–566. [PubMed: 28331068]

32. Vlaskamp DR, Rump P, Callenbach PM, et al. Haploinsufficiency of the STX1B gene is associated with myoclonic astatic epilepsy. Eur J Paediatr Neurol. 2016; 20(3):489–492. [PubMed: 26818399]

33. Caspar SM, Dubacher N, Kopps AM, Meienberg J, Henggeler C, Matyas G. Clinical sequencing: from raw data to diagnosis with lifetime value. Clin Genet. 2017

34. Sobreira N, Schiettecatte F, Valle D, Hamosh A. GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. Hum Mutat. 2015; 36(10):928–930. [PubMed: 26220891]

35. Philippakis AA, Azzariti DR, Beltran S, et al. The Matchmaker Exchange: a platform for rare disease gene discovery. Hum Mutat. 2015; 36(10):915–921. [PubMed: 26295439]
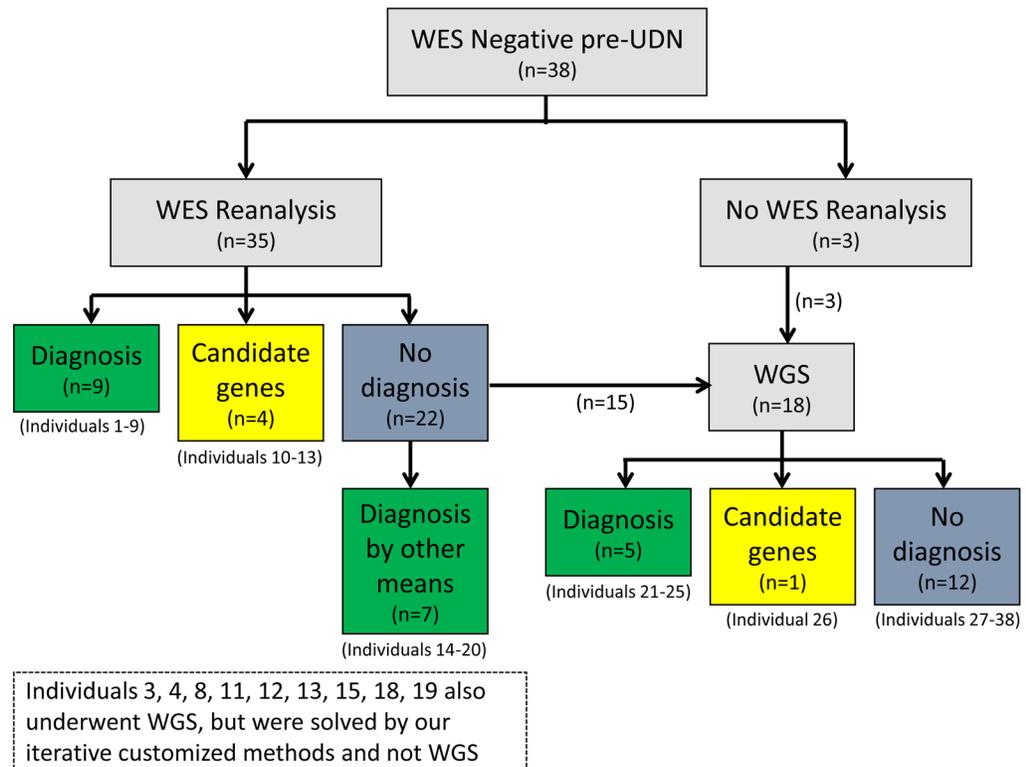
**Figure 1.**
Flowchart illustrating the approach to the ES negatives and the resolution with the different modalities

**Table 1**

Genes that were implicated in Certain, Highly Likely, Tentative Diagnoses and as Candidates in 26/38 Individuals

| Mode of Diagnosis | Certain Diagnoses n=12 | Highly Likely Diagnoses= 6 | Tentative Diagnoses=3 | Candidate Genes=5 |
|---|---|---|---|---|
| Genes detected on ES reanalyses | *AGTPBP1, CACNA1A, EFL1, NACC1, NPHP1* | *CACNA1C, IRF2BPL, MYBPC1* | *HNRNPK* | *CTBS, DROSHA, KRT19, RNF2* |
| Targeted Sanger Sequencing/MLPA | *ANTXR2, PLA2G6* | None | None | None |
| WGS | *HDAC8, MECP2* | *ITPA* | *CAD, SON* | *TBX2* |
| Clinical diagnosis | Oral-facial-digital syndrome, unspecified type | Multiple Pterygium syndrome | None | None |
| Other | Phenotype directed reinterpretation of ES: *HEPACAM* Repeat ES through UDN: *ASXL2* | Chromosomal microarray reinterpretation: 16p11.2 deletion | None | None |

**Table 2**

Details of the 9/36 Individuals that were Resolved by Bioinformatics Reanalyses of pre-UDN ES Data and Phenotyping

| | Individual number and Phenotype | Pre-UDN ES | Reason for Pre-UDN negative ES# | Duke/Columbia Bioinformatics Reanalyses and Annotation of Gene/Variants | Certainty of Diagnosis and Details |
|---|---|---|---|---|---|
| 1 | 14 year-old Caucasian female with growth failure, metaphyseal dysplasia and thrombocytopenia | Research | Variability in Laboratory Reporting (Prioritization) — Research lab identified variant, did not prioritize due to focus on *de novos* | *EFL1* c.379A>G p.T127A Tier 1 newly homozygous | Certain Shwachman-Diamond-like syndrome |
| 2 | 18 year-old Caucasian female with hemiplegic migraine, hypotonia, ataxia, cerebellar atrophy and severe intellectual disability | Research | Analytical Approach (Technical limitation) — Not detected by research laboratory due to unknown reasons | *CACNA1A* c.4055G>T p.R1352L Tier 1 *de novo* in HZ [E] and HZ [OMIM]. Variant previously in ClinVar | Certain Epileptic encephalopathy, early infantile, 42 (MIM#617106) |
| 3 | 14 year-old Caucasian male with renal failure due to nephronophthisis, retinal dystrophy and cerebellar ataxia | Research | Analytical Approach (Technical limitation of ES in CNV detection) — Homozygous deletion of NPHP1 gene not amenable to ES | *NPHP1** Homozygous deletion CNV calling with GATK4 on exome | Certain Nephronophthisis, familial juvenile (MIM#256100) |
| 4 | 8 month-old Caucasian male with congenital hypotonia, muscle weakness, fine tremor, laryngomalacia and motor delay | Commercial | Variability in Laboratory Reporting (Interpretation) — Not reported by clinical lab since phenotype not thought to be a good fit due to lack of arthrogryposis | *MYBPC1** c.776T>C p.L259P Tier 1 *de novo* in HZ [OMIM]. | Highly Likely Phenotypic expansion of Distal Arthrogryposis, Type 1B (MIM#614335); three other similar patients identified, all with congenital hypotonia, tremors that have improved over time, with normal cognition |
| 5 | 20 month-old Caucasian male with infantile spasms, microcephaly, lamellar cataracts, failure to thrive and global developmental delay | Commercial | Knowledge Gap (gene-disease relationship not established) — Reported as candidate gene with VUS | *NACC1*** c.892C>T p.R298W Tier 1 *de novo* in HZ [E] | Certain Neurodevelopmental disorder with epilepsy, cataracts, feeding difficulties and delayed brain myelination (MIM#617393) |
| 6 | 16 month-old Caucasian female with axonal motor neuron disease and cerebellar atrophy | Commercial | Knowledge Gap (gene-disease relationship not established) — Reported as candidate gene with VUS | *AGTPBP1**** c.2492-1G>T IVS16-1G>T c.2892delC p.Y964X LoF variants in LoF depleted gene | Certain Infantile-onset degeneration of central and peripheral nervous systems; nine other patients identified |
| 7 | 6 year-old Caucasian male with developmental delay, hypotonia, followed by progressive neurological regression | Research | Unknown reasons — Not reported for unknown reasons | *IRF2BPL**** c.514 G>T p.E172X Tier 1 *de novo* LoF in LoF depleted gene | Highly Likely *IRF2BPL* associated neurodegenerative disorder; six patients identified with similar phenotypes |

| Individual number and Phenotype | Pre-UDN ES | Reason for Pre-UDN negative ES[#] | Duke/Columbia Bioinformatics Reanalyses and Annotation of Gene/Variants | Certainty of Diagnosis and Details |
|---|---|---|---|---|
| 8 | 7 year-old Caucasian female with epilepsy, hypotonia, profound intellectual disability, cortical visual impairment Subsequently found to have prolonged QT interval | Commercial | Variability in Laboratory Reporting (Interpretation) | *CACNA1C** c.4087G>T p.V1363L Tier 1 *de novo* in HZ [E] and HZ [OMIM]. | Highly Likely Timothy syndrome (MIM#601005); diagnosis on patient supported by additional phenotyping with ECG and Holter monitoring. Three other similar patients identified who presented with epilepsy, intellectual disability, no syndactyly and on further investigation 2/3 had long QT interval, thus representing likely phenotypic expansion of disorder (electrographic investigations pending in third patient) |
| | Not reported by clinical lab since phenotype not thought to be a good fit, due to lack of syndactyly and electrocardiographic abnormalities | | | | |
| 9 | 10 year-old Caucasian female with developmental delay, microcephaly, hypotonia, minor dysmorphic features, conductive hearing loss | Research | Knowledge Gap (gene-disease relationship not established) | *HNRNPK** c.173T>C p.I58T Tier 1 *de novo* | Tentative Au-Kline syndrome (MIM#616580). Functional studies of variant pending |
| | Not reported due to lack of disease association | | | | |

[#] Reasons for negative WES classified into Analytical Approach differences, Knowledge Gap and Variability in Laboratory Reporting, with specific reasons under each category being provided whenever available

* WGS through the UDN also performed, but variant not reported

** Now established to be new disease associated gene

*** Candidate gene pursued with further clinical and functional studies, resulting in multiple affected patients and enough evidence to publish as new disease associated gene

HZ= Hot zone variant

E= Essential gene in mouse

**Table 3**

Diagnostic Resolution of ES Negatives with WGS and Other Modalities of Diagnosis

| Details of individuals resolved with WGS, after negative pre-UDN ES and negative ES reanalyses in UDN | | | | |
|---|---|---|---|---|
| **Individual Number and Phenotype** | **UDN WGS Findings** | **Certainty of Diagnoses and Details** | **Reasons for negative ES/Other pertinent information** | |
| 21 — 7 year-old Hispanic female with acquired microcephaly, speech and motor regression, autism, scoliosis and frequent infections | *de novo* 41 bp deletion in the *MECP2* gene c.1157_1197delTGCCCCC ACCTCCACCTGAGCCCGAGAG CTCCGAGGAGCCCC p.L386HfsTer5 | Certain Rett syndrome (MIM## 312750) | Analytical Approach (Technical limitation of ES in indel detection of >15 bp) | Pre-UDN ES and our reanalyses did not detect variant due to difficulties in indel detection with ES. Phenotyping consistent with a Rett-like syndrome. Manual inspection of ES data enabled visualization of the variant retrospectively |
| 22 — 11 year-old Caucasian female with multiple congenital anomalies, choanal atresia, bilateral sensorineural hearing loss, mild bilateral optic nerve hypoplasia, Klippel-Feil anomaly and global developmental delays | *de novo* 43kb deletion on the X-chromosome encompassing exons 1 and 2 of *HDAC8* and exons 22-32 of *PHKA1*. Confirmed by exon array | Certain Cornelia de Lange syndrome 5 (MIM# 300882) | Analytical Approach (Technical limitation of ES in CNV detection) | Prior CMA reported a 33 kb deletion, with only gene reported in the deletion being *PHKA1*, responsible for X-linked recessive glycogen storage disease type IX |
| 23 — 14 month-old female with refractory epilepsy/ neuromuscular disorder and a family history of the same condition in brother and sister. Parents consanguineous. The individual died at age 20 months with progressive neurological decline | 1.8 kb homozygous deletion of exon 5 of the *ITPA* gene | Highly Likely Epileptic encephalopathy, early infantile, 35 (MIM# 616647) | Analytical Approach (CNV not found on prior WGS due to unknown reasons) | Several regions of homozygosity noted on SNP CMA, encompassing >6.5% of the genome, including the *ITPA* gene. Affected brother had the same deletion on further testing and parents were confirmed to be carriers |
| 24 — 3 year-old Caucasian male with abnormal gait, developmental delay and hypotonia | compound heterozygous VUS in *CAD* c.6320C>G, p.P2107R c.4669C>G, p.L1557V | Tentative Epileptic encephalopathy, early, infantile, 50 (MIM#616457) | Variability in Laboratory Reporting (Interpretation) Not reported due to poor phenotypic fit, since he did not have epilepsy | Being phenotyped for evidence of congenital disorder of glycosylation, since the disorder is due to abnormal glycosylation |
| 25 — 3 year-old Caucasian male with multiple congenital anomalies, cerebellar hypoplasia, hypomyelination, leukomalacia and developmental delays | *de novo* inframe indel in *SON* c.5860_5880delAGCCGCCGCGCAGCCGCACCCCC p.S1992_R1998del | Tentative ZTTK syndrome (MIM#617140) | Analytical Approach (Variant calling of synonymous variants) | Seven amino acid deletion occurs in RS domain of gene, critical for SON function. Due to good phenotypic fit, further functional studies being pursued through collaboration |

**Details of individuals resolved with WGS, after negative pre-UDN ES and negative ES reanalyses in UDN**

| Individual Number and Phenotype | UDN WGS Findings | Certainty of Diagnoses and Details | Reasons for negative ES/Other pertinent information | |
|---|---|---|---|---|
| 14 | *8 year-old African American/Caucasian male with macrosomia, glabellar nevus flammeus, hypertelorism and learning difficulties | UDN ES *ASXL2* *de novo* c.2424delC p.P808fs LoF in LoF depleted gene | Certain Shashi-Pena syndrome (MIM#617190)Shashi-Pena syndrome (MIM#617190) | Knowledge Gap (gene-disease relationship not established) | Networking identified five other cases leading to new gene-disease association |

**Details of the seven cases that were diagnosed by modalities other than ES reanalyses and WGS**

| Individual Number and Phenotype | Modalities to Diagnosis | Certainty of Diagnoses and Details | Reasons for negative ES/Other pertinent information | |
|---|---|---|---|---|
| 15 | 5 year-old Caucasian female with congenital hypothalamic hamartoblastoma, intractable seizures, microcephaly, profound developmental delay | Clinical Targeted sequencing of all known oral-facial-digital (OFD) syndrome genes negative | Certain OFD syndrome Clinical Diagnosis | | UDN WGS negative Phenotyping showed pathognomonic oral, skeletal and radiological features all consistent with an OFD |
| 16 | *3 year-old Pakistani female with developmental regression, cerebellar atrophy Parents are first cousins | Sanger sequencing and MLPA of *PLA2G6* Homozygous 2431-bp deletion with 7-bp insertion (c.-545_-46+1931delinsCGATCTC) in the 5′UTR region | Certain Infantile neuroaxonal dystrophy (MIM#256600) | Analytical Approach (Difficult region of exome) Commercial ES capture kit did not contain probes for the non-coding exon 1 of *PLA2G6* | Changes in neurologic phenotype strongly suggestive of infantile neuroaxonal dystrophy leading to Sanger and MLPA |
| 17 | 6 year-old African American female with macrocephaly, learning difficulties, mild motor incoordination, white matter abnormalities | Radiology reinterpretation of brain MRI identified subcortical cysts and other changes, consistent with *HEPACAM* associated disorder c.592 G>A p.D198N | Certain Megalencephalic leukoencephalopathy with subcortical cysts 2B (MIM#613926) | Variability in Laboratory Reporting (Interpretation) due to incomplete phenotypic information Commercial ES reported variant as VUS in *HEPACAM* inherited from mother | Autosomal recessive and dominant phenotypes associated with *HEPACAM*. Child found to have subcortical cysts not previously detected and found to have classical features of the diagnosis. Mother phenotyped and found to be affected |
| 18 | 26 year-old Caucasian male with intellectual disability, optic nerve abnormalities, appendicular ataxia, speech impairment and history of seizures | Chromosomal microarray reinterpretation due to interim publication, explains part of the phenotype | Highly Likely 16p11.2 deletion syndrome | | The 16p deletion is proximal to the typical 16p11.2 deletion and an interim publication implicated haploinsufficiency of *STX1B* gene within the interval in intellectual disability and seizures |
| 19 | 9 year-old Asian/Caucasian female with pterygiae, cleft palate, poor muscle mass, camptodactyly, club feet, progressive scoliosis, | Clinical | Highly Likely Multiple pterygium syndrome Clinical Diagnosis | | Extensive phenotyping of individual and affected sister confirmed clinical, radiological and muscle biopsy changes consistent with multiple pterygium syndrome |

**Details of individuals resolved with WGS, after negative pre-UDN ES and negative ES reanalyses in UDN**

| Individual Number and Phenotype | UDN WGS Findings | Certainty of Diagnoses and Details | Reasons for negative ES/Other pertinent information |
|---|---|---|---|
| distinctive facial features, and similarly affected sister | | | |
| 20 *18 month-old Hispanic female with progressive joint contractures, gingival hypertrophy, nodules, perianal skin tags and protein losing enteropathy | Sanger sequencing *ANTXR2* homozygous insertion c.1073dupC p.Ala359Cysfs *13 Parents confirmed to be carriers | Certain Infantile Systemic Hyalinosis (MIM#228600) | Analytical Approach (Technical limitations) Commercial ES variant calling software did not identify variant that was adjacent to homopolymeric region |

Phenotyping pathognomonic for infantile systemic hyalinosis and thus Sanger sequencing of *ANTXR2* was performed

CNV= copy number variant, CMA= chromosomal microarray, SNP= Single nucleotide polymorphism

*
Individuals included in previous publications

**Table 4**

Reasons for Negative ES Results in the 23 Genes that were Determined to be Diagnostic (n=18) or a
Candidate Gene (n=5)

| Categories Related to a Negative ES | Subcategory | Reasons | Examples in our study (inclusive of diagnoses and candidate genes) |
|---|---|---|---|
| Analytical Approach (35%) | Difficult Regions of Exome | Variants not detected due to capture kit not containing probes resulting in missed data | *PLA2G6* |
| | Technical Limitations of ES | Variant calling software limitations (Indels, Structural variants and CNVs) | *ANTXR2, NPHP1, MECP2, HDAC8, ITPA* |
| | Variant Filtering/Calling | Stringent filtering, Synonymous variants | *TBX2, SON* |
| Knowledge Gap (35%) | None | Novel Candidate Genes with No Known Disease Association | *KRT19, CTBS, DROSHA, RNF2, HNRNPK AGTPBP1, NACC1, ASXL2* |
| Variability in Laboratory Reporting (22%) | Variants not Prioritized | Laboratory focused on *de novo* variants | *EFL1* |
| | Variant Interpretation | Poor phenotypic fit determination by laboratory | *MYBPC1, CACNA1C, HEPACAM, CAD* |
| Unknown Reasons (8%) | None | Reasons not available from pertinent laboratory | *CACNA1A, IRF2BPL* |