MDPI

Article

Approach to Cohort-Wide Re-Analysis of Exome Data in 1000 Individuals with Neurodevelopmental Disorders

Insa Halfmeyer [1,†], Tobias Bartolomaeus [1,†], Bernt Popp [1,2] , Maximilian Radtke [1], Tobias Helms [3], Julia Hentschel [1], Denny Popp [1] and Rami Abou Jamra [1,*]

1 Institute of Human Genetics, University of Leipzig Medical Center, 04103 Leipzig, Germany
2 Center of Functional Genomics, Berlin Institute of Health at Charité, Universitätsmedizin Berlin, Hessische Straße 4A, 10115 Berlin, Germany
3 Limbus Medical Technologies GmbH, Neuer Markt 9/10, 18055 Rostock, Germany
* Correspondence: rami.aboujamra@medizin.uni-leipzig.de
† These authors contributed equally to this work.

Abstract: The re-analysis of nondiagnostic exome sequencing (ES) has the potential to increase diagnostic yields in individuals with rare diseases, but its implementation in the daily routines of laboratories is limited due to restricted capacities. Here, we describe a systematic approach to re-analyse the ES data of a cohort consisting of 1040 diagnostic and nondiagnostic samples. We applied a strict filter cascade to reveal the most promising single-nucleotide variants (SNVs) of the whole cohort, which led to an average of 0.77 variants per individual that had to be manually evaluated. This variant set revealed seven novel diagnoses (0.8% of all nondiagnostic cases) and two secondary findings. Thirteen additional variants were identified by a scientific approach prior to this re-analysis and were also present in this variant set. This resulted in a total increase in the diagnostic yield of 2.3%. The filter cascade was optimised during the course of the study and finally resulted in sensitivity of 85%. After applying the filter cascade, our re-analysis took 20 h and enabled a workflow that can be used repeatedly. This work is intended to provide a practical recommendation for other laboratories wishing to introduce a resource-efficient re-analysis strategy into their clinical routine.

Keywords: re-analysis; exome sequencing; neurodevelopmental disorder

1. Introduction

Next-generation sequencing technologies have made it possible to bring approaches such as the cost-effective exome sequencing (ES) of individuals with a disease of presumed genetic origin into everyday clinical practice.

ES has a reported diagnostic yield of 30–50% in neurodevelopmental disorders (NDD) [1]. However, many affected individuals remain undiagnosed, which hinders appropriate clinical care. With increasing knowledge of gene–disease associations, the rising number of entries in variant databases [2], the implementation of functional studies and the improvement of bioinformatics tools, the re-analysis of nondiagnostic cases is one way to close this diagnostic gap [3].

The exact increase in the diagnostic yield obtained through re-analysis varies between studies, but was recently summarised in a review as 10% overall [4]. Typically, cohorts of approximately 50–100 individuals are manually re-analysed case by case [5–7]. Although this is readily conceivable in a research context, it is far from the reality of laboratories in a diagnostic setting, with limited staff capacity and a lack of reimbursement options. This raises the following question: how can we achieve the important task of re-analysis that can help to diagnose so many affected individuals?

Others have already shown that, for larger cohorts, a semi-automated re-analysis is more appropriate and significantly reduces the workload per case, while still increasing the diagnostic yield [8,9]. Here, we describe a systematic approach for re-analysing ES

data from a cohort of 1040 individuals. In addition to the increased diagnostic yield, we describe the sensitivity of the filter cascade and its adjustments so that this workflow can be performed repeatedly in a time-effective manner and maximises the clinical impact while maintaining a small burden on the laboratory. Finally, we highlight the pitfalls and develop recommendations to help other laboratories that have to balance day-to-day analyses and subsequent re-analyses.

2. Materials and Methods

We re-analysed individuals with severe, early-onset diseases, mainly with NDD (i.e., intellectual disability, epilepsy, autism) (979/1040, 94.1%, see Figure 1D and Table S4). Expanding to other disease groups was beyond the scope of this work. We included individuals that had been analysed via ES over five years, between February 2017 and January 2022 (hereafter referred to as the "initial analysis"). The final cohort consisted of 1040 affected individuals from 983 families.
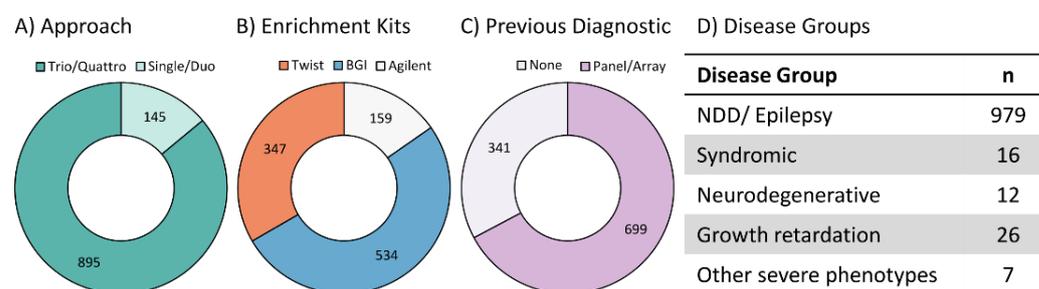

Figure 1. Cohort structure. Cohort structure of 1040 individuals regarding sequencing approach (**A**); enrichment kits used for ES—Agilent SureSelect Human All Exon V6, BGI Exome capture 59 M kit or TWIST Human Core Exome Kit (**B**); genetic tests conducted prior to exome sequencing (**C**); and disorder group (**D**). Numbers refer to the number of individuals. NDD: neurodevelopmental disorders.

In the initial analysis, we identified one or more (L)P disease-causing single-nucleotide variants (SNV) in 138/1040 individuals and (L)P copy-number variants (CNV) in 18/1040 individuals (hereafter referred to as "initially reported variants", see Table S1). This resulted in an ES diagnostic yield of 15% (155/1040, one individual had one SNV and one CNV). This number appears to be low; however, this is due to the multi-gene panel diagnostics, which clarified many cases before ES.

Prior to this study, re-evaluation was requested in some cases by the referring clinician but did not lead to positive reports. Most Trio-Exomes were assessed in a research context and promising candidate variants were reported. In thirteen individuals, such a candidate was subsequently published (see Table S2), making them valid results in a diagnostic setting.

In 159/1040 cases, the DNA was enriched using SureSelect Human All Exon V6 (Agilent Technologies, Santa Clara, CA, USA) (see Figure 1B). In 534/1040 cases, the BGI Exome Capture 59M Kit (BGI, Shenzhen, China) was performed, and in 347/1040 cases, the TWIST Human Core Exome Kit (TWIST Bioscience, San Francisco, CA, USA) was used for DNA enrichment. In 145/1040 cases, only the affected individual ("Single") or the individual with one parent was sequenced ("Duo"), while 895/1040 received Trio- or Quattro-ES (see. Figure 1A). Initial analysis was done using Varvis® genomics software (Limbus Medical Technologies GmbH, Rostock, Germany), using hg19 as a reference.

Prior to the initial ES analysis, 699/1040 individuals received multi-gene panel diagnostics (e.g., TruSightOne, Illumina, 4813 genes) and/or arrays (see Figure 1C). The remaining 341 individuals were the nondiagnostic cases of a larger ES cohort that were subsequently assessed on a research basis. The ES research assessment was regarded as the "initial analysis", since it was the analysis that we re-analysed in this study.

For this study, we reprocessed all cases using an updated bioinformatics pipeline. Sequencing data were aligned to the human genome hg38. Variants were called from the resulting bam files using the GATK HaplotypeCaller [10] (version 4.2.0.0) and SNVs were

annotated using vsWarehouse (Golden Helix, Inc., Bozeman, MT, USA, www.goldenhelix.com) (for further details, see File S1). The median time for BAM realignment (including all conversion and GATK pipeline recommended steps, such as deduplication and base quality score recalibration) was 1.79 h per sample, and the median time for VCF calling was 0.15 h per sample (PowerEdge R7515 Server; CPU: AMD EPYC 7702P 2.00 GHz with 64C/128 T; RAM: 196 GB; disk: 1.6 TB NVMe + 70 TB RAID). For all included subjects, the genomic regions targeted by the respective enrichment design had an average coverage of 149 reads, and >98% were covered by at least 15 reads. Copy number variants based on NGS were excluded from this analysis.

We then applied a filter cascade to identify the most promising variants. The multistep cascade included general filtering steps: (a) only SNVs with reliable quality; (b) only SNVs in diagnostically relevant genes, i.e., in genes that have been associated with a phenotype (morbidgenes.org, monthly updated database; version v2022-03.1 was used in this study and contained 4772 genes); (c) only SNVs in genes with a sufficient phenotypic overlap with the individual symptoms based on HPO terms [11], using the HPOsim score (Limbus Medical Technologies GmbH, Rostock, Germany) at a threshold of 0.1. This score reflects the similarity of the HPO term set of the individual and of the corresponding gene (for further details, see File S1).

Subsequently, we filtered the remaining SNVs based on different inheritance modi (see Figure 2).

Figure 2. Filter cascade and resulting variants. The filter cascade consists of several steps that take into account the quality of the variant, the gene–disease association, and its phenotypic overlap with the symptoms of the corresponding individual. Depending on the zygosity, we filtered by the frequency

of the variant in gnomAD and by other criteria, such as the variant effect, constraints, and presence in ClinVar, also considering segregation information (e.g., *de novo*). Numbers of manually evaluated variants are depicted for each arm of the filter cascade (bars without hatching). Initially reported (L)P variants that are covered by our filter cascade are displayed on the left in the colour of the corresponding filter criterion (bars with hatching). Notably, some variants appear in more than one filter step (e.g., a variant is both *de novo* and has been reported in ClinVar), and thus are counted repeatedly. In nine genes, we identified novel (L)P variants. CV: ClinVar, var.: variant, hem.: hemizygous, hom.: homozygous, LoF: loss of function, (L)P: (likely) pathogenic, MAF: minor allele frequency in gnomAD, pLI: probability of being loss-of-function-intolerant in gnomAD, Z: missense Z score in gnomAD.

Autosomal dominant: All rare heterozygous variants (not in gnomAD [12], release 2.0.1) were evaluated if they were linked to an autosomal dominant mode of inheritance in OMIM [13], and (a) *de novo* missense with a high missense gene constraint (Z score) [14,15], (b) *de novo* missense predicted to be deleterious by at least 4/5 in silico prediction tools (SIFT, PolyPhen2, MutationTaster, MutationAssessor, and FATHMM), (c) predicted to result in a loss of function (LoF) with a high LoF gene constraint (pLI score) [12], or (d) had previously been reported in the ClinVar database [16] as "likely pathogenic" or "pathogenic" ((L)P) by a laboratory other than ours until 7 April 2022.

Autosomal recessive: All homozygous variants were evaluated if they were not found in a homozygous state in gnomAD, with a minor allele frequency (MAF) < 0.01% and (a) predicted to result in a LoF, or (b) reported as (L)P in ClinVar. For compound-heterozygous variants, we filtered for genes with at least two variants that had an MAF < 0.01% in gnomAD and that were both not found in a homozygous state in gnomAD. In addition, the gene had to be linked to an autosomal recessive inheritance in OMIM and one of the variants had to (a) result in an LoF, or (b) be reported as (L)P in ClinVar previously. The other variant did not need to result in an LoF or to be reported as (L)P in ClinVar previously.

X-linked: In male individuals, we filtered for hemizygous variants on the X chromosome that were not found in a hemizygous state in gnomAD, with an MAF < 0.001%, that were (a) *de novo*, (b) predicted to result in an LoF, or (c) have been reported as (L)P in ClinVar. X-linked, dominant disorders are covered by the autosomal dominant filter.

The variants that passed the above-mentioned filter cascade were exported to Excel software (Microsoft Corporation, Redmond, Washington, DC, USA), and manually evaluated one by one by an experienced geneticist. The evaluation included checking the phenotypic overlap of the respective individual and the variant-associated disorder using the OMIM database, HGMD [17], or via literature research in PubMed. The DECIPHER database [18] was used for the visualisation of mutational hotspots. If needed, variant quality was assessed using Integrative Genomics Viewer [19]. For splicing prediction, variants were further assessed using AlamutVisual (Interactive Biosoftware, Rouen, France, v2.7.2) and spliceAI [20]. Classification was done according to the ACMG criteria [21], the ACGS Best Practice Guidelines for Variant Classification [22], and the latest updates published by the ClinGen consortium (https://clinicalgenome.org/working-groups/sequence-variant-interpretation/, accessed on 19 September 2022). If needed, validation via Sanger or Nanopore Sequencing was performed (for further details, see File S1). If relevant variants were identified, the referring clinicians were contacted.

The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of University of Leipzig, Germany (224/16-ek and 402/16-ek). Written informed consent for genetic testing and the publication of findings after providing advice and information about the study was obtained from all study subjects or their legal representatives.

3. Results

3.1. Manual Evaluation of Filtered Variants

We applied the described filter cascade (see Figure 2 and Methods) on the whole cohort, which revealed a list of 802 variants (0.77 variants per case). While some of the variants needed intensive research, others were easy to interpret. The manual evaluation of all variants took around 20 h. The heterozygous variants (see Figure 2) accounted for the largest proportion of all variants (493/802, 61.5%), followed by the compound heterozygous variants (212/802, 26.4%), while homozygous (73/802, 9.1%) and hemizygous (24/802, 3%) variants had smaller proportions.

3.2. Novel Diagnoses through Re-Analysis

Re-analysis of all samples revealed an additional nine (L)P SNVs (hereafter referred to as "novel" variants, see Table S3). Seven variants were disease-causing (primary findings), i.e., 0.8% (7 of 885 nondiagnostic exomes) diagnostic yield. Furthermore, two pathogenic variants were identified in two genes out of a list of 72 genes recommended to be reported as secondary findings by the ACMG guidelines (v3.0) [23]. Seven of these nine novel variants were heterozygous, one was hemizygous, and one was homozygous (see Table 1, and further details in File S1 and Table S3). Of the nine novel variants, in five cases, the gene–disease association was not published at the time of the initial analysis (*FOXP4, LMNB1, MORC2, MSL3, SORD*). The median interval between the initial negative report and the publication of new data on morbidity was nine months (3–53 months). The variants in *KMT2C, MTOR,* and *SDHC* were initially not called with freebayes (v1.1.0-9-g09d4ecf) [24], whereas the GATK HaplotypeCaller (version 4.2.0.0) managed to detect these variants. Moreover, in one variant, the gene (*TTN*) was not on the ACMG secondary findings list at the time of the initial analysis and has only recently been added [23]. We submitted the nine novel variants to the ClinVar database and re-contacted all referring clinicians of affected individuals with newly identified (L)P variants and all requested an updated report.

3.3. Sensitivity and Filter Adjustments

Our re-analysis correctly captured 125 of 147 initially reported diagnostic SNVs among 138 individuals, which resulted in sensitivity of 85% (further details in Table S1). Twenty-two initially reported (L)P SNVs could not be captured with our filter cascade, hereafter referred to as "lost" variants. Nine variants were too frequent in gnomAD (9/22, 41%) and six SNVs (6/22, 27%) were not detectable through our filter criteria (e.g., one homozygous missense variant that is not reported in ClinVar as (L)P; further details in Table S1). Two LoF variants were found in genes with a pLI score < 0.9, and one missense variant was found in a gene with a low Z score, which were filtered out. In addition, two variants had low quality in the re-analysis and one variant was filtered out, as the mode of inheritance in the associated gene did not match the mode of inheritance in OMIM. One variant was lost due to two reasons: a low Z score and mismatch in the mode of inheritance in OMIM.

The Z score threshold for missense variants was optimised during the course of the study. At first, a Z score of 4.0 for *de novo* missense variants was set, which resulted in a workload of 60 heterozygous variants. Compared with the initially reported pathogenic SNVs, this led to the loss of 62% (32/52, sensitivity of 38% for this filter approach) of the initially reported de novo missense variants. If only considering the Z score as a filter criterion, we found that a cut-off of 2.2 only lost three initially reported variants while keeping a moderate workload of 180 variants (see Figure 3A). Out of these three, one was rescued by another filter criterion (reported (L)P in ClinVar).

Table 1. Novel genetic diagnoses through re-analysis.

Gene	Approach (Date of Initial Evaluation)	Symptoms	Variant (Transcript, c-Code, p-Code)	Zygosity	Inheritance	ACMG Criteria (Final Classification)	OMIM Phenotype	Why Not Found Initially
1. *FOXP4*	Trio (08/2017)	Hearing impairment, ventricular septal defect, flattened epiphysis, disproportionate short stature, craniofacial asymmetry	NM_001012426.2:c.1540 G > A, p.Ala514Thr	Heterozygous, *de novo*	Autosomal dominant	PS2, PS3, PS4_MOD, PM2_SUP, PP3 (pathogenic)	-	New gene–disease association after 53 months (4 years and 5 months)
2. *KMT2C*	Trio (09/2017)	Hypothyroidism, mild intellectual disability, mild abnormality of facial shape, mild short stature	NM_170606.3:c.1829_1830delCA, p.Thr610Serfs * 4	Heterozygous, *de novo*	Autosomal dominant	PVS1, PS2_MOD, PS4_SUP, PM2_SUP (pathogenic)	Kleefstra syndrome 2 (#617768)	Updated caller, see also Figure S1
3. *LMNB1*	Trio (12/2019)	Microcephaly, agenesis of the corpus callosum, cerebellar hypoplasia, growth retardation (prenatal)	NM_005573.4:c.97A > G, p.Lys33Glu	Heterozygous, *de novo*	Autosomal dominant	PS2_VSTR, PS3, PS4_MOD, PM2_SUP, PP3 (pathogenic)	Microcephaly 26, primary, autosomal dominant (#619179)	New gene–disease association after 9 months
4. *MORC2*	Trio (05/2019)	Developmental delay, microcephaly	NM_001303256.3:c.79G > A, p.Glu27Lys	Heterozygous, *de novo*	Autosomal dominant	PS2_VSTR, PS3, PS4_MOD, PM2_SUP, PP2 (pathogenic)	Developmental delay, impaired growth, dysmorphic facies, and axonal neuropathy (#619090)	New gene–disease association after 15 months
5. *MSL3*	Trio (01/2018)	Global developmental delay, seizures, chylothorax, mid-aortic syndrome	NM_078629.4:c.973_974delAG, p.Gln326Alafs * 5	Hemizygous, *de novo*	X-chromosomal dominant	PVS1, PS2, PS4_SUP, PM2_SUP (pathogenic)	Basilicata-Akhtar syndrome (#301032)	New gene–disease association after 9 months
6. *MTOR*	Trio (9/2017)	Global developmental delay, macrocephaly	NM_004958.4:c.5911G > A, p.Ala1971Thr	Heterozygous, *de novo*	Autosomal dominant	PS2, PS4_SUP, PM2_SUP, PM5_SUP, PP2, PP3 (likely pathogenic)	Smith-Kingsmore syndrome (#616638)	Updated caller, see also Figure S2
7. *SORD*	Trio (02/2020)	Pain in both legs from the age of 17, ataxia, atrophy of the leg muscles	NM_003104.5:c.757del, p.Ala253Glnfs * 27	Homozygous, maternal and paternal	Autosomal recessive	PVS1, PS3, PM3_VSTR (pathogenic)	Sorbitol dehydrogenase deficiency with peripheral neuropathy (#618912)	New gene–disease association after 3 months
8. *SHDC* *	Trio (12/2017)	Mild global developmental delay, seizures, heterotopia, oral cleft, tall stature, obesity	NM_003001.5:c.377A > G, p.Tyr126Cys	Heterozygous, paternal	Autosomal dominant	PS4_MOD, PM1, PM2_SUP, PP3 (likely pathogenic)	Paragangliomas 3 (#605373)	Updated caller
9. *TTN* *	Trio (09/2020)	Panhypopituitarism, developmental delay, patent ductus arteriosus, scoliosis, short stature, median cleft lip and palate	NM_001267550.2:c.80762_80765delAACA, p.Lys26921Argfs * 5	Heterozygous, maternal	Autosomal dominant	PVS1, PM2_SUP (likely pathogenic)	Cardiomyopathy, dilated, 1G (#604145)	Updated AMCG secondary findings list

Genes marked with * are on the ACMG secondary findings list (v3.0).
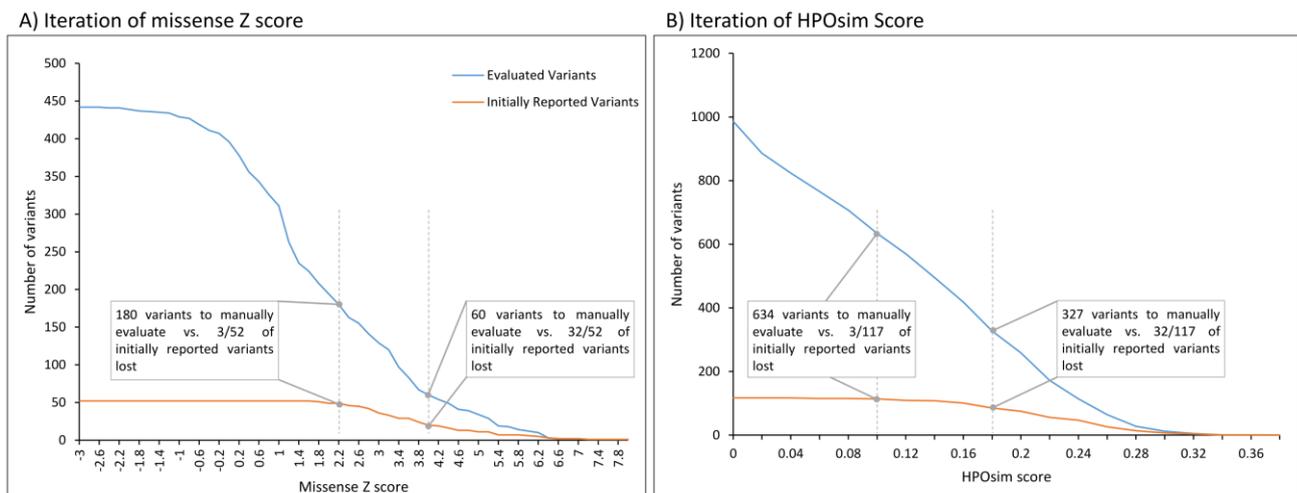
Figure 3. Iteration of missense Z score and HPOsim score. Number of variants to be manually evaluated (blue) and of initially reported (L)P variants (orange) based on missense Z score thresholds (**A**) and HPOsim score thresholds (**B**) as a single filter criterion. A) Depicted lines mark the adapted cut-off of 2.2 and the previously set cut-off of 4.0. B) Depicted lines mark the applied cut-off of 0.1 and an exemplary cut-off of 0.18. Notably, not all variants have an HPOSim score.

Application of the HPOsim score filter with a threshold of 0.1 reduced the variant evaluation burden by 35% (1236 vs. 802 variants; notably, not all variants have an HPOSim score), while three initially reported variants were lost (see Figure 3B). Two of them only partially explained the individuals' symptoms. In the third, the HPO terms used did not result in a sufficiently high overlap. In our opinion, the threshold of 0.1 provides the best balance between clinical sensitivity and workload.

4. Discussion

Here, we present an efficient strategy for the high-throughput re-analysis of ES data in severely affected individuals. This workflow is easily applicable to large cohorts and prioritises the identification of (L)P variants based on ACMG criteria. In our cohort of 1040 affected individuals, this approach resulted in nine clinically relevant findings, with an assessment time of 20 h for the entire cohort. We summarise our recommendations in Figure 4.

❑ Establish a recall-system for nondiagnostic cases
❑ Perform re-analysis after 18-24 months
❑ Re-annotate, not re-process
❑ Integrate HPO-matching
❑ Focus on *de novo* and LoF variants
❑ Trios are most promising

Figure 4. Checkbox for re-analysis of large cohorts.

Our results confirm the general assumption that the re-analysis of older cases is worthwhile, due to updated bioinformatic approaches and protocols (e.g., caller and alignment), the increasing knowledge of gene–disease associations and the rising number of entries in variant databases [2]. In our re-analysis, we did not identify (L)P variants that were reportable at the time of the initial analysis. This means that we did not uncover any human shortcomings. Interestingly, in our re-analysis, five out of the nine novel variants were classified as pathogenic because new data on the morbidity of the gene were published in the meantime. The median interval between the initial report date and the studies indicating the morbidity of the gene was nine months. A re-analysis after 15 months would have been beneficial in four out of five cases. Considering the small number of cases, no exact time interval for the optimal initiation time of a re-analysis can be given based on our data. However, in line with other authors, we suggest re-analysing ES data 18–24 months after a negative report [4,25].

In our cohort, heterozygous variants accounted for the largest proportion of all manually evaluated variants and represent 7/9 novel variants. This is in accordance with reports demonstrating that autosomal dominant inherited disorders account for the largest proportion in NDD [26]. All novel reported heterozygous variants in primary findings occurred *de novo*, so we recommend focusing on this subgroup if the dataset allows. Notably, the compound heterozygous variants accounted for the second-largest proportion of all variants, while not leading to a novel diagnosis. The evaluation of 97 homozygous and hemizygous variants resulted in one novel diagnosis each (*MSL3* in hemizygous, *SORD* in homozygous variants).

In 699/1040 individuals, the ES was performed following negative multi-gene panel diagnostics and/or array analysis. In the 341/1040 remaining cases, no other genetic testing was performed before ES. After ES remained negative in a diagnostic assessment, these 341 individuals were subsequently included and assessed in a research cohort. The initial diagnostic yield of 15% seems to be low. However, this is due to the cohort being preselected to include only individuals who were unremarkable in previous multi-gene panel diagnostics (e.g., TruSightOne, Illumina, 4813 genes) or in a pure diagnostic ES assessment. Additionally, all Trio-Exomes were analysed regarding new gene–disease associations and promising candidates were successively reported in a research setting. In thirteen individuals, such a candidate was published in the time between the initial analysis and this study (see Table S2). As we perform research in parallel to diagnostics at our institute, we add such novel variants during routine operation. Institutions that perform solely diagnostics would only find these thirteen variants during a re-analysis. Thus, the corrected diagnostic yield through re-analysis is 2.3% (20 novel primary findings in 885 nondiagnostic individuals).

The application of an updated analysis pipeline with an alignment against hg38 and new callers did not fail to detect initially reported variants. In contrast, three variants (in *KMT2C, MTOR,* and *SDHC*) were identified that were missed with the initially used caller and software (freebayes and Varvis®). The other six novel variants, as well as the thirteen research variants, could have been identified in the original data. Thus, taking into account the infeasibility of the application of novel bioinformatics pipelines, new alignments and variant calling can be omitted. A re-annotation with the latest information was sufficient to identify 17/20 variants. However, if one would consider the re-processing of the data (alignment, calling, and annotation), we recommend that this should be done only if significant modifications are introduced to an analysis pipeline, including the calling software (e.g., GATK) and the analysis software (in our case, Varvis® vs. GoldenHelix, but other providers exist).

The diagnostic yield of large, cohort-wide re-analyses [8] is below that of case-level re-analyses [5,27]. Our corrected diagnostic yield of 2.3% is in line with a study using a similar approach in a cohort of 4411 individuals that recently reported a 2.7% increase in diagnostic yield [8]. However, this and other large, cohort-wide studies [8,9,28] re-analysed

only nondiagnostic cases. The inclusion of diagnostic cases allowed us to examine the sensitivity of our filter cascade and to address recommendations on how to optimise them.

With the filter cascade described above, we correctly detected 125/147 (85%) initially reported disease-causing variants in the 1040 individuals. Of the 147 initially reported SNVs, 64 were heterozygous missense changes (52 of them *de novo*, see Table S1). The iteration of missense Z score thresholds (see Figure 3) in *de novo* missense variants led to a diagnostic optimum at 2.2, with sensitivity of 94 % (49/52), while maintaining a moderate workload (180 variants to be manually evaluated).

Two of the 22 lost variants were not detected because they failed our gnomAD frequency filter for heterozygous variants (see Table S1). Rescuing one of them by adding variants that are present only once in gnomAD is possible in this filtering step, but results in an additional workload of 86 heterozygous variants. To rescue both variants (including NM_145239.2: c.649dup in *PRRT2*), the needed MAF must be as high as 0.4%, which is not feasible.

The filtering step for likely LoF variants took into account a pLI score > 0.9. One variant in *COL10A1* (pLI: 0) and one in *MECP2* (pLI: 0.894) did not pass this filter and were also not rescued by other filter criteria. An additional eight variants (in *GH1, H1-4, MECP2, NARS1, NRP2, PPM1D, PRRT2*) also did not pass this filter but were rescued by other filter criteria. Thus, if only considering pLI values > 0.9 in genes with LoF variants, we lost ten of 34 initially reported (L)P variants in eight genes (sensitivity of 71% for this filter approach; see also Figure S3). Our filter cascade is partially based on the assumption that severe, early-onset diseases lead to reduced reproduction in affected individuals and that causative variants are thus subject to quantifiable selection pressure (MAF and pLI). The lost variants in *PRRT2* and *COL10A1* are good examples demonstrating that some conditions cannot be precisely distinguished from this group of severe disorders. Do the epileptic seizures of a 6-month-old infant occur in the context of developmental and epileptic encephalopathy, or in the context of *PRRT2*-associated benign epilepsy (OMIM #605751), in which seizures are confined to infancy [29]? While the former is usually associated with reduced reproduction (and high pLI), the latter tends to be unaffected and consequently does not lead to selection pressure. In conditions with reduced penetrance, as described for *COL10A1*-related metaphyseal chondrodysplasia (OMIM #156500), the causative variants are also subject to comparatively less selective pressure [30]. X-linked diseases are also not necessarily subject to the same selection pressure in gnomAD. In our re-analysis, two initially reported variants in *MECP2* (Rett syndrome, OMIM #312750) did not pass the pLI filter. We therefore recommend less stringent filters for X-linked variants in female individuals. The reasoning for the other five genes (*NARS1, PPM1D, GH1, NPR2, H1-4)* that we may have lost with our pLI threshold are further described in Table S1.

Our filtering cascade may be adapted to identify such variants in less severe or less penetrant diseases or in genes that slightly miss the pLI threshold, e.g., by a "white list" of genes with low pLI and LoF as a known pathomechanism (including the abovementioned genes as well as further genes, e.g., *BRCA1* and *BRCA2*). However, such adjustments go beyond the scope of this work, and as 31 of the 34 variants could be identified based on other steps of the filtering cascade, we decided not to adjust the pLI cut-off.

In four individuals, the initial single ES led to the segregation analysis of missense variants and confirmation of *de novo* occurrence. In another male individual, a maternal missense variant in the X-chromosomal gene *ABCD1* was reported as disease-causing. These five variants were not identified with our filter cascade (see Table S1). Additionally, 4/9 of the novel reported variants are *de novo* missense as well. Unless analysed in a trio approach or in a comprehensive variant assessment at the case level, these missense variants (i.e., not in ClinVar, no segregation information, or X-chromosomal) cannot be identified in a stringent, cohort-wide filter cascade. This demonstrates that the trio approach is of lasting advantage, as it also benefits the subsequent re-analysis.

If no trio information is available, these variants will probably continue to be lost in the short term. In the long term, it is likely that further criteria of the ACMG classification,

such as PM1 (variant is located in mutational hotspot or in functional domain) or PS1/PM5 (amino acid exchange at the same amino acid position) will be parameterised and become available for a filter cascade.

Of the 147 (L)P initially reported variants, 22 are compound heterozygous. We lost six of them with our filter cascade (see Table S1) because they are too common (MAF 0.01–0.1%). Raising the threshold for the MAF leads to a substantial increase in the number of variants to be evaluated manually, analogous to the heterozygous variants. Since the group of compound heterozygous variants is already the second largest in our filter cascade and no novel variants were identified, this subgroup is most negligible if resources are limited. However, it should be taken into account that autosomal recessive forms of NDD [31] are not as well clarified and that substantial development in this field is expected.

Application of the HPOsim score resulted in a substantial reduction in variants to be manually evaluated. Review of the HPOsim score curves (see Figure 3B) does not allow the recommendation of a clear threshold. By choosing a threshold of 0.1, we lost only three initially reported variants. These variants do not explain the entire phenotype and are likely to be lost in any phenotype-based filtering approach (see Table S1). There are several algorithms to calculate the phenotypic overlap (e.g., Phenomizer [32]). If the phenotypic overlap of an individual's presentation with a particular disease is not possible, we recommend a stricter assignment of individuals to a disease group (e.g., "focal epilepsy"). Thus, individuals of the same disease group can be re-analysed only with respect to a specific gene panel (e.g., by using SysNDD [33] or PanelApp [34]). This allows a faster evaluation of the cohort or improvement of the sensitivity due to less stringent filters (e.g., gnomAD < 10).

To enable the application of our approach to many other laboratories, we chose OMIM as the database for disorders and retrieved the inheritance modes of these disorders from there. Although a dominant mode of inheritance is described in the literature, two heterozygous variants did not pass our filter cascade because the associated disorders are deposited in OMIM only with a recessive mode of inheritance. While such weaknesses in the databases can hardly be compensated by the users, they are subject to the constant updates in the knowledge of genetic diseases and thus become more and more negligible.

5. Conclusions

As our results show, re-analysis is worthwhile and, most importantly, feasible for large cohorts. In this study, we highlighted pitfalls and provided recommendations based on our experience to facilitate application for other laboratories (see Figure 4). The cut-offs of the above filter cascade ultimately depend on the setting in which the re-analysis takes place. Certainly, the highest possible sensitivity is to be aimed for, but the workload must correspond to the resources of the laboratory.

We recommend a recall system in order to easily queue cases and perform re-analysis on a regular basis. If such systems are not available, software developers should be encouraged to implement recall algorithms, as re-analysis is both feasible and necessary. Smart gene- and disorder-specific filtering would allow us to set less stringent cut-offs regarding in silico predictions, MAF, etc., without increasing the workload, thus making the re-analysis more sensitive and more specific. This would also allow the expansion of the re-analysis to phenotypes that are not severe or of early onset.

Although there were edge cases that were presumably missed with our filter cascade, we conclude from our results that the filters stated above are a reasonable compromise between novel identified variants and needed workforce. In our laboratory, we plan to re-annotate all NGS data and use these filters on a semi-annual basis for cases that have not been analysed in the last 18 months.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/genes14010030/s1, File S1: Supplementary Information [35–59]; Table S1: Initially reported variants, Table S2: Candidate variants, Table S3: Novel variants, Table S4: Cohort structure; Figure S1: IGV screenshot of Nanopore Sequencing data in Patient 1, Figure S2: IGV screenshot of Illumina Sequencing data in Patient 2, Figure S3: Iteration of pLI score

Author Contributions: Conceptualisation, T.B. and R.A.J.; Methodology, B.P., M.R., I.H. and T.H.; Software, B.P., M.R. and I.H.; Validation, J.H. and D.P.; Formal Analysis, I.H.; Data Curation, I.H. and T.B.; Writing—Original Draft Preparation, I.H. and T.B.; Writing—Review and Editing, R.A.J.; Visualisation, I.H. and T.B.; Supervision, R.A.J.; Project Administration, R.A.J. All authors have read and agreed to the published version of the manuscript.

Funding: B.P. is supported by the Deutsche Forschungsgemeinschaft (DFG) through grant PO2366/2–1.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of the University of Leipzig Medical Center (224/16-ek, 26.10.2017, 402/16-ek, 9 July 2021).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study. Written informed consent has been obtained from the patient(s) to publish this paper.

Data Availability Statement: Not applicable.

Conflicts of Interest: T.H. is employed by Limbus, Medical Technologies GmbH, Rostock, Germany. The other authors declare no conflict of interest.

References

1. Seo, G.H.; Lee, H.; Lee, J.; Han, H.; Cho, Y.K.; Kim, M.; Choi, Y.; Choi, J.; Choi, I.H.; Rhie, S.; et al. Diagnostic Performance of Automated, Streamlined, Daily Updated Exome Analysis in Patients with Neurodevelopmental Delay. *Mol. Med.* **2022**, *28*, 38. [CrossRef] [PubMed]

2. Landrum, M.J.; Chitipiralla, S.; Brown, G.R.; Chen, C.; Gu, B.; Hart, J.; Hoffman, D.; Jang, W.; Kaur, K.; Liu, C.; et al. ClinVar: Improvements to Accessing Data. *Nucleic Acids Res.* **2020**, *48*, D835–D844. [CrossRef] [PubMed]

3. Deignan, J.L.; Chung, W.K.; Kearney, H.M.; Monaghan, K.G.; Rehder, C.W.; Chao, E.C. 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.* **2019**, *21*, 1267–1270. [CrossRef] [PubMed]

4. Dai, P.; Honda, A.; Ewans, L.; McGaughran, J.; Burnett, L.; Law, M.; Phan, T.G. Recommendations for next Generation Sequencing Data Reanalysis of Unsolved Cases with Suspected Mendelian Disorders: A Systematic Review and Meta-Analysis. *Genet. Med.* **2022**, *24*, 1618–1629. [CrossRef] [PubMed]

5. Wenger, A.M.; Guturu, H.; Bernstein, J.A.; Bejerano, G. Systematic Reanalysis of Clinical Exome Data Yields Additional Diagnoses: Implications for Providers. *Genet. Med.* **2017**, *19*, 209–214. [CrossRef]

6. Al-Nabhani, M.; Al-Rashdi, S.; Al-Murshedi, F.; Al-Kindi, A.; Al-Thihli, K.; Al-Saegh, A.; Al-Futaisi, A.; Al-Mamari, W.; Zadjali, F.; Al-Maawali, A. Reanalysis of Exome Sequencing Data of Intellectual Disability Samples: Yields and Benefits. *Clin. Genet.* **2018**, *94*, 495–501. [CrossRef]

7. Costain, G.; Jobling, R.; Walker, S.; Reuter, M.S.; Snell, M.; Bowdin, S.; Cohn, R.D.; Dupuis, L.; Hewson, S.; Mercimek-Andrews, S.; et al. Periodic Reanalysis of Whole-Genome Sequencing Data Enhances the Diagnostic Advantage over Standard Clinical Genetic Testing. *Eur. J. Hum. Genet.* **2018**, *26*, 740–744. [CrossRef]

8. Matalonga, L.; Hernández-Ferrer, C.; Piscia, D.; Schüle, R.; Synofzik, M.; Töpf, A.; Vissers, L.E.L.M.; de Voer, R.; Tonda, R.; Johari, M.; et al. Solving Patients with Rare Diseases through Programmatic Reanalysis of Genome-Phenome Data. *Eur. J. Hum. Genet.* **2021**, *29*, 1337–1347. [CrossRef]

9. Liu, P.; Meng, L.; Normand, E.A.; Xia, F.; Song, X.; Ghazi, A.; Rosenfeld, J.; Magoulas, P.L.; Braxton, A.; Ward, P.; et al. Reanalysis of Clinical Exome Sequencing Data. *N. Engl. J. Med.* **2019**, *380*, 2478–2480. [CrossRef]

10. Poplin, R.; Ruano-Rubio, V.; DePristo, M.A.; Fennell, T.J.; Carneiro, M.O.; Van der Auwera, G.A.; Kling, D.E.; Gauthier, L.D.; Levy-Moonshine, A.; Roazen, D.; et al. Scaling Accurate Genetic Variant Discovery to Tens of Thousands of Samples. *BioRxiv* **2017**, 201178. [CrossRef]

11. Köhler, S.; Gargano, M.; Matentzoglu, N.; Carmody, L.C.; Lewis-Smith, D.; Vasilevsky, N.A.; Danis, D.; Balagura, G.; Baynam, G.; Brower, A.M.; et al. The Human Phenotype Ontology in 2021. *Nucleic Acids Res.* **2021**, *49*, D1207–D1217. [CrossRef] [PubMed]

12. Karczewski, K.J.; Francioli, L.C.; Tiao, G.; Cummings, B.B.; Wang, Q.; Collins, R.L.; Laricchia, K.M.; Ganna, A.; Birnbaum, P.; Gauthier, L.D.; et al. Variation across 141,456 Human Exomes and Genomes Reveals the Spectrum of Loss-of-Function Intolerance across Human Protein-Coding Genes. *Nature* **2020**, *581*, 434–443. [CrossRef] [PubMed]

13. Hamosh, A.; Scott, A.F.; Amberger, J.S.; Bocchini, C.A.; McKusick, V.A. Online Mendelian Inheritance in Man (OMIM), a Knowledgebase of Human Genes and Genetic Disorders. *Nucleic Acids Res.* **2005**, *33*, D514–D517. [CrossRef] [PubMed]

14. Lek, M.; Karczewski, K.J.; Minikel, E.V.; Samocha, K.E.; Banks, E.; Fennell, T.; O'Donnell-Luria, A.H.; Ware, J.S.; Hill, A.J.; Cummings, B.B.; et al. Analysis of Protein-Coding Genetic Variation in 60,706 Humans. *Nature* **2016**, *536*, 285–291. [CrossRef] [PubMed]

15. Samocha, K.E.; Robinson, E.B.; Sanders, S.J.; Stevens, C.; Sabo, A.; McGrath, L.M.; Kosmicki, J.A.; Rehnström, K.; Mallick, S.; Kirby, A.; et al. A Framework for the Interpretation of de Novo Mutation in Human Disease. *Nat. Genet.* **2014**, *46*, 944–950. [CrossRef]

16. Landrum, M.J.; Lee, J.M.; Benson, M.; Brown, G.; Chao, C.; Chitipiralla, S.; Gu, B.; Hart, J.; Hoffman, D.; Hoover, J.; et al. ClinVar: Public Archive of Interpretations of Clinically Relevant Variants. *Nucleic Acids Res.* **2016**, *44*, D862–D868. [CrossRef]

17. Stenson, P.D.; Ball, E.V.; Mort, M.; Phillips, A.D.; Shiel, J.A.; Thomas, N.S.T.; Abeysinghe, S.; Krawczak, M.; Cooper, D.N. Human Gene Mutation Database (HGMD): 2003 Update. *Hum. Mutat.* **2003**, *21*, 577–581. [CrossRef]

18. Firth, H.V.; Richards, S.M.; Bevan, A.P.; Clayton, S.; Corpas, M.; Rajan, D.; Van Vooren, S.; Moreau, Y.; Pettett, R.M.; Carter, N.P. DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *Am. J. Hum. Genet.* **2009**, *84*, 524–533. [CrossRef]

19. Robinson, J.T.; Thorvaldsdóttir, H.; Winckler, W.; Guttman, M.; Lander, E.S.; Getz, G.; Mesirov, J.P. Integrative Genomics Viewer. *Nat. Biotechnol.* **2011**, *29*, 24–26. [CrossRef]

20. Jaganathan, K.; Panagiotopoulou, S.K.; McRae, J.F.; Darbandi, S.F.; Knowles, D.; Li, Y.I.; Kosmicki, J.A.; Arbelaez, J.; Cui, W.; Schwartz, G.B.; et al. Predicting Splicing from Primary Sequence with Deep Learning. *Cell* **2019**, *176*, 535–548. [CrossRef]

21. Richards, S.; Aziz, N.; Bale, S.; Bick, D.; Das, S.; Gastier-Foster, J.; Grody, W.W.; Hegde, M.; Lyon, E.; Spector, E.; 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*, 405–424. [CrossRef] [PubMed]

22. Ellard, S.; Baple, E.L.; Berry, I. ACGS Best Practic Guidelines for Variant Classification 2019. 2019. Available online: https://www.acgs.uk.com/media/11631/uk-practice-guidelines-for-variant-classification-v4-01-2020.pdf (accessed on 11 April 2022).

23. Miller, D.T.; Lee, K.; Chung, W.K.; Gordon, A.S.; Herman, G.E.; Klein, T.E.; Stewart, D.R.; Amendola, L.M.; Adelman, K.; Bale, S.J.; et al. ACMG SF v3.0 List for Reporting of Secondary Findings in Clinical Exome and Genome Sequencing: A Policy Statement of the American College of Medical Genetics and Genomics (ACMG). *Genet. Med.* **2021**, *23*, 1381–1390. [CrossRef] [PubMed]

24. Garrison, E.; Marth, G. Haplotype-Based Variant Detection from Short-Read Sequencing. *arXiv* **2012**. [CrossRef]

25. Tan, N.B.; Stapleton, R.; Stark, Z.; Delatycki, M.B.; Yeung, A.; Hunter, M.F.; Amor, D.J.; Brown, N.J.; Stutterd, C.A.; McGillivray, G.; et al. Evaluating Systematic Reanalysis of Clinical Genomic Data in Rare Disease from Single Center Experience and Literature Review. *Mol. Genet. Genom. Med.* **2020**, *8*, e1508. [CrossRef] [PubMed]

26. Wright, C.F.; Fitzgerald, T.W.; Jones, W.D.; Clayton, S.; McRae, J.F.; van Kogelenberg, M.; King, D.A.; Ambridge, K.; Barrett, D.M.; Bayzetinova, T.; et al. Genetic Diagnosis of Developmental Disorders in the DDD Study: A Scalable Analysis of Genome-Wide Research Data. *Lancet* **2015**, *385*, 1305–1314. [CrossRef]

27. Jalkh, N.; Corbani, S.; Haidar, Z.; Hamdan, N.; Farah, E.; Ghoch, J.A.; Ghosn, R.; Salem, N.; Fawaz, A.; Khayat, C.D.; et al. The Added Value of WES Reanalysis in the Field of Genetic Diagnosis: Lessons Learned from 200 Exomes in the Lebanese Population. *BMC Med. Genom.* **2019**, *12*, 11. [CrossRef] [PubMed]

28. Wright, C.F.; McRae, J.F.; Clayton, S.; Gallone, G.; Aitken, S.; FitzGerald, T.W.; Jones, P.; Prigmore, E.; Rajan, D.; Lord, J.; et al. Making New Genetic Diagnoses with Old Data: Iterative Reanalysis and Reporting from Genome-Wide Data in 1,133 Families with Developmental Disorders. *Genet. Med.* **2018**, *20*, 1216–1223. [CrossRef]

29. Chen, Y.; Chen, D.; Zhao, S.; Liu, G.; Li, H.; Wu, Z.-Y. Penetrance Estimation of PRRT2 Variants in Paroxysmal Kinesigenic Dyskinesia and Infantile Convulsions. *Front. Med.* **2021**, *15*, 877–886. [CrossRef]

30. Richmond, C.M.; Savarirayan, R. Schmid Metaphyseal Chondrodysplasia. In *GeneReviews®*; Adam, M.P., Everman, D.B., Mirzaa, G.M., Pagon, R.A., Wallace, S.E., Bean, L.J., Gripp, K.W., Amemiya, A., Eds.; University of Washington: Seattle, WA, USA, 2019.

31. Jamra, R. Genetics of Autosomal Recessive Intellectual Disability. *Med. Genet.* **2018**, *30*, 323–327. [CrossRef]

32. Köhler, S.; Schulz, M.H.; Krawitz, P.; Bauer, S.; Dölken, S.; Ott, C.E.; Mundlos, C.; Horn, D.; Mundlos, S.; Robinson, P.N. Clinical Diagnostics in Human Genetics with Semantic Similarity Searches in Ontologies. *Am. J. Hum. Genet.* **2009**, *85*, 457–464. [CrossRef]

33. Kochinke, K.; Zweier, C.; Nijhof, B.; Fenckova, M.; Cizek, P.; Honti, F.; Keerthikumar, S.; Oortveld, M.A.W.; Kleefstra, T.; Kramer, J.M.; et al. Systematic Phenomics Analysis Deconvolutes Genes Mutated in Intellectual Disability into Biologically Coherent Modules. *Am. J. Hum. Genet.* **2016**, *98*, 149–164. [CrossRef] [PubMed]

34. Martin, A.R.; Williams, E.; Foulger, R.E.; Leigh, S.; Daugherty, L.C.; Niblock, O.; Leong, I.U.S.; Smith, K.R.; Gerasimenko, O.; Haraldsdottir, E.; et al. PanelApp Crowdsources Expert Knowledge to Establish Consensus Diagnostic Gene Panels. *Nat. Genet.* **2019**, *51*, 1560–1565. [CrossRef] [PubMed]

35. Snijders Blok, L.; Vino, A.; den Hoed, J.; Underhill, H.R.; Monteil, D.; Li, H.; Reynoso Santos, F.J.; Chung, W.K.; Amaral, M.D.; Schnur, R.E.; et al. Heterozygous Variants That Disturb the Transcriptional Repressor Activity of FOXP4 Cause a Developmental Disorder with Speech/Language Delays and Multiple Congenital Abnormalities. *Genet. Med.* **2021**, *23*, 534–542. [CrossRef]

36. Koemans, T.S.; Kleefstra, T.; Chubak, M.C.; Stone, M.H.; Reijnders, M.R.F.; de Munnik, S.; Willemsen, M.H.; Fenckova, M.; Stumpel, C.T.R.M.; Bok, L.A.; et al. Functional Convergence of Histone Methyltransferases EHMT1 and KMT2C Involved in Intellectual Disability and Autism Spectrum Disorder. *PLOS Genet.* **2017**, *13*, e1006864. [CrossRef] [PubMed]

37. Kleefstra, T.; de Leeuw, N. Kleefstra Syndrome. In *GeneReviews®*; Adam, M.P., Everman, D.B., Mirzaa, G.M., Pagon, R.A., Wallace, S.E., Bean, L.J., Gripp, K.W., Amemiya, A., Eds.; University of Washington: Seattle, WA, USA, 1993.

38. Li, H. Minimap2: Pairwise Alignment for Nucleotide Sequences. *Bioinformatics* **2018**, *34*, 3094–3100. [CrossRef] [PubMed]

39. Cristofoli, F.; Moss, T.; Moore, H.W.; Devriendt, K.; Flanagan-Steet, H.; May, M.; Jones, J.; Roelens, F.; Fons, C.; Fernandez, A.; et al. De Novo Variants in LMNB1 Cause Pronounced Syndromic Microcephaly and Disruption of Nuclear Envelope Integrity. *Am. J. Hum. Genet.* **2020**, *107*, 753–762. [CrossRef] [PubMed]

40. Parry, D.A.; Martin, C.-A.; Greene, P.; Marsh, J.A.; Ambrose, J.C.; Arumugam, P.; Baple, E.L.; Bleda, M.; Boardman-Pretty, F.; Boissiere, J.M.; et al. Heterozygous Lamin B1 and Lamin B2 Variants Cause Primary Microcephaly and Define a Novel Laminopathy. *Genet. Med.* **2021**, *23*, 408–414. [CrossRef] [PubMed]

41. Guillen Sacoto, M.J.; Tchasovnikarova, I.A.; Torti, E.; Forster, C.; Andrew, E.H.; Anselm, I.; Baranano, K.W.; Briere, L.C.; Cohen, J.S.; Craigen, W.J.; et al. De Novo Variants in the ATPase Module of MORC2 Cause a Neurodevelopmental Disorder with Growth Retardation and Variable Craniofacial Dysmorphism. *Am. J. Hum. Genet.* **2020**, *107*, 352–363. [CrossRef]

42. Brunet, T.; McWalter, K.; Mayerhanser, K.; Anbouba, G.M.; Armstrong-Javors, A.; Bader, I.; Baugh, E.; Begtrup, A.; Bupp, C.P.; Callewaert, B.L.; et al. Defining the Genotypic and Phenotypic Spectrum of X-Linked MSL3-Related Disorder. *Genet. Med.* **2021**, *23*, 384–395. [CrossRef]

43. Basilicata, M.F.; Bruel, A.-L.; Semplicio, G.; Valsecchi, C.I.K.; Aktaş, T.; Duffourd, Y.; Rumpf, T.; Morton, J.; Bache, I.; Szymanski, W.G.; et al. De Novo Mutations in MSL3 Cause an X-Linked Syndrome Marked by Impaired Histone H4 Lysine 16 Acetylation. *Nat. Genet.* **2018**, *50*, 1442–1451. [CrossRef]

44. Smith, L.; Saunders, C.; Dinwiddie, D.; Atherton, A.; Miller, N.; Soden, S.; Farrow, E.; Abdelmoity, A.; Kingsmore, S. Exome Sequencing Reveals De Novo Germline Mutation of the Mammalian Target of Rapamycin (MTOR) in a Patient with Megalencephaly and Intractable Seizures. *J. Genomes Exomes* **2013**, *2013*, 63–72. [CrossRef]

45. Mroske, C.; Rasmussen, K.; Shinde, D.N.; Huether, R.; Powis, Z.; Lu, H.-M.; Baxter, R.M.; McPherson, E.; Tang, S. Germline Activating MTOR Mutation Arising through Gonadal Mosaicism in Two Brothers with Megalencephaly and Neurodevelopmental Abnormalities. *BMC Med. Genet.* **2015**, *16*, 102. [CrossRef] [PubMed]

46. Cortese, A.; Zhu, Y.; Rebelo, A.P.; Negri, S.; Courel, S.; Abreu, L.; Bacon, C.J.; Bai, Y.; Bis-Brewer, D.M.; Bugiardini, E.; et al. Biallelic Mutations in SORD Cause a Common and Potentially Treatable Hereditary Neuropathy with Implications for Diabetes. *Nat. Genet.* **2020**, *52*, 473–481. [CrossRef] [PubMed]

47. Else, T.; Greenberg, S.; Fishbein, L. Hereditary Paraganglioma-Pheochromocytoma Syndromes. In *GeneReviews®*; Adam, M.P., Everman, D.B., Mirzaa, G.M., Pagon, R.A., Wallace, S.E., Bean, L.J., Gripp, K.W., Amemiya, A., Eds.; University of Washington: Seattle, WA, USA, 1993.

48. Roberts, A.M.; Ware, J.S.; Herman, D.S.; Schafer, S.; Baksi, J.; Bick, A.G.; Buchan, R.J.; Walsh, R.; John, S.; Wilkinson, S.; et al. Integrated Allelic, Transcriptional, and Phenomic Dissection of the Cardiac Effects of Titin Truncations in Health and Disease. *Sci. Transl. Med.* **2015**, *7*, 270ra6. [CrossRef] [PubMed]

49. Herman, D.S.; Lam, L.; Taylor, M.R.G.; Wang, L.; Teekakirikul, P.; Christodoulou, D.; Conner, L.; DePalma, S.R.; McDonough, B.; Sparks, E.; et al. Truncations of Titin Causing Dilated Cardiomyopathy. *N. Engl. J. Med.* **2012**, *366*, 619–628. [CrossRef]

50. Tange, O. *GNU Parallel 2018*, 2018; ISBN 978-1-387-50988-1.

51. Di Tommaso, P.; Chatzou, M.; Floden, E.W.; Barja, P.P.; Palumbo, E.; Notredame, C. Nextflow Enables Reproducible Computational Workflows. *Nat. Biotechnol.* **2017**, *35*, 316–319. [CrossRef]

52. McKenna, A.; Hanna, M.; Banks, E.; Sivachenko, A.; Cibulskis, K.; Kernytsky, A.; Garimella, K.; Altshuler, D.; Gabriel, S.; Daly, M.; et al. The Genome Analysis Toolkit: A MapReduce Framework for Analyzing next-Generation DNA Sequencing Data. *Genome Res.* **2010**, *20*, 1297–1303. [CrossRef]

53. Li, H. Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM. *arXiv* **2013**, arXiv:1303.3997.

54. Freed, D.; Aldana, R.; Weber, J.A.; Edwards, J.S. The Sentieon Genomics Tools—A Fast and Accurate Solution to Variant Calling from next-Generation Sequence Data. *BioRxiv* **2017**, 115717. [CrossRef]

55. Bonfield, J.K.; McCarthy, S.A.; Durbin, R. Crumble: Reference Free Lossy Compression of Sequence Quality Values. *Bioinformatics* **2019**, *35*, 337–339. [CrossRef]

56. Kendig, K.I.; Baheti, S.; Bockol, M.A.; Drucker, T.M.; Hart, S.N.; Heldenbrand, J.R.; Hernaez, M.; Hudson, M.E.; Kalmbach, M.T.; Klee, E.W.; et al. Sentieon DNASeq Variant Calling Workflow Demonstrates Strong Computational Performance and Accuracy. *Front. Genet.* **2019**, *10*. [CrossRef]

57. Resnik, P. Using Information Content to Evaluate Semantic Similarity in a Taxonomy. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence—Volume 1*; IJCAI'95; Morgan Kaufmann Publishers Inc.: San Francisco, CA, USA, 1995; pp. 448–453.

58. Lin, D. An Information-Theoretic Definition of Similarity. In *Proceedings of the Fifteenth International Conference on Machine Learning*; ICML'98; Morgan Kaufmann Publishers Inc.: San Francisco, CA, USA, 1998; pp. 296–304.

59. Deng, Y.; Gao, L.; Wang, B.; Guo, X. HPOSim: An R Package for Phenotypic Similarity Measure and Enrichment Analysis Based on the Human Phenotype Ontology. *PLOS ONE* **2015**, *10*, e0115692. [CrossRef] [PubMed]