# Clinical utility of Whole Exome Sequencing for rare Mendelian disorders: phenotypic-driven strategy for a high diagnostic yield and identification of 48 novel variants

Nikolaos Marinakis[1], Danai Veltra[1], Maria Sviggou[1], Christalena Sofocleous[1], Kyriaki Kekou[1], Eirini Tsoutsou[1], Konstantina Kosma[1], and Jan Traeger-Synodinos[1]

[1]National and Kapodistrian University of Athens

November 1, 2020

## Abstract

About 6,000 to 7,000 different rare disorders with suspected genetic etiologies have been described and in almost 4,500 of them the causative gene(s) have been identified. The advent of Next-Generation Sequencing (NGS) technologies has revolutionized genomic research and diagnostics, representing a major advance in identification of pathogenic genetic variations. WES facilitates a faster and more cost-effective route for definite diagnosis of rare genetic disorders, minimizing previous "diagnostic odysseys" for the patients. Due to the limitation that WES is not reimbursed in Greece, we aimed to minimize cost per patient/family through applying WES in the proband, followed by targeted family segregation studies when necessary. Furthermore, for variant filtration and interpretation we applied a phenotypic-driven strategy in close collaboration with clinical geneticists or referring clinicians. In this study we report the clinical application of WES in the diagnosis of 162 cases referred to investigate patients with undiagnosed genetic disorders. The overall molecular diagnostic yield reached 52.5%. Our experience as an academic diagnostic laboratory using WES, although limited to the last two years, allowed characterization of 94 pathogenic variants in 85 positive cases, 48 of which were novel, contributing information to the community of disease and variant databases.

Marinakis Nikolaos ORCID iD: 0000-0002-7078-3253

**Clinical utility of Whole Exome Sequencing for rare Mendelian disorders: phenotypic-driven strategy for a high diagnostic yield and identification of 48 novel variants**

**Marinakis Nikolaos[1], Veltra Danai[1], Sviggou Maria[1], Sofocleous Christalena[1,2], Kekou Kyriaki[1], Tsoutsou Eirini[1], Kosma Konstantina[1], Traeger-Synodinos Jan[1]**

[1]Department of Medical Genetics, St. Sophia's Children's Hospital, National and Kapodistrian University of Athens, Athens, Greece

[2] Research University Institute for the Study of Genetic and Malignant Disorders in Childhood, St. Sophia's Children's Hospital, National and Kapodistrian University of Athens, Athens, Greece

**Correspondence:**

Marinakis Nikolaos, M.Sc.

Laboratory of Medical Genetics

School of Medicine

National and Kapodistrian University of Athens

St. Sophia's Children's Hospital

nikomari@med.uoa.gr

**Running title :**

Whole Exome Sequencing for a high diagnostic yield and identification of 48 novel variants

**Abstract**

About 6,000 to 7,000 different rare disorders with suspected genetic etiologies have been described and in almost 4,500 of them the causative gene(s) have been identified. The advent of Next-Generation Sequencing (NGS) technologies has revolutionized genomic research and diagnostics, representing a major advance in identification of pathogenic genetic variations. WES facilitates a faster and more cost-effective route for definite diagnosis of rare genetic disorders, minimizing previous "diagnostic odysseys" for the patients. Due to the limitation that WES is not reimbursed in Greece, we aimed to minimize cost per patient/family through applying WES in the proband, followed by targeted family segregation studies when necessary. Furthermore, for variant filtration and interpretation we applied a phenotypic-driven strategy in close collaboration with clinical geneticists or referring clinicians. In this study we report the clinical application of WES in the diagnosis of 162 cases referred to investigate patients with undiagnosed genetic disorders. The overall molecular diagnostic yield reached 52.5%. Our experience as an academic diagnostic laboratory using WES, although limited to the last two years, allowed characterization of 94 pathogenic variants in 85 positive cases, 48 of which were novel, contributing information to the community of disease and variant databases.

**Keywords:**

Whole Exome Sequencing, diagnostic yield, rare Mendelian disorders, phenotypic-driven strategy, novel variants

*Introduction*

Rare diseases (RDs) affect approximately 6-8% of the population worldwide (Aymé, 2012; Baird et al., 1988). To date, about 6,000 to 7,000 different rare diseases with suspected genetic etiologies have been described (Sawyer et al., 2016) and in almost 4,500 of them the causative gene(s) have been identified (OMIM April 2020 updated, Amberger et al., 2019). The advent of Next-Generation Sequencing (NGS) technologies has revolutionized genomic research and diagnostics, representing a major advance in the identification of pathogenic genetic variations (Baynam et al., 2016; Hartman et al., 2019; Lim et al., 2015; Rabbani et al., 2012; Smith et al., 2019; J. Taylor et al., 2019).

Whole Exome Sequencing (WES) involves the analysis of protein-coding regions of the total human genome (representing <2%), which is currently believed to include the majority of causative variants in monogenic diseases (Adams & Eng, 2018). WES application for clinical diagnostics has completely changed the landscape of medical genetics. Compared to conventional testing strategies, WES facilitates a faster and more cost-effective route for definite diagnosis of rare genetic diseases, minimizing previous "diagnostic odysseys" experienced by many patients and their families (Bertier et al., 2016; Iglesias et al., 2014; Niguidula et al., 2018; Yang et al., 2013). Several studies have showcased a high WES diagnostic rate, ranging from 25% to 40%, (Jalkh et al., 2019; Reuter et al., 2019; Sawyer et al., 2016; Seaby et al., 2016; Valencia et al., 2015). This is likely influenced by a variety of factors including the strategy of analysis (proband alone versus trio of proband and parents), family history, type of disorder, age of onset, and acuity of the clinical evaluation (Taylor et al., 2015). Correct interpretation of WES data presents a challenge, entailing filtration and classification of ˜30.000 variants distributed across the coding genome of every case.

When it comes to genetic services, the diagnostic yield and clinical utility of WES are highly dependent on the close collaboration of a multidisciplinary team, including clinical geneticists (or referring clinicians) and laboratory geneticists. In our experience, definitive diagnosis was reached in more than half of the patients referred for WES, highlighting the importance of applying stringent testing and bioinformatic protocols, complimented by a robust phenotypic-driven strategy for variant evaluation.

*Material and Methods*

*Patient cohort*

During a two-year period (2018-2019), 162 patients with a clinically suspected genetic condition were referred to our laboratory for WES. The cohort comprised of 57 children <5 years old (35.2%), 74 adolescents (5-18 years) (45.7%) and 31 adults (19.1%). Males (n=83) and females (n=79) were equally represented and within all age groups. The primary reasons for referral, included neurodevelopmental abnormalities (developmental delay, intellectual disability, growth deficiency etc.) and concerned 47 cases (29%). Twenty-nine cases (18%) were referred for skeletal & connective tissue abnormalities, and 24 cases (15%) for congenital anomalies/syndromic features (Figure 1). The remaining 62 cases were referred for various reasons as presented in Figure 1 where "others" comprise of patients with endocrine, immune, sex or hematological abnormalities and a germline tumor case. Nine (out of 162) patients had a negative result with array-CGH analysis prior to WES.

In Greece, the cost of WES is not reimbursed by the Greek National Health System or private insurance companies. To minimize the expense for the patients and their families, WES was offered only for the probands, followed by targeted Sanger sequencing of parental/familial samples as and when appropriate for the evaluation of family segregation for likely pathogenic and causative variants.

Written informed consent prior to all tests were provided by the patients and/or their parents/legal guardians after extensive clinical evaluation and counseling by relevant medical specialists (Clinical Geneticists, Pediatricians, Neurologists, Cardiologists and others). All samples were given a unique patient code and all data was processed and stored according to the guidelines of the General Data Protection Regulation (GDPR).

*Whole Exome Sequencing*

Genomic DNA was extracted from peripheral blood samples according to standard methods. WES was performed using the Whole Exome Solution kit by Sophia Genetics, SA, following the manufacturer's recommendations. The resulting libraries were subjected to paired-end sequencing on an Illumina NextSeq 500 platform. Library preparation and sequencing were outsourced (Genotypos-Science Labs, Athens). Raw data were converted to FastQ files for computational analysis which included read alignment to the human genome reference sequence (GRCh37/hg19), duplicate marking, base quality score recalibration, indel realignment and variant calling, using the *SOPHiA DDM®* platform (Sophia Genetics, SA). Variant annotation with ANNOVAR algorithm (K. Wang et al., 2010) and variant filtration process were performed using VarAFT v2.16 (http://varaft.eu) application (Desvignes et al., 2018).

Variant filtering involved a six-step phenotypic-driven strategy as presented in Figure 2 and summarized as follows: (1) Variant filtering by phenotype and/or in silico gene list; (2) Variant filtering by population frequency (Minor Allele Frequency, MAF [?]1%); (3) Variant filtering by variant type and/or ClinVar pathogenicity and/or in-silico pathogenic prediction tools; (4) Variant filtering by genotype; (5) Variant filtering by American College of Medical Genetics and Genomics (ACMG) classification; (6) Clinical/phenotypic (re)evaluation of retained variant(s).

The accepted WES data quality control parameters included a mean depth of coverage >50x, with >98% regions at 25x. Compared to the human genome reference hg19, ~30.000 variants were identified per exome including only exonic and flanking (+/- 10bp) regions. Subsequent to variant filtration using criteria outlined in Fig. 1, distinct variants were selected and set for further phenotype/genotype assessment and evaluation by a multidisciplinary team including the referring clinician

*Sanger confirmation, family segregation and reporting strategy*

Neither Sanger confirmation nor family segregation was required for all cases. Verification was deemed necessary for variant(s) with borderline quality parameters (e.g. coverage) to preclude a false positive result and short INDELS (over 5 bp) to define their exact size and position. Variants(s) with clear pathogenicity, such as nonsense, frameshift or canonical splice site variants and variant(s) previously reported as pathogenic

3

in similar cases required no family segregation studies, except in cases where trans or cis configuration of variants was important for the final resolution of genotypes.

Further analysis was also necessary for assessing VUS and determine *de novo* or inherited nature of variants, parental origin and final classification to likely pathogenic or benign. In the presence of one pathogenic/likely pathogenic variant associated with recessive inheritance exonic deletions/duplications along the candidate gene were investigated using Multiplex Ligation-dependent Probe Amplification (MLPA) or other techniques used for Copy Number Variant (CNVs) detection (e.g. SNP arrays). Recommendations for re-evaluation of WES data and clinical findings after 1-2 years were made to cases where no variant could explain the patient's phenotype and cases when parental samples were unavailable to resolve the VUS detected.

*Results*

## Patients Cohort and Diagnostic yield

A total of 85 out of 162 patients received a definitive genetic diagnosis, including 9 of 85 cases with a previously negative array-CGH, (Table 1), demonstrating an overall diagnostic yield of 52.5%. Specifically, 40 patients (24.6%) received a definitive diagnosis based on the identification of highly scored known variants in phenotype-relevant genes. In 62 cases where family segregation studies were performed, WES findings were confirmed and mode of inheritance supported a definitive diagnosis in 45 more patients. Two cases (1.2%) with VUS remained unresolved as parents refused DNA analysis. In the remaining 58 (35.9%) cases no candidate causative variant was identified. From amongst the 77 unresolved cases, 75 were advised to proceed to clinical and bioinformatics re-evaluation after 1-2 years (Bruel et al., 2019; Ewans et al., 2018; Jalkh et al., 2019).

The majority of patients resolved (58/85 68%) concerned an autosomal dominant disease, followed by 20 (24%) patients with an autosomal recessive disease and 7 (8%) with X-linked diseases. The highest diagnostic yield (83%) was achieved among the sub-cohort of patients referred with dermatological abnormalities (epidermolysis bullosa, ichthyosis etc.) and neuromuscular diseases (muscular dystrophies, myopathies etc.). Furthermore, >60% of cases with multiple congenital anomalies/syndromic (Noonan syndrome, Kabuki syndrome, Sotos syndrome etc.) and skeletal/connective tissue abnormalities (Marfan syndrome, Ehlers-Danlos syndrome, skeletal dysplasia etc), were resolved. The lowest diagnostic yield (22%) was observed in cases with cardio and/or vascular abnormalities (Fig. 3). In 77 (47.53%) of cases no relevant variant could be identified, even following variant filtering and analysis with all combinations described (Figure 2). In four patients with complex clinical presentations, additional variants in likely phenotype-relevant genes were identified alongside the primary disease-associated genotype, as summarized in Table 2 and described in Discussion.

## Characteristics of variants identified

According to ACMG classification (Richards et al., 2015) 95 pathogenic or likely pathogenic variants were detected in 85 patients (Table 1). Remarkably, 48 of the 95 (~50%) variants were not reported in any of the public databases and were submitted to the ClinVar database https://www.ncbi.nlm.nih.gov/clinvar/ (Landrum et al., 2014). The variants comprised 56 (59%) missense, 10 (11%) nonsense, 19 (20%) frameshift, 1 (1%) in-frame deletion, 8 (8%) splice site variants and 1 (1%) whole exon deletion.

*Discussion*

In this study we report the clinical application of WES in the diagnosis of 162 cases referred to a single academic genetics center. The overall molecular diagnostic yield reached 52.5%, slightly higher of that reported in many previous studies (Jalkh et al., 2019; Sawyer et al., 2016; Seaby et al., 2016; Valencia et al., 2015). This rate was observed to vary depending on the disease category (Figure 3), but is also likely influenced by the strategy applied. In our department, the strategy included: (1) Analytical clinical description of the patient's phenotype along with comprehensive family and medical history to include as many data; (2) Stringent phenotype-driven strategies adopted for variant filtration and interpretation (Figure 2); (3) Close collaborations between clinicians and laboratory geneticists for variant evaluation

and phenotype-genotype correlation; (4) Family segregation strategies with Sanger sequencing to support evaluation of variant pathogenicity when necessary. Besides potentially optimizing the diagnostic yield, the strategy supported a reduction of the number of variants in need of classification and an acceptable turn-around time and reduced the overall cost to the patient, paramount in the absence of NGS-reimbursement in Greece. Every attempt was made to use new and/or updated annotation and pathogenicity prediction tools and variant databases, of great value in the context of variant evaluation. Recognizing the importance of data sharing, 48 novel pathogenic or likely pathogenic variants were identified in this cohort and uploaded to public databases.

A few cases in this cohort demonstrated that complex phenotypes may be caused by complex genotypes, potentially complicating both clinical evaluation and data interpretation. For example, in cases 8039 and 9151 (Table 2), the findings in *TGFB3* and *MYH7* genes respectively, were classified as VUS due to the lack of cardiological data at the time of WES analysis and the report included a recommendation for subsequent re-evaluation.

Rare genetic phenomena including variable expressivity, incomplete penetrance and genetic heterogeneity, should also be considered when assessing WES results. At least 2 cases had secondary variants potentially contributing to the clinical expression of the primary condition. In case 9081 (Table 2), referred with polycystic kidney disease at the age of 31, the detection of the co-inheritance of the pathogenic *PKD2:* c.1837C>T variant associated with autosomal dominant PKD2, and variant c.6992T>A in the*PKHD1* gene, (associated with autosomal recessive PKD4) could explain the relative early onset of the disease, although*PKHD1* :c.6992T>A classification is VUS (Bergmann et al., 2011).

Case 9115 (Table 1) presented with spastic paraplegia with lower limb spasticity and weakness. WES identified a known pathogenic variant c.1245+5G>A (Polymeris et al., 2016) of paternal origin in*SPAST* gene, along with a known modifier polymorphism c.131C>T of maternal origin in the same gene (Parodi et al., 2018). Usually variants in *SPAST* are associated with spastic paraplegia type 4, where the age of onset ranges between 20 to 60 years; the infantile onset in this case might be attributed to the presence of the modifying polymorphism *in trans* .

Case 9045 (Table 1), highlights the importance of adjunct genetic analysis, such as classic karyotyping, since *ZFPM2* pathogenic variant are disease causative only in 46,XY individuals. A 6 years-old female with a 46,XY karyotype, was referred for Disorders of Sex Development (DSD). A novel, likely pathogenic variant c.192T>G of maternal origin was detected in the*ZFPM2* (*FOG2* ), related to 46,XY sex reversal 9 (SRXY9). Functional studies have suggested that failure of testis development is most likely caused by impaired or no interactions of the mutant ZFPM2 (proteins) with *GATA4* gene, which regulates early testis development (Bashamboo et al., 2014).

Pleiotropy was another interesting observation in this cohort.*GJB2* gene is associated with autosomal recessive deafness or autosomal dominant Vohwinkel syndrome (De Zwart-Storm et al., 2011). Two unrelated patients with deafness (Table 1) were homozygous for pathogenic variants c.71G>A (case 8016) and c.35delG (case 8039) in *GJB2* . Interestingly, a likely pathogenic variant c.524C>G was revealed in the same gene in case 9088 (Table 1) with ichthyosis and keratitis but no hearing impairment; heterozygosity of *GJB2:* c.524C>G likely explains the dermatological findings. Unlike recessive *GJB2* variants, in which the spectrum and phenotype-genotype correlations have been analyzed clearly, only a few studies of *GJB2* dominant variants have been reported. However approximately two-thirds of dominant *GJB2*mutations are shown to cause syndromic hearing loss associated with diverse skin lesions(H. Wang et al., 2017).

In cases where one gene is linked to more than one disease or disease variants, like *PTPN11,* distinction may be based on differential clinical features or previous reports of the specific variant.*PTPN11 is* usually associated with Noonan syndrome 1, as additionally supported by findings concerning 5 unrelated patients in this study (8112, 8133, 9080, 9138, 9148) (Table 1). In case 9165 (Table 1), a *PTPN11* variant c.1381G>T was associated with Leopard syndrome 1 (Carcavilla et al., 2013), but the patients' diagnosis was based on the phenotypic findings rather than the genotype (Osawa et al. 2009). Diligent clinical examination,

5

biochemical tests and other pre-WES genetic tests, such as conventional and molecular karyotype, all have important roles in supporting the interpretation of WES results.

In conclusion, classical genetic diagnosis strategies involve targeted single gene/exon analysis often with negative or inconclusive results. Within the referred period that our academic diagnostic laboratory uses WES, 85 cases were resolved upon the detection of 94 pathogenic variants 48 of which were novel contributing valuable information to the genotype spectrum of the disease. Despite the technical and diagnostic challenges associated with NGS, (phenotyping, clinical and genetic heterogeneity and variable expressivity), WES provides a unique opportunity to resolve the genetic etiology of disorders, end diagnostic odysseys and support the provision of appropriate medical management and genetic counseling. When used within the context of effective multidisciplinary collaboration between clinicians and laboratories, the clinical WES diagnostic tool becomes efficient and cost-effective and thus appropriate for first-tier genomic analysis.

However, it is important to recognize the limitations of WES, including low coverage of coding regions, inefficient capture of GC-rich regions, misalignment of reads to reference genome, presence of pseudogenes, homologous or repetitive regions, large deletions/duplications, complex rearrangements, triplet repeat and imprinting variants (Bertier et al. 2016; Taylor et al. 2015; White et al. 2017). Polygenic and/or multi-factorial diseases and limited data on VUS in coding genes not yet associated with human diseases hinder interpretation and call for alternative approaches, such as functional studies. Finally, Whole Genome Sequencing potentially addresses many of the limitations of WES and can additionally detect non-coding (deep intronic or non-coding RNAs) and structural variants (Smith et al., 2019).

**Conflict of interest:**

The authors declare no potential conflict of interests.

**Data availability statement:**

The data that support the findings of this study are openly available in ClinVar (SUB5285601) (https://www.ncbi.nlm.nih.gov/clinvar/). Further data that support the findings of this study are available on request from the corresponding author. These data are not publicly available due to privacy or ethical restrictions.

**References**

Adams, D. R., & Eng, C. M. (2018). Next-generation sequencing to diagnose suspected genetic disorders. In *New England Journal of Medicine* . https://doi.org/10.1056/NEJMra1711801

Amberger, J. S., Bocchini, C. A., Scott, A. F., & Hamosh, A. (2019). OMIM.org: Leveraging knowledge across phenotype-gene relationships.*Nucleic Acids Research* . https://doi.org/10.1093/nar/gky1151

Ayme, S. (2012). State of the art of rare disease activities in Europe: A EUCERD perspective. In *Orphanet Journal of Rare Diseases* . https://doi.org/10.1186/1750-1172-7-S2-A1

Baird, P. A., Anderson, T. W., Newcombe, H. B., & Lowry, R. B. (1988). Genetic disorders in children and young adults: A population study.*American Journal of Human Genetics* .

Bashamboo, A., Brauner, R., Bignon-Topalovic, J., Lortat-Jacob, S., Karageorgou, V., Lourenco, D., Guffanti, A., & McElreavey, K. (2014). Mutations in the FOG2/ZFPM2 gene are associated with anomalies of human testis determination. *Human Molecular Genetics* . https://doi.org/10.1093/hmg/ddu074

Baynam, G., Pachter, N., McKenzie, F., Townshend, S., Slee, J., Kiraly-Borri, C., Vasudevan, A., Hawkins, A., Broley, S., Schofield, L., Verhoef, H., Walker, C. E., Molster, C., Blackwell, J. M., Jamieson, S., Tang, D., Lassmann, T., Mina, K., Beilby, J., . . . Goldblatt, J. (2016). The rare and undiagnosed diseases diagnostic

service - Application of massively parallel sequencing in a state-wide clinical service. *Orphanet Journal of Rare Diseases* . https://doi.org/10.1186/s13023-016-0462-7

Bergmann, C., Von Bothmer, J., Bruchle, N. O., Venghaus, A., Frank, V., Fehrenbach, H., Hampel, T., Pape, L., Buske, A., Jonsson, J., Sarioglu, N., Santos, A., Ferreira, J. C., Becker, J. U., Cremer, R., Hoefele, J., Benz, M. R., Weber, L. T., Buettner, R., & Zerres, K. (2011). Mutations in multiple PKD genes may explain early and severe polycystic kidney disease. *Journal of the American Society of Nephrology* . https://doi.org/10.1681/ASN.2010101080

Bertier, G., Hetu, M., & Joly, Y. (2016). Unsolved challenges of clinical whole-exome sequencing: a systematic literature review of end-users' views. *BMC Medical Genomics* . https://doi.org/10.1186/s12920-016-0213-6

Bruel, A. L., Nambot, S., Quere, V., Vitobello, A., Thevenon, J., Assoum, M., Moutton, S., Houcinat, N., Lehalle, D., Jean-Marcais, N., Verloes, A., Karsenti, A., Goldenberg, A., Jacquette, A., Jouret, B., Laudier, B., Coubes, C., Francannet, C., Lehalle, D., . . . Thauvin-Robinet, C. (2019). Increased diagnostic and new genes identification outcome using research reanalysis of singleton exome sequencing. *European Journal of Human Genetics* . https://doi.org/10.1038/s41431-019-0442-1

Carcavilla, A., Santome, J. L., Pinto, I., Sanchez-Pozo, J., Guillen-Navarro, E., Martin-Frias, M., Lapunzina, P., & Ezquieta, B. (2013). LEOPARD Syndrome: A Variant of Noonan Syndrome Strongly Associated With Hypertrophic Cardiomyopathy. *Revista Espanola de Cardiologia (English Edition)* . https://doi.org/10.1016/j.rec.2012.09.015

De Zwart-Storm, E. A., Van Geel, M., Veysey, E., Burge, S., Cooper, S., Steijlen, P. M., Martin, P. E., & Van Steensel, M. A. M. (2011). A novel missense mutation in GJB2, p.Tyr65His, causes severe Vohwinkel syndrome. *British Journal of Dermatology* . https://doi.org/10.1111/j.1365-2133.2010.10058.x

Desvignes, J. P., Bartoli, M., Delague, V., Krahn, M., Miltgen, M., Beroud, C., & Salgado, D. (2018). VarAFT: A variant annotation and filtration system for human next generation sequencing data.*Nucleic Acids Research* . https://doi.org/10.1093/nar/gky471

Ewans, L. J., Schofield, D., Shrestha, R., Zhu, Y., Gayevskiy, V., Ying, K., Walsh, C., Lee, E., Kirk, E. P., Colley, A., Ellaway, C., Turner, A., Mowat, D., Worgan, L., Freckmann, M. L., Lipke, M., Sachdev, R., Miller, D., Field, M., . . . Roscioli, T. (2018). Whole-exome sequencing reanalysis at 12 months boosts diagnosis and is cost-effective when applied early in Mendelian disorders. *Genetics in Medicine* . https://doi.org/10.1038/gim.2018.39

Hartman, P., Beckman, K., Silverstein, K., Yohe, S., Schomaker, M., Henzler, C., Onsongo, G., Lam, H. C., Munro, S., Daniel, J., Billstein, B., Deshpande, A., Hauge, A., Mroz, P., Lee, W., Holle, J., Wiens, K., Karnuth, K., Kemmer, T., . . . Thyagarajan, B. (2019). Next generation sequencing for clinical diagnostics: Five year experience of an academic laboratory. *Molecular Genetics and Metabolism Reports* . https://doi.org/10.1016/j.ymgmr.2019.100464

Iglesias, A., Anyane-Yeboa, K., Wynn, J., Wilson, A., Truitt Cho, M., Guzman, E., Sisson, R., Egan, C., & Chung, W. K. (2014). The usefulness of whole-exome sequencing in routine clinical practice. *Genetics in Medicine* . https://doi.org/10.1038/gim.2014.58

Jalkh, N., Corbani, S., Haidar, Z., Hamdan, N., Farah, E., Abou Ghoch, J., Ghosn, R., Salem, N., Fawaz, A., Djambas Khayat, C., Rajab, M., Mourani, C., Moukarzel, A., Rassi, S., Gerbaka, B., Mansour, H., Baassiri, M., Dagher, R., Breich, D., . . . Chouery, E. (2019). The added value of WES reanalysis in the field of genetic diagnosis: Lessons learned from 200 exomes in the Lebanese population. *BMC Medical Genomics* . https://doi.org/10.1186/s12920-019-0474-y

Landrum, M. J., Lee, J. M., Riley, G. R., Jang, W., Rubinstein, W. S., Church, D. M., & Maglott, D. R. (2014). ClinVar: Public archive of relationships among sequence variation and human phenotype.*Nucleic Acids Research* . https://doi.org/10.1093/nar/gkt1113

7

Lim, E. C. P., Brett, M., Lai, A. H. M., Lee, S. P., Tan, E. S., Jamuar, S. S., Ng, I. S. L., & Tan, E. C. (2015). Next-generation sequencing using a pre-designed gene panel for the molecular diagnosis of congenital disorders in pediatric patients. *Human Genomics* . https://doi.org/10.1186/s40246-015-0055-x

Niguidula, N., Alamillo, C., Shahmirzadi Mowlavi, L., Powis, Z., Cohen, J. S., & Farwell Hagman, K. D. (2018). Clinical whole-exome sequencing results impact medical management. *Molecular Genetics and Genomic Medicine* . https://doi.org/10.1002/mgg3.484

Parodi, L., Fenu, S., Barbier, M., Banneau, G., Duyckaerts, C., Du Montcel, S. T., Monin, M. L., Said, S. A., Guegan, J., Tallaksen, C. M. E., Sablonniere, B., Brice, A., Stevanin, G., Depienne, C., & Durr, A. (2018). Spastic paraplegia due to SPAST mutations is modified by the underlying mutation and sex. *Brain* . https://doi.org/10.1093/brain/awy285

Polymeris, A. A., Tessa, A., Anagnostopoulou, K., Rubegni, A., Galatolo, D., Dinopoulos, A., Gika, A. D., Youroukos, S., Skouteli, E., Santorelli, F. M., & Pons, R. (2016). A series of Greek children with pure hereditary spastic paraplegia: clinical features and genetic findings. *Journal of Neurology* . https://doi.org/10.1007/s00415-016-8179-z

Rabbani, B., Mahdieh, N., Hosomichi, K., Nakaoka, H., & Inoue, I. (2012). Next-generation sequencing: Impact of exome sequencing in characterizing Mendelian disorders. In *Journal of Human Genetics* . https://doi.org/10.1038/jhg.2012.91

Reuter, C. M., Kohler, J. N., Bonner, D., Zastrow, D., Fernandez, L., Dries, A., Marwaha, S., Davidson, J., Brokamp, E., Herzog, M., Hong, J., Macnamara, E., Rosenfeld, J. A., Schoch, K., Spillmann, R., Loscalzo, J., Krier, J., Stoler, J., Sweetser, D., . . . Wheeler, M. T. (2019). Yield of whole exome sequencing in undiagnosed patients facing insurance coverage barriers to genetic testing. *Journal of Genetic Counseling* . https://doi.org/10.1002/jgc4.1161

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W. W., Hegde, M., Lyon, E., Spector, E., Voelkerding, K., & Rehm, H. L. (2015). 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. *Genetics in Medicine* . https://doi.org/10.1038/gim.2015.30

Sawyer, S. L., Hartley, T., Dyment, D. A., Beaulieu, C. L., Schwartzentruber, J., Smith, A., Bedford, H. M., Bernard, G., Bernier, F. P., Brais, B., Bulman, D. E., Warman Chardon, J., Chitayat, D., Deladoey, J., Fernandez, B. A., Frosk, P., Geraghty, M. T., Gerull, B., Gibson, W., . . . Boycott, K. M. (2016). Utility of whole-exome sequencing for those near the end of the diagnostic odyssey: Time to address gaps in care. In *Clinical Genetics* . https://doi.org/10.1111/cge.12654

Seaby, E. G., Pengelly, R. J., & Ennis, S. (2016). Exome sequencing explained: A practical guide to its clinical application.*Briefings in Functional Genomics* . https://doi.org/10.1093/bfgp/elv054

Smith, H. S., Swint, J. M., Lalani, S. R., Yamal, J. M., de Oliveira Otto, M. C., Castellanos, S., Taylor, A., Lee, B. H., & Russell, H. V. (2019). Clinical Application of Genome and Exome Sequencing as a Diagnostic Tool for Pediatric Patients: a Scoping Review of the Literature. In *Genetics in Medicine* . https://doi.org/10.1038/s41436-018-0024-6

Taylor, J. C., Martin, H. C., Lise, S., Broxholme, J., Cazier, J. B., Rimmer, A., Kanapin, A., Lunter, G., Fiddy, S., Allan, C., Aricescu, A. R., Attar, M., Babbs, C., Becq, J., Beeson, D., Bento, C., Bignell, P., Blair, E., Buckle, V. J., . . . McVean, G. (2015). Factors influencing success of clinical genome sequencing across a broad spectrum of disorders. *Nature Genetics* . https://doi.org/10.1038/ng.3304

Taylor, J., Craft, J., Blair, E., Wordsworth, S., Beeson, D., Chandratre, S., Cossins, J., Lester, T., Nemeth, A. H., Ormondroyd, E., Patel, S. Y., Pagnamenta, A. T., Taylor, J. C., Thomson, K. L., Watkins, H., Wilkie, A. O. M., & Knight, J. C. (2019). Implementation of a genomic medicine multi-disciplinary team

approach for rare disease in the clinical setting: A prospective exome sequencing case series. *Genome Medicine* . https://doi.org/10.1186/s13073-019-0651-9

Valencia, C. A., Husami, A., Holle, J., Johnson, J. A., Qian, Y., Mathur, A., Wei, C., Indugula, S. R., Zou, F., Meng, H., Wang, L., Li, X., Fisher, R., Tan, T., Hogart Begtrup, A., Collins, K., Wusik, K. A., Neilson, D., Burrow, T., ... Zhang, K. (2015). Clinical Impact and Cost-Effectiveness of Whole Exome Sequencing as a Diagnostic Tool: A Pediatric Center's Experience. *Frontiers in Pediatrics* . https://doi.org/10.3389/fped.2015.00067

Wang, H., Wu, K., Yu, L., Xie, L., Xiong, W., Wang, D., Guan, J., & Wang, Q. (2017). A novel dominant GJB2 (DFNA3) mutation in a Chinese family. *Scientific Reports* . https://doi.org/10.1038/srep34425

Wang, K., Li, M., & Hakonarson, H. (2010). ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research* . https://doi.org/10.1093/nar/gkq603

White, S. J., Laros, J. F. J., Bakker, E., Cambon-Thomsen, A., Eden, M., Leonard, S., Lochmuller, H., Matthijs, G., Mattocks, C., Patton, S., Payne, K., Scheffer, H., Souche, E., Thomassen, E., Thompson, R., Traeger-Synodinos, J., Van Vooren, S., Janssen, B., & den Dunnen, J. T. (2017). Critical points for an accurate human genome analysis. In *Human Mutation* . https://doi.org/10.1002/humu.23238

Yang, Y., Muzny, D. M., Reid, J. G., Bainbridge, M. N., Willis, A., Ward, P. A., Braxton, A., Beuten, J., Xia, F., Niu, Z., Hardison, M., Person, R., Bekheirnia, M. R., Leduc, M. S., Kirby, A., Pham, P., Scull, J., Wang, M., Ding, Y., ... Eng, C. M. (2013). Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *New England Journal of Medicine* . https://doi.org/10.1056/NEJMoa1306555

**Figure 1: Patient cohort percentages classified according to the clinical features of the primary referral.**

**Figure 2: The six-step phenotypic-driven strategy applied for variant filtration and interpretation.**

(1) Targeted analysis of the WES data, focusing on a list of genes associated with the phenotype of the patients indicated by the clinical features according to HPO (Human Phenotype Ontology) (https://hpo.jax.org/app/) and/or OMIM (Online Mendelian Inheritance in Man) (https://www.omim.org/) and/or in-house in-silico gene lists, which were supplemented with genes from literature searches and established protein-protein interaction networks; (2) Exclusion of variants with Minor Allele Frequency (MAF) >1% according to the gnomAD database (https://gnomad.broadinstitute.org/) and/or the 1000 Genomes Project database (http://www.1000genomes.org/), (3) Subsequent filtering of remaining variants based on their category and gene/genomic localization, initially discarding synonymous variants, intronic variants, variants in intergenic regions and, 3' and 5' UTR regions. (At this step, synonymous variants were kept on a separately list for further investigation if appropriate); (4) The resulting variant list (including rare, nonsense, missense, splice site, and frameshift INDELs), was subsequently filtered according to the in-silico pathogenicity prediction, phenotype correlation, mode of inheritance and whether variants had already been recorded and classified using the VarSome database (https://varsome.com/). Human Splicing Finder (https://www.genomnis.com/access-hsf) was used to assess suspected splice site variants along with a concomitant review of the literature; (5) variant classification according to the American College of Medical Genetics and Genomics (ACMG) recommendations. The five categories comprise pathogenic, likely pathogenic, Variant of Uncertain Significance (VUS), likely benign and benign variants; (6) Finally, a multidisciplinary team of molecular biologists, laboratory geneticists and clinical geneticists, evaluated the variant(s) most relevant to the patient's phenotype and decided on the variants to be disclosed in the final report.

**Figure 3: Diagnostic yield (%) per category of disorder.**

Based on 100% in each disorder category, the diagnostic rate is shown with black box where a molecular diagnosis was made or with grey where no diagnosis made.

**Hosted file**

Tables.pdf available at https://authorea.com/users/372015/articles/490176-clinical-utility-of-whole-exome-sequencing-for-rare-mendelian-disorders-phenotypic-driven-strategy-for-a-high-diagnostic-yield-and-identification-of-48-novel-variants