



OPEN

Five years' experience of the clinical exome sequencing in a Spanish single center

A. Arteche-López^{1,2}, A. Ávila-Fernández¹, R. Riveiro Álvarez¹, B. Almoguera¹, A. Bustamante Aragonés¹, I. Martín-Merida¹, M. A. López Martínez¹, A. Giménez Pardo¹, C. Vélez-Monsalve¹, J. Gallego Merlo¹, I. García Vara¹, F. Blanco-Kelly¹, S. Tahsin Swafiri¹, I. Lorda Sánchez¹, M. J. Trujillo Tiebas¹ & C. Ayuso¹✉

Nowadays, exome sequencing is a robust and cost-efficient genetic diagnostic tool already implemented in many clinical laboratories. Despite it has undoubtedly improved our diagnostic capacity and has allowed the discovery of many new Mendelian-disease genes, it only provides a molecular diagnosis in up to 25–30% of cases. Here, we comprehensively evaluate the results of a large sample set of 4974 clinical exomes performed in our laboratory over a period of 5 years, showing a global diagnostic rate of 24.62% (1391/4974). For the evaluation we establish different groups of diseases and demonstrate how the diagnostic rate is not only dependent on the analyzed group of diseases (43.12% in ophthalmological cases vs 16.61% in neurological cases) but on the specific disorder (47.49% in retinal dystrophies vs 24.02% in optic atrophy; 18.88% in neuropathies/paraparesias vs 11.43% in dementias). We also detail the most frequent mutated genes within each group of disorders and discuss, on our experience, further investigations and directions needed for the benefit of patients.

Exome sequencing (ES) has revolutionized the diagnosis of genetic disorders, and has contributed to the discovery of new Mendelian-disease genes¹. Although nearly 85% of the disease causing variants are located in the coding regions or in the canonical splicing sites², a global molecular diagnosis is nowadays only established in up to 25%–30% of cases^{3–7}. However, this ES diagnostic capacity is variable and it depends on the analysed disorder^{5,7}, the type of study performed (TRIO (31–37%) vs Singleton (21–22%))^{6,8} and the cohort investigated^{9,10}.

The clinical exome sequencing -which comprises genes associated to known clinical phenotypes and covers >98% of variants identified on targeted next-generation sequencing (NGS) panels¹¹- has been implemented in many clinical genetics laboratories, becoming a robust and cost-efficient diagnostic tool¹², despite its inherent challenges and limitations¹³.

A recent overview of the different genetic approaches used to evaluate undiagnosed diseases has been published, emphasizing on how NGS can be incorporated in routine clinical practice¹⁴. The Genetics Department of the Fundación Jimenez Diaz University Hospital (FJD-HU) in Madrid (Spain) is a well-established and experienced genetics laboratory that integrates cytogenetic and molecular techniques. Before the NGS-era, the analysis of point variants and the sequencing of genes were offered by Sanger sequencing. Gradually, after the NGS implementation, its protocols have been adapted to this new technology. Nowadays, we offer all our patients an initial clinical evaluation and a molecular testing for each disorder being the NGS and, specifically CES, the first diagnostic approach for many of them.

Here, we present a 5-year retrospective review of 4974 clinical exomes performed (from October 2016 to April 2021), analysed and informed in our laboratory. We describe our NGS cases, their distribution, and the global and the specific diagnostic rate (DR) achieved for each type of disorder. We also remark the importance of having an experienced team of clinical and molecular geneticists and emphasise the importance of appropriate communication with fellow clinicians, in an increasingly complex and demanding field.

¹Department of Genetics, Health Research Institute–Jimenez Diaz Foundation University Hospital (IIS-FJD), Avda. Reyes Católicos 2, 28040 Madrid, Spain. ²Department of Genetics, University Hospital, 12 de Octubre, Madrid, Spain. ✉email: cayuso@fjd.es

Methods

Patients and samples. All patients were referred to the Genetics Department of the Fundación Jimenez Diaz University Hospital (Madrid, Spain) with a clinical suspicion of a genetic disorder from different Clinical Departments. All underwent full clinical genetic evaluations. Only if indicated and after an adequate pre-test genetic counselling, patients were molecularly studied. Written informed consent was obtained from each participant in accordance with institutional requirements. The study was reviewed and approved by the Research Ethics Committee of HU-FJD and fulfilled the principles of the Declaration of Helsinki.

Whole peripheral blood samples were collected from probands ($n = 4974$). Data on sex, age and available clinical indication for genetic testing was retrospectively recollected.

Genomic DNA was extracted following standard procedures. From October 2016 to April 2021, clinical exomes were performed using the TruSight One Sequencing panel (TSO; *Illumina Inc, San Diego, CA, USA*) ($N = 612$), and the Clinical Exome Solution (CES; *Sophia Genetics, USA*) ($N = 4362$) that capture 4813 and 4900 genes related to inherited diseases, respectively. Over 95% of the captured genes overlapped in both approaches. The detection of Copy Number Variants (CNVs) was only possible in samples sequenced with CES.

All samples were sequenced on a NextSeq500 instrument. Bioinformatics analysis for the detection of single-nucleotide variants (SNVs) and small insertions/deletions (indels) in TSO samples was performed using the Burrows-Wheeler alignment (BWA) Enrichment App, which provides industry-standard alignment. Variant calling was based on the Genome Analysis Toolkit (GATK). An in-house Sophia Genetics pipeline was used for the analysis of single nucleotide variants in CES samples. The detection of CNVs in CES samples was also performed using an in-house Sophia Genetics method based on a hidden Markov model. Specifically, the method evaluates the coverage levels of the target regions across all samples within the same sequencing run. The coverage is normalized by sample and target region, thus enabling CNV calling. A minimum of 8 samples per run is required.

Informed NGS variants in the first 825 clinical exomes of fully characterized cases were validated by Sanger sequencing before reporting¹⁵. In the remaining samples, only indels and bad-quality NGS variants (filter \neq PASS, QUAL < 100 , coverage $< 20X$ or variant fraction $< 20\%$) were further confirmed.

For CES samples, all CNVs were confirmed before being reported, either by Multiplex ligation-dependent Probe amplification (MLPA) or by array-CGH (60 k), depending on the availability of the MLPA kit and the size of the CNVs.

The analysis of the mitochondrial DNA or the clinical exome analysis in trio was not performed in any of the analysed patients.

Panel content and variant's interpretation. The Variant Studio (*-Illumina-*) and the SOPHiA DDM™ (*Sophia Genetics, USA*) platforms were used for the analysis of variants.

Variant interpretation was performed by our team of experienced molecular geneticists, making use of population and clinical databases (gnomAD, ClinVar), the literature and other tools such as Varsome or InterVar. Variants were classified according to the ACMG criteria in the 5 established categories¹⁶: benign (B), likely benign (LB), unknown clinical significance (VUS), likely pathogenic (LP) and pathogenic (P). Only P and LP variants, as well as VUS clearly related with the phenotype were reported.

In some cases, when only a LP or P variant was detected in a recessive related-gene, additional testing such as MLPA, array CGH or Sanger sequencing of uncovered regions was performed to attempt detecting the second allele, specifically in samples analysed by TSO as CNVs are not detected with this technology.

Clinical virtual panels were developed based on public database (OMIM, Genomics England PanelApp, DisGeNet, GTR) and medical specific literature. Specifically, retinal dystrophy panels were based on the RetNet database. Each panel was periodically re-evaluated—at least once a year- and thus, the panel content changed over time during the study.

Data analysis. “Molecular diagnosis” was defined as the presence of a P or LP variant in a dominant or X linked gene; as two P or LP, one P and one LP variant in a recessive gene, respectively, sex regardless. This group was used to define the global DR in our cohort and among the different groups of diseases.

“Potential molecular diagnosis” was established with a VUS in a dominant gene potentially explaining the phenotype, two VUS in a recessive gene potentially explaining the phenotype or VUS in a X-linked gene.

“Partial molecular diagnosis” includes cases with only one P or LP variant in a recessive gene (monoallelic) potentially explaining the phenotype.

“No molecular diagnosis” was defined when no variants were reported or when either a monoallelic P, LP or VUS in recessive gene or a VUS in a dominant gene not clearly related with the clinical suspicion was detected.

For data analysis, we established different groups of diseases based on the more demanding Clinical Departments (Ophthalmology, Neurology, Cardiology, Endocrinology, Nephrology) and/or on the more demanding diseases (Retinal disorders, Hearing impairment, Skeletal dysplasia, Cystic Fibrosis, Familial hypercholesterolemia and Miscellanea group). Within the majority of them, different subgroups of diseases were established.

A summary an overview of the distribution of the 4974 analysed cases are shown in Fig. 1.

Results

Clinical exome cases and its distribution by groups of disease. A total of 4974 patients (2389 females, 48.03%, and 2585 males, 51.97%) were analysed. The mean age (\pm SD) at diagnosis was 36.1 ± 22.8 years, ranging from 0 to 96 years of age.

Since the implementation of the NGS in the laboratory, we have faced a continuous and exponential growth in the number of clinical exomes requested. In the last 3 years, we approximately performed 1300–1400 clinical exomes per year (Fig. 2).

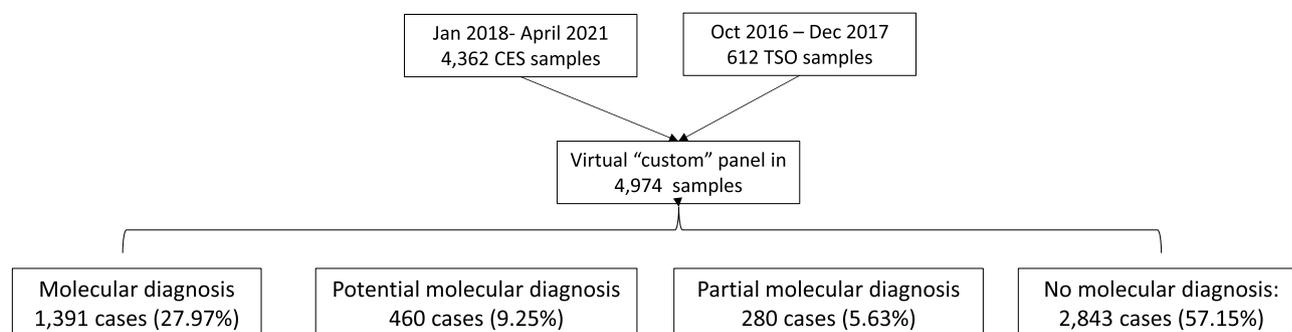


Figure 1. Summary and overview of the 4974 analysed cases. “Molecular diagnosis”: a P or LP variant in a dominant or X linked gene; as two P or LP, one P and one LP variant in a recessive gene, respectively, sex regardless; “Potential molecular diagnosis”: a VUS in a dominant gene potentially explaining the phenotype, two VUS in a recessive gene potentially explaining the phenotype or VUS in a X-linked gene. “Partial molecular diagnosis”: only one P or LP variant in a recessive gene (monoallelic) potentially explaining the phenotype. “No molecular diagnosis”: no variants or either a monoallelic P, LP or VUS in recessive gene or a VUS in a dominant gene not clearly related with the clinical suspicion were reported.

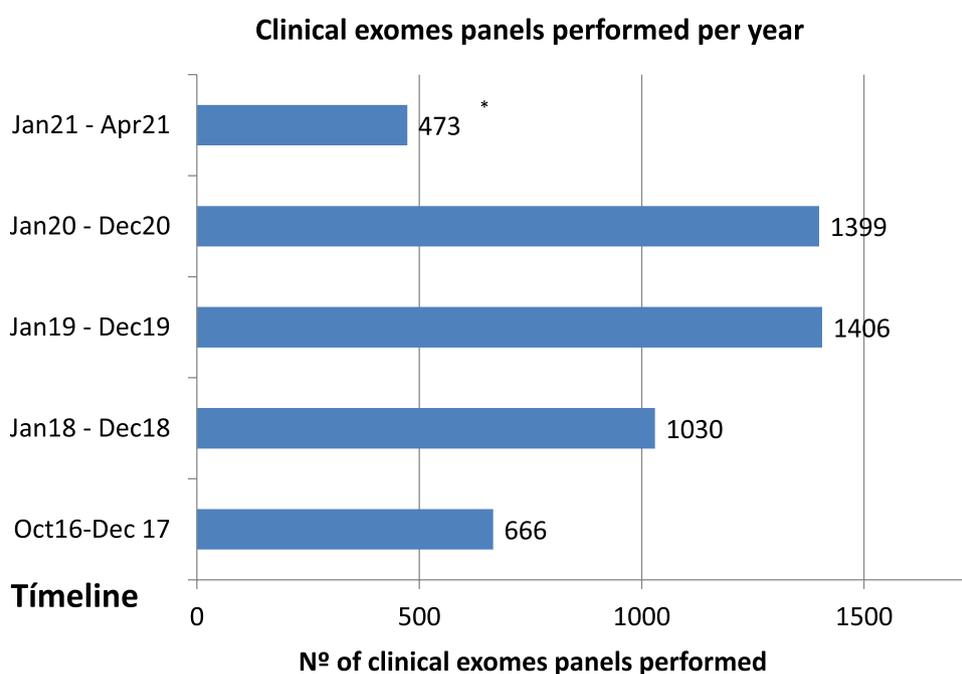


Figure 2. Number of clinical exomes panels performed in our laboratory per year. * Data of only 4 months.

We analysed a total of 4974 clinical exome panels. The two more frequent groups of diseases were Ophthalmological ($n = 1904$; 38.3%) and Neurological disorders ($n = 1529$; 30.74%), followed by Cardiological cases including those with clinical suspicion of Marfan syndrome, Noonan or other Rasopathies ($n = 356$; 7.16%). A total of 272 (5.47%) Endocrinological disorders, and 150 (3.02%) Kidney disorder cases were analysed. Skeletal dysplasias ($n = 134$), Hearing impairment ($n = 133$), Cystic Fibrosis ($n = 80$) and Familial Hypercholesterolemia ($N = 70$) corresponded to 2.69%, 2.67%, 1.61% and 1.41% of total cases, respectively. The remaining 346 cases (6.96%) corresponded to a miscellany of disorders that includes Autoinflammatory, Metabolic and Dermatology cases among others (Fig. 3).

Diagnostic rates among established groups. A molecular and a potential molecular diagnosis was established in 27.97% ($n = 1391/4974$) and 9.25% ($n = 460/4974$) of total cases, respectively.

In 280 cases (5.63%; $280/4974$), a partial molecular diagnosis was achieved. No molecular diagnosis was established in 57.16% of samples ($2843/4974$), including few cases with a VUS in a recessive reported gene with doubtful association with the clinical suspicion (Table 1).

Among the molecular and potential molecular diagnosis group in CES samples ($n = 1592$), a deletion or duplication was detected in 29 cases (1.8%), of which 23 (12 deletions and 11 duplications) belonged to the

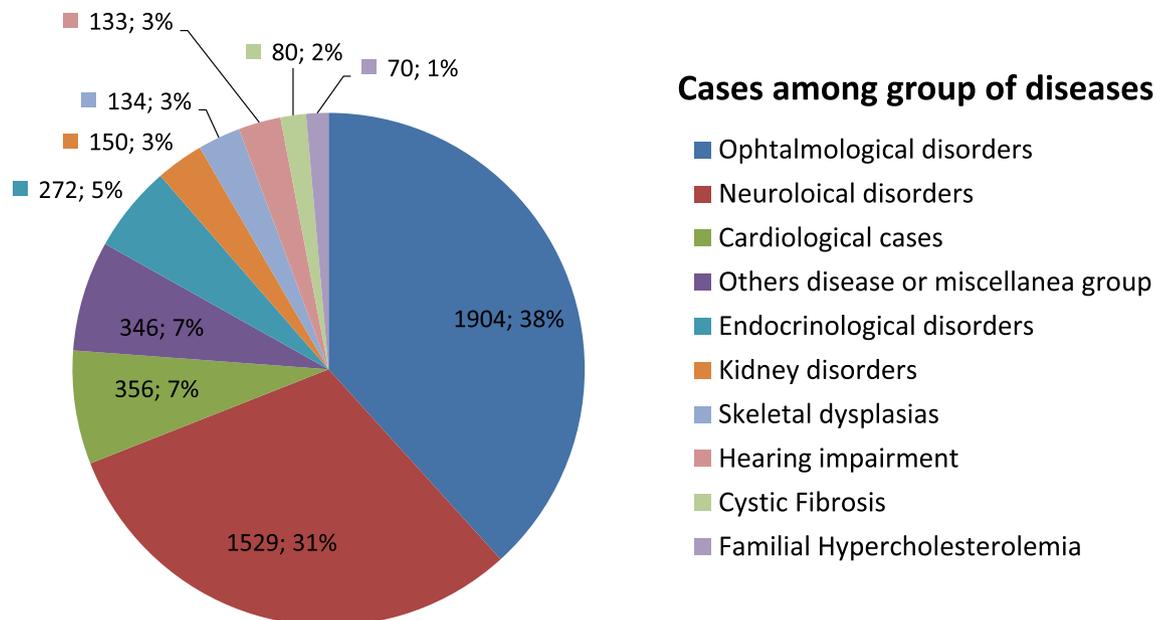


Figure 3. Analysed cases among groups of diseases. N = number of analysed cases; %: percentage of total (n = 4974) analysed cases.

Category	N	% Total
No molecular diagnosis	2843	57.15%
Molecular diagnosis (Solved cases)	1391	27.97%
Potential molecular diagnosis	460	9.25%
Partial molecular diagnosis	280	5.63%
Total	4974	100.00%

Table 1. Molecular, potential molecular, partial molecular and no molecular diagnosis cases in the analysed cohort.

molecular diagnosis group (1.9%; n = 23/1191). Interestingly, the chromosomes with more detected CNVs were chromosome 15 and X (8 vs 4 cases respectively), followed by chromosomes 1 and 17 (3 cases each).

Diagnostics rates based on group of disease. We observed that the DR was highly dependent on the indication of study and on the knowledge of the disease. The best molecular DR was achieved in cases derived from Ophthalmology (n = 821/1904; 43.12%) followed by familial hypercholesterolemia cases (n = 27/70; 38.57%) and hearing impairment (n = 31/133; 25.56%). In Nephrology (n = 38/150) and Neurological cases (n = 254/1529) we achieved a DR of 25.33% and 16.61%, respectively (Fig. 4 and Table 2).

As shown in Table 2, the DR achieved is significantly different when comparing retinal dystrophy (n = 680/1432; 47.49%) to optic atrophy (n = 43/179; 24.02%) or ataxia (n = 37/196; 18.88%) or intellectual disability (114/692; 16.47%) to dementia (n = 12/105; 11.43%), among others (Table 2).

Distribution of genes involved in molecular diagnosis. Specifically, in molecular diagnosis cases and excluding cases with deletions/duplications, genes observed 10 or more times and the five most mutated genes among the different disorders are shown in Table 3.

Discussion

Over the last 5 years we have implemented clinical exome sequencing as the first line test for many Mendelian disorders. This has led to a continuous and exponential growth in the number of cases analysed (Fig. 2), reflecting the increasing impact and the importance of genetics in the medical field¹⁷.

ES is a robust and cost-efficient diagnostic tool but, based on large cohort studies, it only provides an overall molecular diagnosis in 25–30% of cases^{4,6}. In our cohort, we obtained a molecular diagnosis in 1391 patients out of the 4974 analysed cases, corresponding to an overall DR of 24.62%. Although this DR could seem lower than reported, we should consider that an important number of potential molecular cases will be likely reclassified to molecular diagnosis cases after family segregation and/or after the upgrade of the VUS, thus, increasing our DR. It has been reported that approximately 74.6% and 25.4% of VUS are further respectively reclassified to benign/likely benign and to pathogenic/likely pathogenic variants¹⁸. Therefore, if we consider that 70% of our

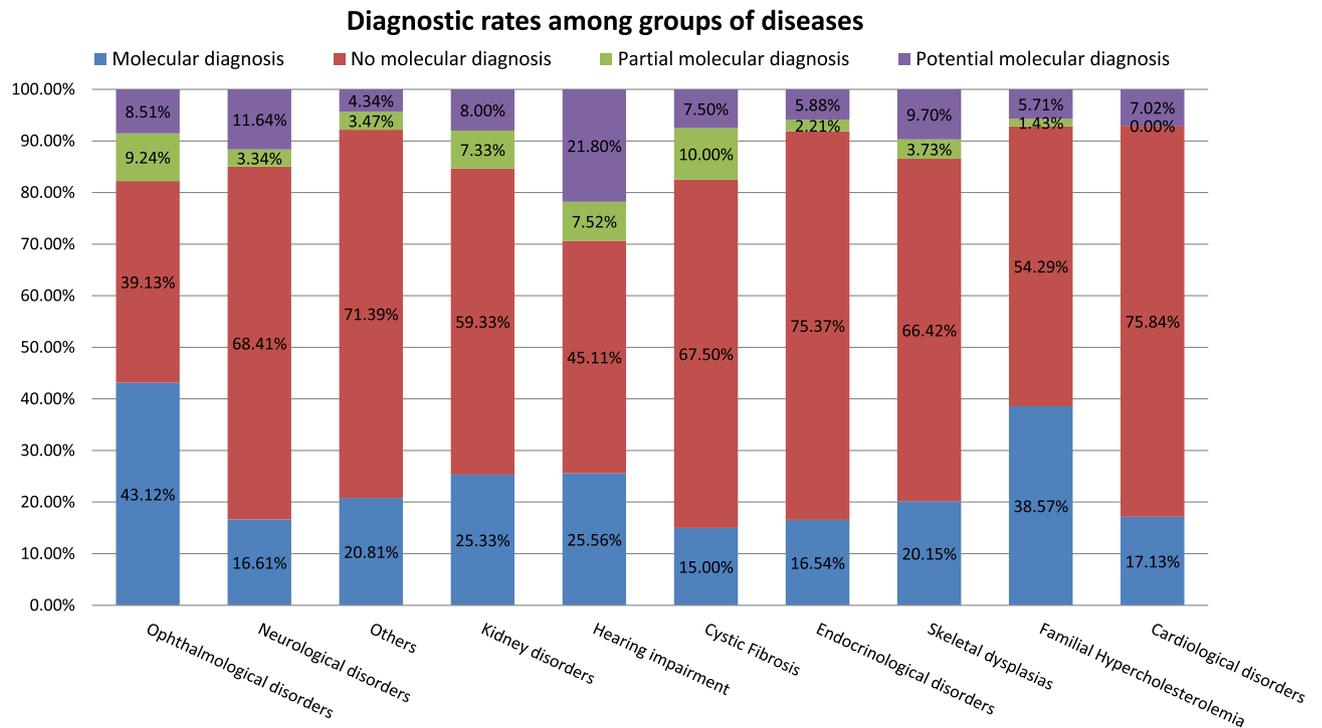


Figure 4. Diagnostic rates among groups of diseases. “Molecular diagnosis”: a P or LP variant in a dominant or X linked gene; as two P or LP, one P and one LP variant in a recessive gene, respectively, sex regardless; “Potential molecular diagnosis”: a VUS in a dominant gene potentially explaining the phenotype, two VUS in a recessive gene potentially explaining the phenotype or VUS in a X-linked gene. “Partial molecular diagnosis”: only one P or LP variant in a recessive gene (monoallelic) potentially explaining the phenotype. “No molecular diagnosis”: no variants or either a monoallelic P, LP or VUS in recessive gene or a VUS in a dominant gene not clearly related with the clinical suspicion were reported.

potential molecular diagnosis cases ($n = 460$) correspond to cases with a VUS in a dominant gen, we could estimate that 17% of potential molecular cases will be reclassified to molecular cases, giving a global DR of around 29% (1469/4974), which falls into the rates previously reported by other groups.

As already described, the DR of ES is very variable depending on the analysed disorder⁷. Based on the literature, it is estimated to be approximately of 36% in developmental disorders^{19,20} but around 50% and 58% in hearing loss and retinal dystrophies, respectively^{21,22}. In our cohort, we had the higher DR in ophthalmological and familial hypercholesterolemia cases (43.12% and 38.57% respectively) compared to other disorders such as neurological cases (16.61%), cardiology (17.13%) or others (20.81%) (Table 2). However, we noticed that the DR was disorder-specific rather than dependent on the group of diseases. For instance, we observed that the DR in retinal dystrophies (47.49%) was almost twice than in optic atrophy (24.02%), and in neuropathies/paraparesias was 18.88% while it was 11.43% in dementia cases. Probably, these differences are a direct consequence of the general knowledge about the genes associated to the disease, of our capacity to detect the causal gene and of the probability of a non-Mendelian aetiology of the disorder. The latter is reflected for example in the DR of dementia (11.43%) or parkinsonism (11.43%) vs retinal dystrophy (49.49%) or hearing impairment (25.56%). It is important to point out that we are not evaluating digenic inheritance or dual/multiple diagnoses, being those one of the main limitations of our study.

CNVs have been associated with a broad range of pathological conditions, accounting approximately for 13% of the human genome alterations²³. They could not be detected in our TSO analysed samples; however, at least in four TSO cases with high clinical suspicion and one allele detected in a disease-related gene, a definitive molecular diagnosis was established thanks to further MLPA analysis. Within the total 4362 CES samples analysed, we detected a deletion/duplication in 29 cases, of which 23 CNVs were detected within the molecular diagnosis group of CES samples, corresponding to a 1.9% (23/1191). Probably, we missed others due to the limitations of the CNVs detection by ES, which depends on the capture regions, the homogeneity and the coverage obtained for each sample.

The NGS diagnosis success strongly relies on the analysis and the strategy used for gene selection and variant filtering²⁴. The pipeline used during the analysis is also essential to improve diagnosis, as reflected in our capacity to increase the DR of previously unsolved cases in 2.5%, 3.2% and 4.4% of cancer, cardiovascular and retinal dystrophy cases respectively, as recently reported by our group²⁵. In addition, reanalysis of ES data represents a powerful approach to identify additional diagnosis and improve patient care^{26–28}. According to recent studies, exome reanalysis improves the DR in approximately 12%, with reported increases ranging from 5 to 26%²⁹. It is known that this increased DR is mainly due to the discovery of new disease-genes, the use of better bioinformatics tools, such as large databases and improved pipelines that help in the reclassification of variants,

	Molecular diagnosis		Potential molecular diagnosis		Partial molecular diagnosis		No molecular diagnosis		Total
	N	%	N	%	N	%	N	%	
Ophthalmological disorders	821	43.12	162	8.51	176	9.24	745	39.13	1904
Retinal dystrophy	680	47.49	132	9.22	159	11.10	461	32.19	1432
Optic Atrophy	43	24.02	4	2.23	5	2.79	127	70.95	179
Ocular malformation	76	34.55	20	9.09	12	5.45	112	50.91	220
Corneal dystrophy	22	30.14	6	8.22		0.00	45	61.64	73
Neurological disorders	254	16.61	178	11.64	51	3.34	1046	68.41	1529
Parkinsonism	8	11.43	5	7.14	4	5.71	53	75.71	70
Neuropathies/Paraparesia	37	18.88	29	14.80	10	5.10	120	61.22	196
Neurology others	12	20.69	2	3.45		0.00	44	75.86	58
Myopathies	34	18.89	19	10.56	5	2.78	122	67.78	180
Intellectual disability	114	16.47	76	10.98	17	2.46	485	70.09	692
Epilepsy	23	14.56	28	17.72	10	6.33	97	61.39	158
Dementia	12	11.43	7	6.67	1	0.95	85	80.95	105
Ataxia	14	20.00	12	17.14	4	5.71	40	57.14	70
Others diseases or Miscellanea group	72	20.81	15	4.34	12	3.47	247	71.39	346
Cardiological disorders	61	17.13	25	7.02		0.00	270	75.84	356
Rasopathies	7	20.59		0.00		0.00	27	79.41	34
Marfan syndrome	5	13.51		0.00		0.00	32	86.49	37
Cardiology	49	17.19	25	8.77		0.00	211	74.04	285
Endocrinological disorders	45	16.54	16	5.88	6	2.21	205	75.37	272
Kidney disorders	38	25.33	12	8.00	11	7.33	89	59.33	150
Nephrology others	34	27.42	11	8.87	10	8.06	69	55.65	124
Alport syndrome	4	15.38	1	3.85	1	3.85	20	76.92	26
Hearing impairment	34	25.56	29	21.80	10	7.52	60	45.11	133
Familial hypercholesterolemia	27	38.57	4	5.71	1	1.43	38	54.29	70
Skeletal dysplasia	27	20.15	13	9.70	5	3.73	89	66.42	134
Cystic Fibrosis	12	15.00	6	7.50	8	10.00	54	67.50	80
Total	1391	27.97	460	9.25	280	5.63	2843	57.16	4974

Table 2. Molecular, potential molecular, partial molecular and no molecular diagnosis cases of each group of diseases. Molecular diagnosis cases were used to establish the DR among each group and subgroup of disease. Data are ordered from highest to lowest according to the number of molecular diagnosis cases. Solved cases correspond to potential and molecular diagnosis cases.

as previously demonstrated by our group^{30,31}, and the availability of additional phenotype data, among others¹⁶. In this line, prioritized features and genes using Human Phenotype Ontology terminology have been proposed as a helping diagnostic tool^{32,33}. In addition, good communication with the clinicians ordering tests and an continuous genomic data sharing, is essential for the discovery of new genes -as already demonstrated with the GeneMatcher tool³⁴⁻³⁶-and for a better interpretation of variants that will definitely improve our diagnosis process³. Given the evolving nature of the genomic field, it is also essential to implement the exome reanalysis as part of routine clinical practice^{25,29,37}.

Big efforts are still needed to improve our genetic DR¹⁴. Although ES is gradually being optimized, it has limitations and in some cases generates laboratory's errors difficult to avoid³⁸. In this situation, whole genome sequencing (WGS) is becoming an attractive alternative as it is presumably able to detect more potential pathogenic variants than ES³⁹⁻⁴¹. Thanks to its greater breadth of coverage, genome sequencing (GS) is better than ES to detect single nucleotide variants and CNVs, also allowing the detection of variants in non-coding regions and chromosome rearrangements, all of which have been demonstrated to be involved in human disease^{3,42,43}. It is not casual that if we exclude cystic fibrosis cases due to its high prevalence in the general population, the higher partial molecular diagnosis is achieved in ophthalmological (9.2%), kidney (7.33%) and hearing impairment (7.52%) cases, in which the highest DR is reported in the literature. Undoubtedly, the analyses of non-coding regions or improved CNVs would help us to detect the second missed allele, increasing our global DR with the corresponding benefit for the patients.

Conclusions

Here, we present the DR of the CES we routinely perform in our clinical practice. Probably, the case selection, the continuous communication with clinicians and the expertise of our genetics team, make the difference in favour of our global estimated DR of 29%. Nevertheless, this overall DR is still low and direct efforts on the

Group of disorders	N	%
Ophthalmological disorders	821	60.01
<i>ABCA4</i>	137	16.69
<i>USH2A</i>	84	10.23
<i>BEST1</i>	31	3.78
<i>OPA1</i>	29	3.53
<i>PRPH2</i>	27	3.29
<i>CRB1</i>	21	2.56
<i>PAX6</i>	19	2.31
<i>CNGB3</i>	17	2.07
<i>EYS</i>	17	2.07
<i>RHO</i>	16	1.95
<i>NR2E3</i>	15	1.83
<i>RPGR</i>	15	1.83
<i>PRPF31</i>	14	1.71
<i>MYO7A</i>	13	1.58
<i>RS1</i>	12	1.46
<i>CHM</i>	11	1.34
<i>CRX</i>	10	1.22
<i>PROM1</i>	10	1.22
<i>BBS1</i>	10	1.22
<i>SNRNP200</i>	10	1.22
<i>TGFBI</i>	10	1.22
Neurological disorders	235	17.18
<i>ANKRD11</i>	9	3.83
<i>CLCN1</i>	7	2.98
<i>NOTCH3</i>	6	2.55
<i>MECP2</i>	6	2.55
<i>KRIT1</i>	5	2.13
<i>SCN1A</i>	5	2.13
Others or Miscellanea group	72	5.26
<i>ATP7B</i>	5	6.94
<i>HMBS</i>	4	5.56
<i>ACVRL1</i>	4	5.56
<i>NF1</i>	3	4.17
<i>ENG</i>	3	4.17
Cardiological disorders	61	4.46
<i>MYBPC3</i>	10	16.39
<i>TTN</i>	8	13.11
<i>MYH7</i>	5	8.20
<i>DSP</i>	3	4.92
<i>FBN1</i>	3	4.92
Endocrinological disorders	45	3.29
<i>GCK</i>	11	24.44
<i>HNF1A</i>	6	13.33
<i>FGFR1</i>	3	6.67
<i>AR</i>	3	6.67
<i>PDX1</i>	2	4.44
<i>SPRY4</i>	2	4.44
<i>HNF4A</i>	2	4.44
Kidney disorders	37	2.70
<i>COL4A4</i>	10	27.03
<i>COL4A3</i>	8	21.62
<i>PKHD1</i>	2	5.41
<i>HNF1B</i>	2	5.41
Hearing impairment	33	2.41
<i>MYO7A</i>	5	15.15
Continued		

Group of disorders	N	%
<i>GJB2</i>	4	12.12
<i>MYO6</i>	3	9.09
<i>COL4A3</i>	2	6.06
<i>TECTA</i>	2	6.06
Familial Hypercholesterolemia	27	1.97
<i>LDLR</i>	25	92.59
Skeletal dysplasia	25	1.83
<i>COL2A1</i>	5	20.00
<i>COL1A1</i>	4	16.00
<i>SHOX</i>	2	8.00
Cystic Fibrosis	12	0.88
<i>CFTR</i>	12	100.00
Total	1368	100.00

Table 3. Most frequent mutated genes in molecular diagnosis cases excluding deletions/duplications cases (n = 1368) among the different disorders and ordered by frequency. Only mutated genes observed 10 or more times or the five most mutated genes (if found more than once) among the different disorders are shown.

analysis and interpretation of NGS are required. Large databases, improved pipelines, data sharing and further investigations for a deeper understanding of ES, GS and other focused genetics techniques are needed in order to improve our global DR in the benefit of patients.

Data availability

All data generated or analysed during this study are included in this published article.

Received: 19 May 2022; Accepted: 4 November 2022

Published online: 10 November 2022

References

- Bamshad, M. J. *et al.* Exome sequencing as a tool for Mendelian disease gene discovery. *Nat. Rev. Genet.* **12**, 745–755 (2011).
- Choi, M. *et al.* Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc. Natl. Acad. Sci. USA* **106**, 19096–19101 (2009).
- Burdick, K. J. *et al.* Limitations of exome sequencing in detecting rare and undiagnosed diseases. *Am. J. Med. Genet. A* **182**, 1400–1406 (2020).
- Yang, Y. *et al.* Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *N. Engl. J. Med.* **369**, 1502–1511 (2013).
- Retterer, K. *et al.* Clinical application of whole-exome sequencing across clinical indications. *Genet. Med.* **18**, 696–704 (2016).
- Lee, H. *et al.* Clinical exome sequencing for genetic identification of rare mendelian disorders. *JAMA J. Am. Med. Assoc.* **312**, 1880–1887 (2014).
- Hartman, P. *et al.* Next generation sequencing for clinical diagnostics: Five year experience of an academic laboratory. *Mol. Genet. Metab. Rep.* **19**, 100464 (2019).
- Farwell, K. D. *et al.* Enhanced utility of family-centered diagnostic exome sequencing with inheritance model-based analysis: results from 500 unselected families with undiagnosed genetic conditions. *Genet. Med.* **17**, 578–586 (2014).
- Alfares, A. *et al.* A multicenter clinical exome study in unselected cohorts from a consanguineous population of Saudi Arabia demonstrated a high diagnostic yield. *Mol. Genet. Metab.* **121**, 91–95 (2017).
- Kuperberg, M. *et al.* Utility of whole exome sequencing for genetic diagnosis of previously undiagnosed pediatric neurology patients. *J. Child. Neurol.* **31**, 1534–1539 (2016).
- LaDuca, H. *et al.* Exome sequencing covers >98% of mutations identified on targeted next generation sequencing panels. *PLoS ONE* **12**, e0170843 (2017).
- Stark, Z. *et al.* A prospective evaluation of whole-exome sequencing as a first-tier molecular test in infants with suspected monogenic disorders. *Genet. Med.* **18**, 1090–1096 (2016).
- Corominas, J. *et al.* Clinical exome sequencing—Mistakes and caveats. *Hum. Mutat.* <https://doi.org/10.1002/HUMU.24360> (2022).
- Schuler, B. A. *et al.* Lessons learned: next-generation sequencing applied to undiagnosed genetic diseases. *J. Clin. Invest.* **132**, e154942 (2022).
- Arteche-López, A. *et al.* Sanger sequencing is no longer always necessary based on a single-center validation of 1109 NGS variants in 825 clinical exomes. *Sci. Rep.* **2021**(11), 1–7 (2021).
- Richards, 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.* **17**, 405–423 (2015).
- Niguidula, N. *et al.* Clinical whole-exome sequencing results impact medical management. *Mol. Genet. Genomic. Med.* **6**, 1068–1078 (2018).
- Harrison, S. M. & Rehm, H. L. Is 'likely pathogenic' really 90% likely? Reclassification data in ClinVar. *Genome Med.* **11**, 72 (2019).
- Srivastava, S. *et al.* Meta-analysis and multidisciplinary consensus statement: exome sequencing is a first-tier clinical diagnostic test for individuals with neurodevelopmental disorders. *Genet. Med.* **21**, 2413–2421 (2019).
- Gieldon, L. *et al.* Diagnostic value of partial exome sequencing in developmental disorders. *PLoS ONE* **13**, e0201041 (2018).
- Cabanillas, R. *et al.* Comprehensive genomic diagnosis of non-syndromic and syndromic hereditary hearing loss in Spanish patients. *BMC Med. Genomics* **11**, 58 (2018).
- Lee, K. *et al.* High diagnostic yield of whole exome sequencing in participants with retinal dystrophies in a clinical ophthalmology setting. *Am. J. Ophthalmol.* **160**, 354–363.e9 (2015).

23. Stankiewicz, P. & Lupski, J. R. Structural variation in the human genome and its role in disease. *Annu. Rev. Med.* **61**, 437–455 (2010).
24. Shamseldin, H. E. *et al.* Increasing the sensitivity of clinical exome sequencing through improved filtration strategy. *Genet. Med.* **19**, 593–598 (2017).
25. Romero, R. *et al.* An evaluation of pipelines for DNA variant detection can guide a reanalysis protocol to increase the diagnostic ratio of genetic diseases. *npj Genomic Med.* **7**, 1–10 (2022).
26. Salfati, E. L. *et al.* Re-analysis of whole-exome sequencing data uncovers novel diagnostic variants and improves molecular diagnostic yields for sudden death and idiopathic diseases. *Genome Med.* **11** (2019).
27. Tan, N. B. *et al.* Evaluating systematic reanalysis of clinical genomic data in rare disease from single center experience and literature review. *Mol. Genet. Genomic Med.* **8**, e1508 (2020).
28. Liu, P. *et al.* Reanalysis of clinical exome sequencing data. *N. Engl. J. Med.* **380**, 2478–2480 (2019).
29. Ji, J., Leung, M. L., Baker, S., Deignan, J. L. & Santani, A. Clinical exome reanalysis: Current practice and beyond. *Mol. Diagn. Ther.* **25**, 529–536 (2021).
30. Iancu, I. F. *et al.* Prioritizing variants of uncertain significance for reclassification using a rule-based algorithm in inherited retinal dystrophies. *npj Genomic Med.* **2021**(6), 1–9 (2021).
31. Iancu, I. F. *et al.* Aggregated genomic data as cohort-specific allelic frequencies can boost variants and genes prioritization in non-solved cases of inherited retinal dystrophies. *Int. J. Mol. Sci.* **23**, 8431 (2022).
32. Deelen, P. *et al.* Improving the diagnostic yield of exome-sequencing by predicting gene–phenotype associations using large-scale gene expression analysis. *Nat. Commun.* **10** (2019).
33. Perea-Romero, I. *et al.* NGS and phenotypic ontology-based approaches increase the diagnostic yield in syndromic retinal diseases. *Hum. Genet.* **140**, 1665–1678 (2021).
34. Towne, M. C. *et al.* Diagnostic testing laboratories are valuable partners for disease gene discovery: 5-year experience with GeneMatcher. *Hum. Mutat.* **43**, 772–781 (2022).
35. McWalter, K., Torti, E., Morrow, M., Juusola, J. & Retterer, K. Discovery of over 200 new and expanded genetic conditions using GeneMatcher. *Hum. Mutat.* **43**, 760–764 (2022).
36. Hamosh, A. *et al.* The impact of GeneMatcher on international data sharing and collaboration. *Hum. Mutat.* **43**, 668–673 (2022).
37. Deignan, J. L. *et al.* Points to consider in the reevaluation and reanalysis of genomic test results: A statement of the American College of Medical Genetics and Genomics (ACMG). *Genet. Med.* **21**, 1267–1270 (2019).
38. Corominas, J. *et al.* Clinical exome sequencing—Mistakes and caveats. *Hum. Mutat.* **43**, 1041–1055 (2022).
39. Belkadi, A. *et al.* Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. *Proc. Natl. Acad. Sci. USA* **112**, 5473–5478 (2015).
40. Meienberg, J., Bruggmann, R., Oexle, K. & Matyas, G. Clinical sequencing: is WGS the better WES?. *Hum. Genet.* **135**, 359–362 (2016).
41. Lionel, A. C. *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**, 435–443 (2018).
42. Klopocki, E. & Mundlos, S. Copy-number variations, noncoding sequences, and human phenotypes. *Annu. Rev. Genomics Hum. Genet.* **12**, 53–72 (2011).
43. Turro, E. *et al.* Whole-genome sequencing of patients with rare diseases in a national health system. *Nature* **583**, 96–102 (2020).

Acknowledgements

The authors would also like to thank to the clinical staff of University Hospital Foundation Jimenez Diaz, who made this study possible.

Author contributions

Wrote the first draft of the manuscript: A.A.L. Contribution to data preparation and acquisition: A.A.L., M.A.L.M., A.G.P., C.V., J.G.M., I.G.V. Contribution to the design of the study and analysis of data: A.A.L., A.A.F., C.A.G. Critical review of the manuscript for important intellectual content: A.A.F., R.R.A., M.J.T., B.A., I.L.S., I.M.M., C.A.G., F.B.K. Agreement with manuscript results and conclusions: All authors. Supervision: C.A.G. All the authors have read the manuscript and agree on publication.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to C.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022