Genomic diagnosis for children with intellectual disability and/or developmental delay

Kevin M. Bowling, PhD[1], Michelle L. Thompson, PhD[1], Michelle D. Amaral, PhD[1], Candice R. Finnila, PhD[1], Susan M. Hiatt, PhD[1], Krysta L. Engel, PhD[1], J. Nicholas Cochran, PhD[1], Kyle B. Brothers, MD, PhD[3], Kelly M. East, MS, CGC[1], David E. Gray, MS[1], Whitley V. Kelley, MS, CGC[1], Neil E. Lamb, PhD[1], Edward J. Lose, MD[2], Carla A. Rich, MA[3], Shirley Simmons, RN[2], Jana S. Whittle, BS[1,4], Benjamin T. Weaver, BS[1,2], Amy S. Nesmith, BS[1], Richard M. Myers, PhD[1], Gregory S. Barsh, MD, PhD[1], E. Martina Bebin, MD, MPA[2], Gregory M. Cooper, PhD[1*]

[1]HudsonAlpha Institute for Biotechnology, Huntsville, AL, USA
[2]University of Alabama at Birmingham, Birmingham, AL, USA
[3]University of Louisville, Louisville, KY, USA
[4]University of Alabama in Huntsville, Huntsville, AL, USA

Kevin M. Bowling (kbowling@hudsonalpha.org), Michelle L. Thompson (mthompson@hudsonalpha.org), Michelle D. Amaral (mamaral@hudsonalpha.org), Candice R. Finnila (cfinnila@hudsonalpha.org), Susan M. Hiatt (shiatt@hudsonalpha.org), Krysta L. Engel (kengel@hudsonalpha.org), J. Nicholas Cochran (ncochran@hudsonalpha.org), Kyle B. Brothers (kyle.brothers@louisville.edu), Kelly M. East (keast@hudsonalpha.org), David E. Gray (dgray@hudsonalpha.org), Whitley V. Kelley (wkelley@hudsonalpha.org), Neil E. Lamb (nlamb@hudsonalpha.org), Edward J. Lose (elose@uab.edu), Carla A. Rich (carla.rich@louisville.edu), Shirley Simmons (shirley.simmons@childrensal.org), Jana S. Whittle (jwhittle@hudsonalpha.org), Benjamin T. Weaver (weaver@hudsonalpha.org), Richard M. Myers (rmyers@hudsonalpha.org), Gregory S. Barsh (gbarsh@hudsonalpha.org), E. Martina Bebin (ebebin@uab.edu), Gregory M. Cooper (gcooper@hudsonalpha.org)

*Corresponding Author: 601 Genome Way, Huntsville, AL 35806; 256-327-9490; gcooper@hudsonalpha.org

ABSTRACT

Background: Developmental disabilities have diverse genetic causes that must be identified to facilitate precise diagnoses. We describe genomic data from 371 affected individuals, 309 of which were sequenced as proband-parent trios. **Methods:** Whole exome sequences (WES) were generated for 365 individuals (127 affected) and whole genome sequences (WGS) were generated for 612 individuals (244 affected). **Results:** Pathogenic or likely pathogenic variants were found in 100 individuals (27%), with variants of uncertain significance in an additional 42 (11.3%). We found that a family history of neurological disease, especially the presence of an affected 1st degree relative, reduces the pathogenic/likely pathogenic variant identification rate, reflecting both the disease relevance and ease of interpretation of *de novo* variants. We also found that improvements to genetic knowledge facilitated interpretation changes in many cases. Through systematic reanalyses we have thus far reclassified 15 variants, with 11.3% of families who initially were found to harbor a VUS, and 4.7% of families with a negative result, eventually found to harbor a pathogenic or likely pathogenic variant. To further such progress, the data described here are being shared through ClinVar, GeneMatcher, and dbGAP.

Conclusion: Our data strongly support the value of large-scale sequencing, especially WGS within proband-parent trios, as both an effective first-choice diagnostic tool and means to advance clinical and research progress related to pediatric neurological disease.

KEYWORDS

Developmental Delay, Intellectual Disability, *De novo,* clinical sequencing, CSER

59 **BACKGROUND**

60 Developmental delay, intellectual disability, and related phenotypes (DD/ID) affect 1-2% of
61 children and pose medical, financial, and psychological challenges [1]. While many are genetic
62 in origin, a large fraction of cases are not diagnosed, with many families undergoing a
63 "diagnostic odyssey" involving numerous ineffective tests over many years. A lack of diagnoses
64 undermines counseling and medical management and slows research towards improving
65 educational or therapeutic options.

66

67 Standard clinical genetic testing for DD/ID includes karyotype, microarray, Fragile X, single
68 gene, gene panel, and/or mitochondrial DNA testing [2]. The first two tests examine an
69 individual's entire genome with low resolution, while the latter offer higher resolution but over
70 a small fraction of a person's genome. Whole exome or genome sequencing (WES or WGS) can
71 provide both broad and high-resolution identification of genetic variants, and hold great
72 promise as effective diagnostic assays [3].

73

74 As part of the Clinical Sequencing Exploratory Research (CSER) consortium [4], we have
75 sequenced 371 individuals with one or more DD/ID-related phenotypes. 100 affected
76 individuals (27%) were found to harbor a pathogenic or likely pathogenic (P/LP) variant, most of
77 which were *de novo*. 16% of P/LP variants were identified upon re-analysis that took place after
78 initial assessment and results return, supporting the value of systematic reanalysis of variant
79 data to maximize clinical effectiveness. We also describe 21 variants of uncertain significance
80 (VUS) in 19 genes not currently associated with disease but which are intriguing candidates. The

genomic data we generated and shared through dbGAP [5], ClinVar [6], and GeneMatcher [7] may prove useful to other clinical genetics labs and researchers. Our experiences and data strongly support the value of large-scale sequencing for clinical and research progress related to pediatric neurological disease.

METHODS

IRB approval and monitoring

Review boards at Western (20130675) and the University of Alabama at Birmingham (X130201001) approved and monitored this study.

Study participant population

Participants were enrolled at North Alabama Children's Specialists in Huntsville, AL. A parent or legal guardian was required to give consent for all probands, and assent was obtained from those probands who were capable. Probands were required to be at least two years of age, weigh at least 9 kilos (19.8 lbs), and be affected with developmental and/or intellectual delays; more detailed information regarding enrollment, including phenotypic criteria, is provided in the Supplemental Methods.

Whole exome and genome sequencing

Blood samples were sent for sequencing at the HudsonAlpha Genomic Services Laboratory (http://gsl.hudsonalpha.org). Genomic DNA was isolated from peripheral blood and WES (Nimblegen v3) or WGS was conducted to a mean depth of 71X or 35X, respectively, with >80% of bases covered at 20X. WES was conducted on Illumina HiSeq 2000 or 2500 machines; WGS was done on Illumina HiSeq Xs. Reads were aligned and variants called according to standard protocols [8, 9]. A robust relationship inference algorithm (KING) was used to confirm familial relationships [10].

125 **WGS CNV calling**

126 CNVs were called from WGS bam files using ERDS [11] and read Depth [12]. Overlapping calls
127 with at least 90% reciprocity, less than 50% segmental duplications, and that were observed in
128 five or fewer unaffected parents were retained and subsequently analyzed for potential disease
129 relevance. All CNVs found within 5 kb of a known DD/ID gene, within 5 kb of an OMIM disease-
130 associated gene [13], or intersecting one or more exons of any gene were subject to manual
131 curation.

132

133 **Filtering and reanalysis**

134 Using filters related to call quality, allele frequency, and impact predictions, we searched for
135 rare, damaging *de novo* variation, or inherited X-linked, recessive, or compound heterozygous
136 variation in affected probands, with modifications for probands with only one (duos) or neither
137 (singletons) biological parent available for sequencing.

138

139 Potential secondary variants (i.e., medically relevant but not associated with the proband's
140 DD/ID) were also sought within parents. We assessed variants in 56 genes flagged by the
141 American College of Medical Genetics and Genomics (ACMG) as potentially harboring medically
142 actionable, highly penetrant genetic variation [14], those associated with recessive disease in
143 OMIM [13], and carrier status for *CFTR*, *HBB,* and *HEXA*.

144

145 We also searched for all those variants listed as pathogenic or likely pathogenic in ClinVar [6],
146 regardless of inheritance or affected status. Further details for variant annotation and filtration

147 are supplied in Supplemental Methods.

148

149 For reanalysis, variants were reannotated with additional data, including updated versions of
150 ClinVar [6], ExAC [15], DDG2P [16], and gene or variant lists identified in publications related to
151 DD/ID genetics [17-19], and refiltered as described above and in Supplemental Methods.
152 Candidate variants found in genes that were either not known to associate with disease or were
153 found in individuals with phenotypes dissimilar from previously reported associations were
154 submitted to GeneMatcher (https://genematcher.org/) [7].

155

156 **Variant classification**

157 Variants were classified into one of five categories: pathogenic, likely pathogenic, VUS, likely
158 benign or benign. Our study began prior to publication of the formal classification system
159 proposed by the ACMG [20], although our evidence and interpretation criteria are conceptually
160 similar. Multiple lines of evidence, with mode of inheritance, allele frequency in population
161 databases, and quality of previously reported disease associations weighing most heavily, are
162 required to support assignments of pathogenicity. The Supplemental Methods contains a
163 detailed description of our assertion criteria, and these criteria are also available via ClinVar [6].
164 The key annotations, including mode of inheritance, allele frequencies, PubMed identifiers, and
165 computational inferences of variant effect, used to support the disease relevance of each
166 variant are supplied in Table S2.

167

168

169 **Variant validation**

170 WES and WGS were carried out under a research protocol and were not completed within a
171 CAP/CLIA laboratory. All variants found to be medically relevant and returnable were validated
172 by Sanger sequencing in an independent CLIA laboratory (Emory Genetics Laboratory) before
173 being returned to participants, although these validated variant results are not CLIA-compliant
174 as the input DNA was originally isolated in a research laboratory.

175

176 **Analysis of trios as singletons**

177 For probands subjected to WGS as part of trios, we removed parental genotype information
178 from their associated VCFs and subsequently filtered to identify variants that are expected to
179 be extremely rare in the general population and/or affect genes known to associate with
180 disease (Figure 2, Table S3). Scores from the Combined Annotation Dependent Depletion
181 (CADD) algorithm [21] were subsequently used to rank P/LP variants within the filtered variant
182 subset from each relevant proband. See Supplemental Methods for details.

183

184 **Functional Assays**

185 RNA isolation, cDNA synthesis, qPCR and western blotting were conducted according to
186 standard protocols. Details are provided in Supplemental Methods.

187

188

189

190

RESULTS

Demographics of study population

We enrolled 339 families (977 individuals total) with at least one proband with an unexplained diagnosis of a DD/ID-related phenotype (see "Study participant population" in Supplemental Methods). 284 participating families were enrolled with both biological parents. 261 of these families had one affected proband, while 21 families had two affected probands, and an additional two families had three affected probands. As each proband (including siblings within a family) was used to anchor a proband-parent "trio" as an analytical unit, our study includes a total of 309 trios from 284 families. We also enrolled 35 proband-parent "duos" that included one proband and one biological parent. Additionally, we enrolled two families with one biological parent and two affected probands (4 "duos"), and one duo family with three affected probands (3 "duos"), leading to a total of 42 "duos" from 38 families. Finally, we enrolled 17 "singleton" families in which no parents were available for testing; for 14 of these only one proband was tested, and in three families, two affected siblings were sequenced (total of 20 "singleton" probands).

During the course of this study, a decision to replace WES with WGS was made. In total, WES was performed on 365 individuals (127 affected) and WGS was performed on 612 individuals (244 affected). WES and WGS were sequenced to an average depth of 71X and 35X, respectively, with >80% of bases covered ≥ 20X in both experiment types. DNA from probands subjected to WES was also analyzed via a SNP array to detect copy-number variants (CNV) if clinical array testing had not been previously performed.

213 The study population had a mean age of 11 years and was 58% male. Affected individuals
214 displayed symptoms described by 333 unique HPO [22] terms with over 90% of individuals
215 displaying intellectual disability, 69% with speech delay, 45% with seizures, and 20% with
216 microcephaly or macrocephaly. 18% of affected individuals had an abnormal brain magnetic
217 resonance imaging (MRI) result and 81% of individuals had been subjected to genetic testing
218 prior to enrollment in this study (Table 1).
219
220 **DD/ID-associated genetic variation**
221 WES and WGS data were processed with standard protocols to produce variant lists in each
222 family that were subsequently annotated and filtered, and filtered variant lists were subject to
223 manual review (see Methods). KING, a robust relationship inference algorithm, was used to
224 confirm familial relationships [10]. Variant pathogenicity was classified based on allele
225 frequency, inheritance status, published reports, computational deleteriousness predictions,
226 and other sources of evidence; these assertion criteria are described in detail in the
227 Supplemental Methods. All variants described here were confirmed by Sanger sequencing (see
228 Methods) in probands and available family members before being returned to participants.
229
230 100 (27%) of the 371 probands had P/LP variants, while an additional 42 (11.3%) harbored a
231 VUS (Table 2). Given that most probands had been previously tested via microarray prior to
232 their enrollment in this study, large CNVs classified as a VUS or greater were detected in only 11
233 affected individuals (Table 2; Table S2; Figure S1).
234

235 Most (76%) P/LP variation occurred *de novo*, while 12% of individuals inherited P/LP variants as
236 compound heterozygotes or homozygotes (Figure S2A). An additional 5% were males with an X-
237 linked maternally inherited P/LP variant. Finally, 7% of participants who harbored a P/LP result
238 were sequenced with one or no biological parent and thus have unknown inheritance (Figure
239 S2A). Most P/LP variants were missense mutations (52%), while 39% were nonsense or
240 frameshift, 7% were predicted to disrupt splicing, and 2% led to inframe deletion (Figure S2B).
241 Variants that were classified as a VUS or greater were identified in 97 genes, excluding large
242 CNVs, with variants in 23 (24%) of these genes observed in two or more unrelated individuals
243 (Tables S1 and S2).
244
245 **Pathogenic/likely pathogenic variant rates across families of varying structure and phenotypic**
246 **complexity**
247 Affected individuals were categorized into one of three analytical structures based on the
248 number of parents that were sequenced along with the proband(s): proband-parent trios (309);
249 duos with one parent (42); and proband-only singletons (20). A P/LP result was found in 29.1%
250 of trio individuals, 19% of duo individuals, and 15% of singletons (Table 1).
251
252 We believe that at least some of the decline in P/LP variant yield in duos and singletons reflects
253 the analytical benefits of trio sequencing to efficiently highlight *de novo* variation. However,
254 given that one or both biological parents were unavailable or unwilling to participate in duo or
255 singleton analyses, the P/LP rate comparisons among trios/duos/singletons may be confounded
256 by other disease-associated factors (depression, schizophrenia, ADHD, etc.). For example, most

257 (11 of 20) of the singleton probands were adopted owing to death or disability associated with
258 neurological disease in their biological parents. To assess the relationship between
259 identification of a P/LP variant and family history, we separated all probands into three types:
260 simplex families in which there was only one affected proband and no 1st to 3rd degree relatives
261 reported to be affected with any neurological condition (n=93); families in which the enrolled
262 proband had no affected 1st degree relatives but with one or more reported 2nd or 3rd degree
263 relatives who were affected with a neurological condition (n=85); and multiplex families in
264 which the proband had at least one first degree relative affected with a neurological condition
265 (n=123) (Table S4). Thirty-eight probands with limited or no family history information were
266 excluded from this analysis.
267
268 P/LP variants were found in 24 (20%) of the 123 multiplex families (20 out of 97 trios), in
269 contrast with 35 (37.6%) of 93 simplex families (31 out of 80 trios), suggesting a P/LP
270 identification rate that is twice as high for simplex, relative to multiplex, families. While larger
271 sample sizes are needed to confirm this effect, the rate difference is significant whether or not
272 all enrolled families (p=0.002) or only those sequenced as trios (p=0.008) are considered. Rates
273 in families that were neither simplex nor multiplex (i.e., proband lacks an affected 1st degree
274 relative but has one or more affected 2nd or 3rd degree relatives) were intermediate, with 26%
275 of all such families having a P/LP result (28% of trios). Of relevance to the trio/duo/singleton
276 comparison described above, 11 of 13 (85%) singletons for which we had family history
277 information had an affected 1st degree relative, in contrast with 41% for duos and 39% for trios

278 (Table S4). This enrichment for affected 1st degree relatives likely contributed to the generally
279 reduced rate of P/LP variants in singletons observed here.
280
281 Multiplex family findings include examples of both expected and unexpected inheritance
282 patterns. For example, two affected male siblings were found to be hemizygous for a nonsense
283 mutation in *PHF6* (Börjeson-Forssman-Lehmann syndrome MIM:301900) inherited from their
284 unaffected mother. In another family, we found the proband to be compound heterozygous for
285 two variants in *GRIK4*, with one allele inherited from each parent. Interestingly, both the
286 mother and father of this proband report psychiatric illness, and extended family history of
287 psychiatric phenotypes is notable. While these data are insufficient to conclude that they are
288 indeed causative, it is plausible that the observed psychiatric phenotypes are at least partially
289 attributable to the variation in *GRIK4* found in this family. We also observed independent *de*
290 *novo* causal variants within two families. Affected siblings in family 00135 each harbored a
291 returnable *de novo* variant in a different gene, including a VUS in *SPR* (Dystonia MIM:612716)
292 and a pathogenic variant in *RIT1* (Noonan syndrome MIM:615355), while two probands (00075-
293 C and 00078-C) who were second degree relatives to one another harbored independent
294 pathogenic *de novo* variants, one each in *DDX3X* (X-linked ID MIM:300958) and *TCF20* (Table
295 S2).
296
297 **Alternative mechanisms of disease**
298 While the majority of DD/ID-associated genetic variation found here is predicted to lead to
299 missense, frameshift, or nonsense effects (Figure S2B), a subset of probands harbor variants

300 predicted to disrupt splicing, and in some cases, potentially alternative mechanisms of disease.
301 As an example, we sequenced an affected 14-year-old girl (00003-C, Table S2) who presented
302 with severe ID, seizures, speech delay, autism and stereotypic behaviors. WES revealed an SNV
303 within the splice acceptor site of intron 2 in *MECP2* (c.27-6C>G, MIM:312750), identical to a
304 previously observed *de novo* variant in a 5-year-old female with several features of Rett
305 syndrome, but who lacked deceleration of head growth and exhibited typical growth
306 development [23]. Laccone, et al. showed by qPCR that the variant produces a cryptic splice
307 acceptor site that adds five nucleotides to the mRNA resulting in a frameshift (p.R9fs24X) [23].
308 It is likely that both the canonical and cryptic splice sites function, allowing for most *MECP2*
309 transcripts to produce full-length protein, resulting in the milder Rett phenotype observed in
310 the individual described here and the girl described by Laccone and colleagues [23].
311
312 In another affected proband (00126-C), we identified compound heterozygous variants in *ALG1*
313 (Table S2). This proband has phenotypes consistent with ALG1-CDG (congenital disorder of
314 glycosylation MIM:608540) including severe ID, hypotonia, growth retardation, microcephaly,
315 and seizures, and was included as part of a comprehensive study of *ALG1*-associated
316 phenotypes [24]. The paternally inherited missense mutation (c.773C>T (p.S258L)) has been
317 previously reported as pathogenic [25], while the maternally inherited variant, which has not
318 been observed before (c.1187+3A>G), is three bases downstream of an exon/intron junction
319 (Figure 1A). We performed qPCR from patient blood RNA and found that intron 11 of *ALG1* is
320 completely retained in both the proband and the mother (Figure 1A-D). The retention of intron
321 11 results in a stop-gain after adding 84 nucleotides (28 codons).

322 In a separate family consisting of affected maternal half siblings (00218-C and 00218-S, Table
323 S2, Figure 1E) we found a variant in a canonical splice acceptor site (c.505-2A>G) of *MTOR*
324 intron 4. The half siblings described here both have ID; the younger sibling has no seizures but
325 has facial dysmorphism, speech delay, and autism, while his older sister exhibits seizures. We
326 presume that the maternal half siblings inherited the splice variant from their mother, for
327 whom DNA was not available, who was reported to exhibit seizures. We conducted qPCR and
328 Sanger sequencing using blood-derived RNA from both siblings, finding transcripts that included
329 an additional 134 nucleotides from the 3' end of intron 4, ultimately leading to the addition of
330 20 amino acids before a stop-gain (Figure 1F-H, Figure S3). Because the stop-gain occurs early in
331 protein translation, this splice variant likely leads to *MTOR* loss-of-function. Mutations in *MTOR*
332 associate with a broad spectrum of phenotypes including epilepsy, hemimegalencephaly, and
333 intellectual disability [26]. However, previously reported pathogenic variants in *MTOR* are all
334 missense and suspected to result in gain-of-function [27]. Owing to this mechanistic
335 uncertainty, we have classified this splice variant as a VUS. However, given the overlap between
336 phenotypes observed in this family and previously reported families, we find this variant to be
337 highly intriguing and suggestive that *MTOR* loss-of-function variation may also lead to disease.
338 *MTOR* is highly intolerant of mutations in the general population (RVIS [28] score of 0.09%)
339 supporting the hypothesis that loss-of-function is deleterious and likely leads to disease
340 consequences.
341
342
343

344 **Proband-only versus trio sequencing**

345 Our trio-based study design allows rapid identification of *de novo* variants, which are enriched
346 among variants that are causally related to deleterious, pediatric phenotypes [29]. However, we
347 also assessed to what extent our P/LP rate would differ if we had only enrolled probands. Thus,
348 and to avoid the confounding of family history differences among trios, duos, and singletons
349 (see above), we subjected variants found by WGS within all trio-based probands to various
350 filtering scenarios blinded to parental status and assessed the CADD score [21] ranks of *de novo*
351 variants previously classified as P/LP (Figure 2; Table S3). While parentally informed filters were
352 the most sensitive and efficient (e.g., >60% of P/LP variants were the top-ranked variant among
353 the list of all *de novo* events in each respective proband), filters defined without parental
354 information were also effective. For example, among all rare, protein-altering (i.e., missense,
355 nonsense, frameshift, or canonical splice-site) mutations found in genes associated with
356 Mendelian disease via OMIM [13] or associated with DD/ID via DECIPHER [16], 20% of P/LP
357 variants were the top-ranked variant in the given proband, most ranked among the top 5, and
358 >80% ranked among the top 25. These data suggest that most P/LP variants could be found
359 within probands analyzed without parental information, although additional curation time,
360 likely in proportion to the drops in P/LP variant rank within any given filtered subset, would be
361 required (Table S3).

362

363 In contrast to P/LP variants, VUSs would have been more difficult to identify without parental
364 sequencing (Figure S4), owing to the fact that many VUSs do not affect genes known to
365 associate with disease. Also, those VUSs that do affect genes known to associate with disease

366 tended to have lesser computationally estimated effects, and therefore lower CADD ranks [21];
367 if they were more overtly deleterious, they would likely have been classified as P/LP. Discovery
368 of candidate or novel disease associations, many of which are likely to eventually be shown as
369 robust, is thus substantially more effective within trios.
370

371 **Secondary findings in participating parents**

372 We found genetic variation unrelated to DD/ID, i.e., secondary findings, in 8.7% of parents
373 (Table S5). One and a half percent of parents were found to harbor a P/LP variant related to a
374 self-reported secondary condition, such as variants in *SLC22A5* that underlie a primary carnitine
375 deficiency (MIM:212140). We also examined 56 genes identified by the ACMG as potentially
376 harboring actionable secondary findings [14], revealing P/LP variants in 12 parents (2.0%), a
377 rate similar to that observed in other cohorts [14, 30]. Finally, we performed a limited carrier
378 screening assessment, identifying 28 (4.6%) parents as carriers of P/LP variation in *HBB* (Sickle
379 cell anemia MIM:603903), *HEXA* (Tay-Sachs disease MIM:272800), or *CFTR* (Cystic fibrosis
380 MIM:219700). We also assessed parents as mate pairs and searched for genes in which both
381 are heterozygous for a P/LP recessive allele. These analyses yielded one parental pair (among
382 285 total) as carriers for variants in *ATP7B*, associated with Wilson disease (MIM:277900).
383

384 **Reanalysis of WES and WGS data**

385 To exploit steady increases in human genetic knowledge, we performed systematic reanalyses
386 of WES/WGS data. We approached reanalysis in three ways: 1) systematic reanalysis of old
387 data, with the goal of reassessing each dataset every 12 months after initial analysis; 2) mining

388 of variant prompted by new DD/ID genetic publications; and 3) use of GeneMatcher [7] to aid in
389 the interpretation of variants in genes of uncertain disease significance.
390
391 As shown in Table 3, these efforts led to an increase in pathogenicity score for 15 variants in 17
392 individuals. In nine cases, a new publication became available that allowed a variant that had
393 not been previously reported, or that was previously reported as a VUS, to be reclassified as
394 P/LP. Three additional changes were a result of discussions facilitated by GeneMatcher [7],
395 while the remaining upgrades resulted from reductions in filter stringency (changes to read
396 depth and batch allele frequency) or clarification of the clinical phenotype. Among all 44 VUSs
397 thus far identified, five (11.3%) have been upgraded. The most rapid change affected a *de novo*
398 variant in *DDX3X*, which was upgraded from VUS to pathogenic approximately one month after
399 initial assessment, while a *de novo* disruption of *EBF3* was upgraded from VUS to pathogenic
400 approximately 2.5 years after initial assessment. VUSs associated with DD/ID, especially when
401 identified via parent-proband trio sequencing, thus have considerable potential for upgrade.
402 Additionally, of the 211 families who originally received a negative result, P/LP variation was
403 identified for 10 (4.7%) through reanalysis. These data show that regular reanalysis of both
404 uncertain and negative results is an effective mechanism to improve diagnostic yield.
405
406 **Identification of novel candidate genes**
407 We have identified 21 variants within 19 genes with no known disease association but which
408 are interesting candidates. For example, in one proband (00265-C) we identified an early
409 nonsense variant (c.2140C>T (p.R714X), CADD score 44) in *ROCK2*, with reduction of ROCK2

410 protein confirmed by western blot (Figure S5). ROCK2 is a conserved Rho-associated
411 serine/threonine kinase involved in a number of cellular processes including actin cytoskeleton
412 organization, proliferation, apoptosis, extracellular matrix remodeling and smooth muscle cell
413 contraction, and has an RVIS [28]score placing it among the top 17.93% most intolerant genes
414 [31]. As a second example, in two unrelated probands (00310-C and 00030-C), we identified *de*
415 *novo* variation in *NBEA*, a nonsense variant at codon 2213 (of 2946, c.6637C>T (p.R2213X),
416 CADD score 52), and a missense at codon 946 (c.2836C>T (p.H946Y), CADD score 25.6). *NBEA* is
417 a kinase anchoring protein with roles in the recruitment of cAMP dependent protein kinase A to
418 endomembranes near the trans-Golgi network [32]. The RVIS score [28] of *NBEA* is 0.75%.
419 While these variants remain VUSs, the fact that they are *de novo*, predicted to be deleterious,
420 and affect genes under strong selective conservation in human populations, suggests they have
421 a good chance to be disease-associated.
422
423 **DISCUSSION**
424 We have sequenced 371 individuals with various DD/ID-related phenotypes. 27% percent of
425 these individuals harbored a P/LP variant, most of which were *de novo* and protein-altering. We
426 found that the P/LP yield is impacted by presence of neurological disease in family members, as
427 our success rate drops from 38% for probands without any affected relatives to 19.5% for
428 probands with one or more affected 1st degree relatives. These data are consistent with the
429 observation of higher causal variant yields in simplex families relative to multiplex families
430 affected with autism [33]. It in part reflects the eased interpretation of *de novo* causal variation
431 relative to inherited, and likely in many cases variably expressive or incompletely penetrant,

432 causal variation (e.g., 16p12) [34].

433

434 127 probands were subject to WES and 244 were subject to WGS. The P/LP identification rate
435 was not significantly different between the two assays when considering only SNVs or small
436 indels (p=0.30). However, WGS is a better assay for detection of CNVs [35] and, while our
437 patient population is depleted for large causal CNVs owing to prior array or karyotype testing,
438 we have identified CNVs that we classified as P/LP in eight individuals.

439

440 We have also demonstrated the value of systematic reanalysis, which has thus far yielded P/LP
441 variants for an additional 17 individuals (17% of total P/LP variants, 4.6% of total probands).
442 Given the rates of progress in Mendelian disease genetics [36] and the development of new
443 genomic annotations, we believe that systematic reanalysis of genomic data should become
444 standard practice. While the costs and logistical demands for implementation at large scales are
445 unclear, re-analysis has the potential to considerably increase diagnostic yields over time (e.g.,
446 in our study, ~8% for cases > 1 year removed from initial analysis). Furthermore, as more
447 pathogenic coding and non-coding variants are found, the reanalysis benefit potential is largest
448 for WGS relative to WES; the former typically has slightly better coverage of coding exons in
449 both our data (Table S6) and previous studies [35], and re-analysis of pathogenic non-coding
450 variation is impossible with WES.

451

452 Our data clearly suggests trio-based sequencing as more sensitive and analytically efficient than
453 proband-only sequencing, supporting the value of trios in clinical diagnostics; as sequencing

454 costs continue to drop, testing parents should eventually be offered routinely. Further, VUSs
455 and novel candidates are more difficult to identify without parental sequence data, and
456 proband-only approaches will ultimately confer less benefit in terms of discovery of new
457 disease associations. However, current sequencing costs, when coupled to overall priorities
458 (e.g., per-patient yield vs. total number of diagnoses) may lead to variability in decision-making
459 about how to best allocate resources. For example, tripling per-patient sequencing costs will,
460 under many realistic cost scenarios, lead to fewer total diagnoses within a given total budget
461 even though the per-patient diagnostic yield is higher and curation time reduced for trios
462 relative to singletons. Our retrospective analyses, in which we evaluated ranks of pathogenic
463 variants under various filtering parameters, may provide useful information in making these
464 decisions. Trade-offs in curation time, which will correlate with P/LP variant ranks, and
465 sensitivity can be estimated empirically, in relative terms, using these data (Figure 2; Table S3).
466
467 Variation detected through our studies has already helped lead to the discovery of at least one
468 new disease association, as we identified two patients that harbor *de novo* variants in *EBF3*, a
469 highly conserved transcription factor involved in neurodevelopment that is relatively intolerant
470 to mutations in the general population (RVIS [28]: 6.78%). Through collaboration with other
471 researchers via GeneMatcher [7], we were able to identify a total of 10 DD/ID-affected
472 individuals who harbor *EBF3* variants, supporting that *de novo* disruption of *EBF3* function leads
473 to neurodevelopmental phenotypes [37]. It is our hope that the other VUSs described here,
474 shared via ClinVar [6] and GeneMatcher [7], will also help to facilitate new associations.
475

476 **CONCLUSIONS**

477 We have demonstrated the benefits of genomic sequencing to identify disease-associated
478 variation in probands with developmental disabilities who are otherwise lacking a precise
479 clinical diagnosis. Indeed, by combining genomic breadth with resolution capable of detecting
480 SNVs, indels, and CNVs in a single assay, WGS is a highly effective choice as the first diagnostic
481 test, rather than last resort, for unexplained developmental disabilities. The ability for WGS to
482 serve as a single-assay replacement for WES and microarrays underscores its value as a
483 frontline test. Furthermore, the benefits and effectiveness of WGS testing is likely to grow over
484 time both by accelerating research (for example into the discovery of smaller pathogenic CNVs
485 and pathogenic SNVs outside of coding exons), and by facilitating more effective reanalysis, a
486 process which we show to be an essential component to maximize diagnostic yield.

487

488 **ABBREVIATIONS**

489 WGS, whole genome sequencing; WES, whole exome sequencing; CSER, Clinical Sequencing
490 Exploratory Research; CNV, copy number variant; DD/ID, Developmental delay/Intellectual
491 disability; P/LP, Pathogenic/likely pathogenic; VUS, variant of uncertain significance

492

493 **DECLARATIONS**

494 *Ethics approval and consent to participate:* Review boards at Western (20130675) and the
495 University of Alabama at Birmingham (X130201001) approved and monitored this study.
496 *Consent for publication:* A parent or legal guardian was required to give consent to participate
497 in the study and inclusion of their data for publication, and assent was obtained from those

498 children who were capable. *Availability of data and material:* The genomic data generated for
499 this work is available through dbGAP [5], ClinVar [6], and GeneMatcher [7]. *Competing*
500 *interests:* The authors declare that they have no competing interests. *Funding:* This work was
501 supported by grants from the US National Human Genome Research Institute (NHGRI;
502 UM1HG007301) and the National Cancer Institute (NCI; R01CA197139). *Author contributions:*
503 GMC, RMM, GSB, NEL and KBB designed the study and guided its implementation. KMB, MDA,
504 CRF, SMH, and MLT performed genomic analyses to identify disease-linked variation, prepared
505 variant reports, and oversaw variant review. DEG called CNVs from WGS data. DEG, BTW and
506 JSW contributed to bioinformatics analyses. ASN contributed to data acquisition. WVK, KME, SS,
507 EJL, and EMB recruited study participants, collected blood samples and clinical information, and
508 returned results. KBB, CAR, GSB aided in variant review. KLE and JNC performed functional
509 validation studies. KMB, MDA, CRF, SMH, MLT, and GMC wrote the manuscript. All authors
510 contributed to and approved the manuscript.
511
512 *Acknowledgements:* We are grateful to the patients and their families who contributed to this
513 study. We thank the HudsonAlpha Software Development and Informatics team and the
514 Genome Sequencing Center who contributed to data acquisition and analysis. We would also
515 like to thank Dr. Jeremy Herskowitz in the Department of Neurology at University of Alabama at
516 Birmingham for discussions about ROCK2.
517
518
519

520 **REFERENCES**

521 1. Boyle CA, Boulet S, Schieve LA, Cohen RA, Blumberg SJ, Yeargin-Allsopp M, et al. Trends
522 in the prevalence of developmental disabilities in US children, 1997-2008. Pediatrics.
523 2011;127:1034-42.

524 2. Sun F, Oristaglio J, Levy SE, Hakonarson H, Sullivan N, Fontanarosa J, et al. Genetic
525 Testing for Developmental Disabilities, Intellectual Disability, and Autism Spectrum
526 Disorder. Rockville (MD): Agency for Healthcare Research and Quality (US). 2015 Jun
527 (Technical Briefs, No. 23) Available from:
528 https://www.ncbi.nlm.nih.gov/books/NBK304462/.

529 3. Stavropoulos D, Merico D, Jobling R, Bowdin S. Whole-genome sequencing expands
530 diagnostic utility and improves clinical management in paediatric medicine. npj Genomic
531 Medicine. 2016;1:1-9.

532 4. Green RC, Goddard KA, Jarvik GP, Amendola LM, Appelbaum PS, Berg JS, et al. Clinical
533 Sequencing Exploratory Research Consortium: Accelerating Evidence-Based Practice of
534 Genomic Medicine. Am J Hum Genet. 2016;98:1051-66.

535 5. Mailman MD, Feolo M, Jin Y, Kimura M, Tryka K, Bagoutdinov R, et al. The NCBI dbGaP
536 database of genotypes and phenotypes. Nat Genet. 2007;39:1181-6.

537 6. Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S, et al. ClinVar: public
538 archive of interpretations of clinically relevant variants. Nucleic Acids Res.
539 2016;44:D862-8.

540 7. Sobreira N, Schiettecatte F, Valle D, Hamosh A. GeneMatcher: a matching tool for
541 connecting investigators with an interest in the same gene. Hum Mutat. 2015;36:928-
542 30.
543 8. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform.
544 Bioinformatics. 2009;25:1754-60.
545 9. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for
546 variation discovery and genotyping using next-generation DNA sequencing data. Nat
547 Genet. 2011;43:491-8.
548 10. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship
549 inference in genome-wide association studies. Bioinformatics. 2010;26:2867-73.
550 11. Zhu M, Need AC, Han Y, Ge D, Maia JM, Zhu Q, et al. Using ERDS to infer copy-number
551 variants in high-coverage genomes. Am J Hum Genet. 2012;91:408-21.
552 12. Miller CA, Hampton O, Coarfa C, and Milosavljevic A. ReadDepth: a parallel R package
553 for detecting copy number alterations from short sequencing reads. PLoS One.
554 2011;6:e16327.
555 13. Hamosh A, Scott AF, Amberger JS, Bocchini CA, and McKusick VA. Online Mendelian
556 Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders.
557 Nucleic Acids Res. 2005;33:D514-7.
558 14. Green RC, Berg JS, Grody WW, Kalia SS, Korf BR, Martin CL, et al. ACMG
559 recommendations for reporting of incidental findings in clinical exome and genome
560 sequencing. Genet Med. 2013;15:565-74.

561 15. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of
562 protein-coding genetic variation in 60,706 humans. Nature. 2016;536:285-91.
563 16. Deciphering Developmental Disorders Study. Prevalence and architecture of de novo
564 mutations in developmental disorders. Nature. 2017; doi:10.1038/nature21062.
565 17. Epi4K Consirum, Epilepsy Phenome/Genome Project, Allen AS, Berkovic SF, Cossette P,
566 Delanty N, et al. De novo mutations in epileptic encephalopathies. Nature.
567 2013;501:217-21.
568 18. Karaca E, Harel T, Pehlivan D, Jhangiani SN, Gambin T, Coban Akdemir Z, et al. Genes
569 that Affect Brain Structure and Function Identified by Rare Variant Analyses of
570 Mendelian Neurologic Disease. Neuron. 2015;88:499-513.
571 19. Samocha KE, Robinson EB, Sanders SJ, Stevens C, Sabo A, McGrath LM, et al. A
572 framework for the interpretation of de novo mutation in human disease. Nat Genet.
573 2014;46:944-50.
574 20. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines
575 for the interpretation of sequence variants: a joint consensus recommendation of the
576 American College of Medical Genetics and Genomics and the Association for Molecular
577 Pathology. Genet Med. 2015;17:405-24.
578 21. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework
579 for estimating the relative pathogenicity of human genetic variants. Nat Genet.
580 2014;46:310-5.

581 22. Kohler S, Doelken SC, Mungall CJ, Bauer S, Firth HV, Bailleul-Forestier I, et al. The Human
582 Phenotype Ontology project: linking molecular biology and disease through phenotype
583 data. Nucleic Acids Res. 2014;42:D966-74.
584 23. Laccone F, Huppke P, Hanefeld F, Meins M. Mutation spectrum in patients with Rett
585 syndrome in the German population: Evidence of hot spot regions. Hum Mutat.
586 2001;17:183-90.
587 24. Ng BG, Shiryaev SA, Rymen D, Eklund EA, Raymond K, Kircher M, et al. ALG1-CDG:
588 Clinical and Molecular Characterization of 39 Unreported Patients. Hum Mutat.
589 2016;37:653-60.
590 25. Dupre T, Vuillaumier-Barrot S, Chantret I, Sadou Yaye H, Le Bizec C, Afenjar A, et al.
591 Guanosine diphosphate-mannose:GlcNAc2-PP-dolichol mannosyltransferase deficiency
592 (congenital disorders of glycosylation type Ik): five new patients and seven novel
593 mutations. J Med Genet. 2010;47:729-35.
594 26. Baulac S. mTOR signaling pathway genes in focal epilepsies. Prog Brain Res.
595 2016;226:61-79.
596 27. Moller R, Wechkuysen S, Chipaux M, Marsan E, Taly V, Bebin ME, et al. Germline and
597 somatic mutations in MTOR gene in focal cortical dysplasia and epilepsy. Neurology:
598 Genetics. In press;2:e118.
599 28. Petrovski S, Wang Q, Heinzen EL, Allen AS, Goldstein DB. Genic intolerance to functional
600 variation and the interpretation of personal genomes. PLoS Genet. 2013;9:e1003709.
601 29. Vissers LE, de Ligt J, Gilissen C, Janssen I, Steehouwer M, de Vries P, et al. A de novo
602 paradigm for mental retardation. Nat Genet. 2010;42:1109-12.

603 30. Johnston JJ, Rubinstein WS, Facio FM, Ng D, Singh LN, Teer JK, et al. Secondary variants
604 in individuals undergoing exome sequencing: screening of 572 individuals identifies
605 high-penetrance mutations in cancer-susceptibility genes. Am J Hum Genet. 2012;91:97-
606 108.
607 31. Loirand G. Rho Kinases in Health and Disease: From Basic Science to Translational
608 Research. Pharmacol Rev. 2015;67:1074-95.
609 32. Castermans D, Wilquet V, Parthoens E, Huysmans C, Steyaert J, Swinnen L, et al. The
610 neurobeachin gene is disrupted by a translocation in a patient with idiopathic autism. J
611 Med Genet. 2003;40:352-6.
612 33. Klei L, Sanders SJ, Murtha MT, Hus V, Lowe JK, Willsey AJ, et al. Common genetic
613 variants, acting additively, are a major source of risk for autism. Mol Autism. 2012;3:9.
614 34. Quintans B, Ordonez-Ugalde A, Cacheiro P, Carracedo A, Sobrido MJ. Medical genomics:
615 The intricate path from genetic variant identification to clinical interpretation. Appl
616 Transl Genom. 2014;3:60-7.
617 35. Belkadi A, Bolze A, Itan Y, Cobat A, Vincent QB, Antipenko A, et al. Whole-genome
618 sequencing is more powerful than whole-exome sequencing for detecting exome
619 variants. Proc Natl Acad Sci U S A. 2015;112:5473-8.
620 36. Chong JX, Buckingham KJ, Jhangiani SN, Boehm C, Sobreira N, Smith JD, et al. The
621 Genetic Basis of Mendelian Phenotypes: Discoveries, Challenges, and Opportunities. Am
622 J Hum Genet. 2015;97:199-215.

623 37. Harms FL, K.M. G, Hardigan AA, Kortum F, Shukla A, Alawi M, et al. Mutations in EBF3
624 disturb transcriptional profiles and underlie a novel syndrome of intellectual disability,
625 ataxia and facial dysmorphism. BioRxiv. 2016; https://doi.org/10.1101/067454.
626 38. Deciphering Developmental Disorders Study. Large-scale discovery of novel genetic
627 causes of developmental disorders. Nature. 2015;519:223-8.
628 39. Farach LS, Northrup H. KIAA2022 nonsense mutation in a symptomatic female. Am J
629 Med Genet A. 2016;170:703-6.
630 40. Shang L, Cho MT, Retterer K, Folk L, Humberson J, Rohena L, et al. Mutations in ARID2
631 are associated with intellectual disabilities. Neurogenetics. 2015;16:307-14.
632 41. Wortmann SB, Zietkiewicz S, Kousi M, Szklarczyk R, Haack TB, Gersting SW, et al. CLPB
633 mutations cause 3-methylglutaconic aciduria, progressive brain atrophy, intellectual
634 disability, congenital neutropenia, cataracts, movement disorder. Am J Hum Genet.
635 2015;96:245-57.
636 42. Siekierska A, Isrie M, Liu Y, Scheldeman C, Vanthillo N, Lagae L, et al. Gain-of-function
637 FHF1 mutation causes early-onset epileptic encephalopathy with cerebellar atrophy.
638 Neurology. 2016;86:2162-70.
639 43. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, et al. Reference
640 sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and
641 functional annotation. Nucleic Acids Res. 2016;44:D733-45.
642 44. Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A
643 global reference for human genetic variation. Nature. 2015;526:68-74.

45. *Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP)*. [cited 2013; Available from: http://evs.gs.washington.edu/EVS.

666 **FIGURE AND TABLE LEGENDS**

667 **Figure 1: Intronic variants in *ALG1* and *MTOR* disrupt splicing and introduce early stop codons.**
668 (A) Diagram showing the region of *ALG1* surrounding the variant found in the proband and
669 mother, an A>G transition 3 nucleotides downstream from the splicing donor site of intron 11.
670 E=exon. (B) PCR F and PCR R indicate the position of the oligos used to amplify the region from
671 patient derived cDNA. The control sample is from the RNA extracted from blood of an unrelated
672 individual that did not harbor the variant. Control reactions lacking RT were also performed and
673 did not show the PCR product containing the fully retained intron (data not shown). (C and D)
674 qPCR analysis shows that the variant leads to inclusion of the entire intron 11. Controls are two
675 unrelated individuals and the father of the proband. The affected individuals are the proband
676 and mother. (E) Diagram showing the region of *MTOR* surrounding the variant, an A>G
677 transition 2 nucleotides upstream of the splicing acceptor site. E=exon. (F) The region
678 surrounding intron 4 was amplified using PCR F and PCR R (position indicated in E), and shows
679 partial retention of the intron. The retained partial intron was not detected in control reactions
680 lacking RT (data not shown). (G and H) qPCR from blood RNA shows that the 5' splice site is not
681 affected by the variant, but that the 3' acceptor site is, leading to partial retention (134bp) of
682 intron 4. Controls included unrelated individuals and the maternal half aunt of the proband.
683 Affected individuals are the proband and half-sibling. For all qPCR analyses RNA was extracted
684 from blood and $\Delta\Delta C_T$ values were calculated as a percent of affected individuals and normalized
685 to *GAPDH*. The sequences of all oligos used are found in Table S7.

686

687 **Figure 2: Ranks of pathogenic/likely pathogenic variants filtered without parental data**

688 **relative to trio-defined *de novo* events.** Most pathogenic/likely pathogenic variants, even

689 under models that only consider population frequencies (e.g., "Rare"), rank (based on CADD)

690 among the top 25 hits in a patient, and many rank as the top hit. Restrictions to rare coding

691 variants and/or those affecting OMIM/DDG2P [13, 16] genes further enrich for causal variants

692 among top candidates, making diagnosis feasible without parents.

693

694 **Table 1: Pathogenic/Likely pathogenic rates by clinical annotation and family structure among**

695 **the 371 DD/ID-affected individuals**

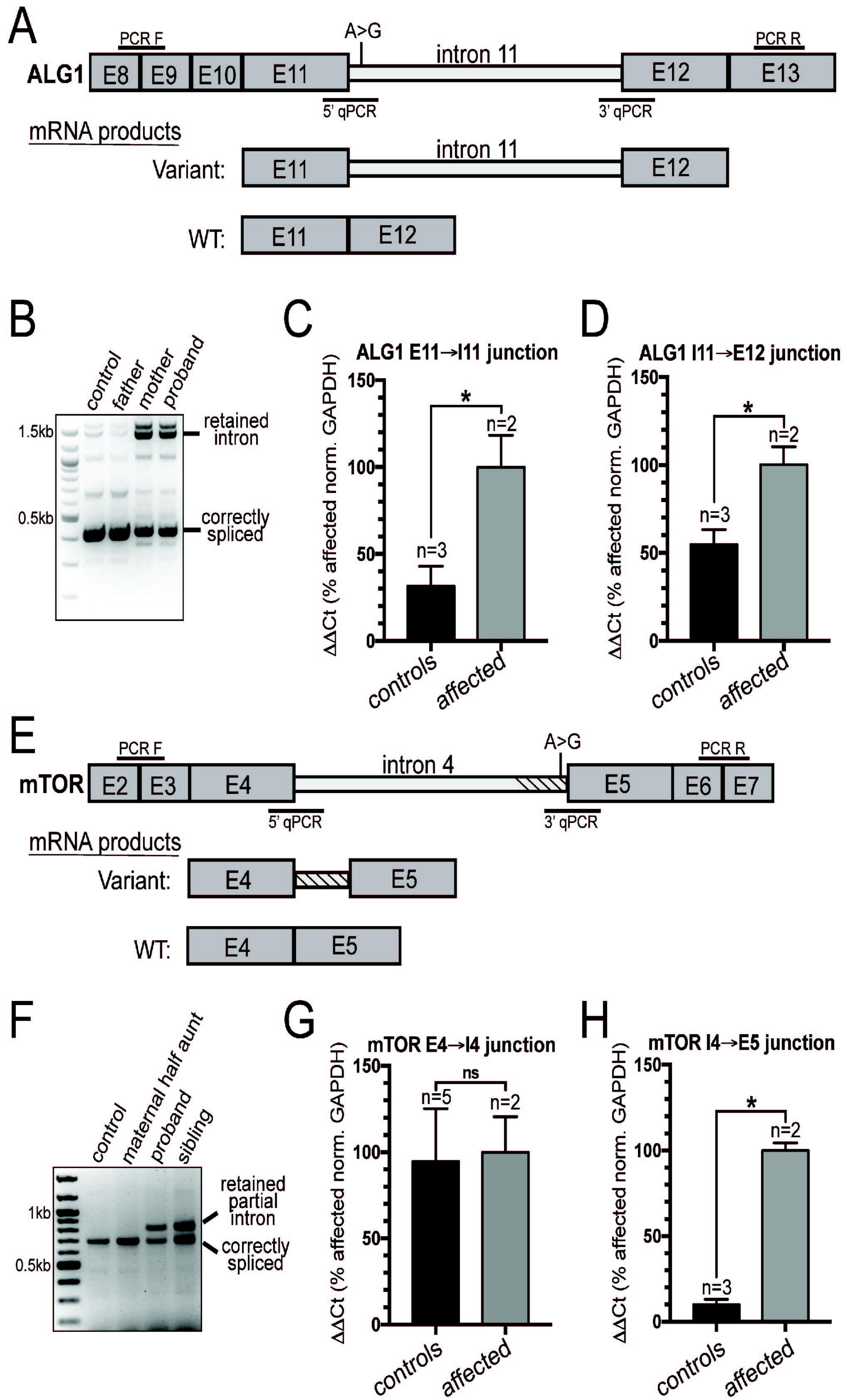
696

697 **Table 2: Results of exome and/or genome sequencing for 371 DD/ID-affected individuals**

698

699 **Table 3: Variants with an increase in pathogenicity score due to reanalysis.**

700


A
PCR F
A>G
PCR R
ALG1
E8
E9
E10
E11
intron 11
E12
E13
5' qPCR
3' qPCR
mRNA products
Variant:
E11
intron 11
E12
WT:
E11
E12
B
control
father
mother
proband
1.5kb
0.5kb
retained intron
correctly spliced
C
ALG1 E11→I11 junction
ΔΔCt (% affected norm. GAPDH)
150
100
50
0
*
n=2
n=3
controls
affected
D
ALG1 I11→E12 junction
ΔΔCt (% affected norm. GAPDH)
150
100
50
0
*
n=2
n=3
controls
affected
E
PCR F
A>G
PCR R
mTOR
E2
E3
E4
intron 4
E5
E6
E7
5' qPCR
3' qPCR
mRNA products
Variant:
E4
E5
WT:
E4
E5
F
control
maternal half aunt
proband
sibling
1kb
0.5kb
retained partial intron
correctly spliced
G
mTOR E4→I4 junction
ΔΔCt (% affected norm. GAPDH)
150
100
50
0
ns
n=5
n=2
controls
affected
H
mTOR I4→E5 junction
ΔΔCt (% affected norm. GAPDH)
150
100
50
0
*
n=2
n=3
controls
affected

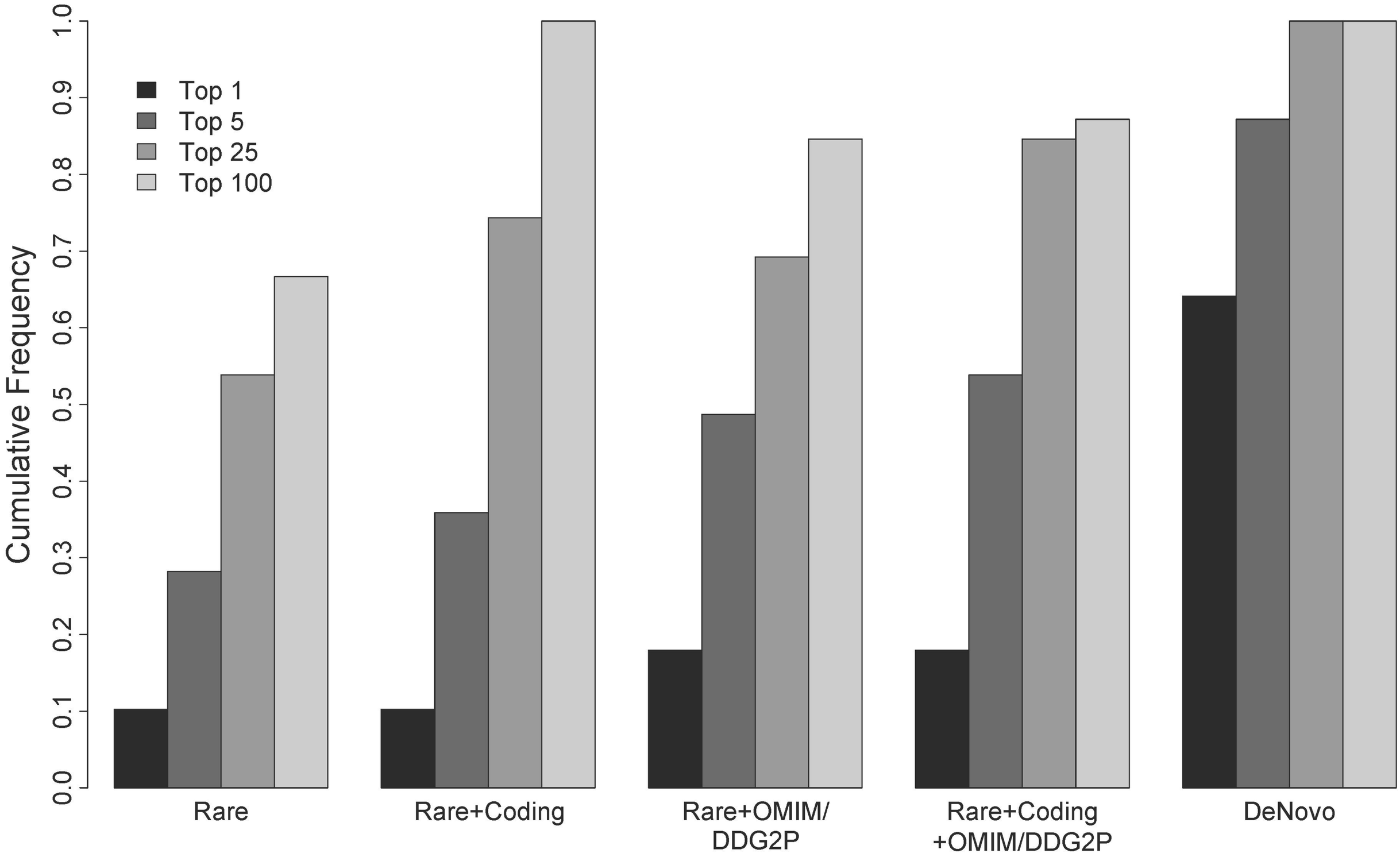

Cumulative Frequency
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
Top 1
Top 5
Top 25
Top 100
Rare
Rare+Coding
Rare+OMIM/
DDG2P
Rare+Coding
+OMIM/DDG2P
DeNovo

Table 1. Pathogenic/Likely pathogenic rates by clinical annotation and family structure among the 371 DD/ID-affected individuals

CHARACTERISTIC	% Individuals (No. of individuals)	% individuals with P/LP result (No. of individuals)
AGE OF INDIVIDUAL		
2-5 years	25.8% (96)	27.1% (26/96)
6-12 years	44.5% (165)	25.4% (42/165)
13-18 years	16.5% (61)	32.8% (20/61)
19-40 years	12.7% (47)	32.0% (15/47)
>40 years	0.54% (2)	0.00% (0/2)
Average age of subject (age range)	10.56 (2 to 54 years)	
SEX OF INDIVIDUAL		
Male	57.7% (214)	24.3% (52/214)
Female	42.3% (157)	32.5% (51/157)
CLINICAL SPECIFICS		
Intellectual disability (mild, moderate, severe) (HP:0001256, HP:0002342, HP:0010864)	92.7% (344)	28.5% (98/344)
Speech delay (HP:0000750)	68.7% (255)	27.1% (69/255)
Seizures (HP:0001250)	45.3% (168)	30.9% (52/168)
Facial dysmorphism (HP:0001999)	30.2% (112)	29.5% (33/112)
Autism spectrum disorder (HP:000729)	25.6% (95)	18.9% (18/95)
Hypotonia (HP:0001252)	20.2% (75)	34.6% (26/75)
Positive Brain MRI	17.5% (65)	28.1% (18/64)
Macrocephaly (HP:0000256)	9.70% (36)	25.0% (9/36)
Microcephaly (HP:0000252)	9.16% (34)	47.0% (16/34)
ADHD (HP:0007018)	7.28% (27)	25.9% (7/27)
Failure to thrive (HP:0001508)	5.90% (22)	27.3% (6/22)
Short stature (HP:0004322)	4.85% (18)	44.4% (8/18)
PREVIOUS GENETIC TESTING		
Microarray	59.8% (222)	27.5% (61/222)
Single Gene/Gene Panel	38.3% (142)	30.3% (43/142)
Karyotype	29.1% (108)	36.1% (39/108)
Fragile-X	27.2% (101)	27.7% (28/101)
Mito DNA Screen	7.55% (28)	25.0% (7/28)
FAMILY STRUCTURE		
Trio	83.3% (309)	29.1% (90/309)
Duo	11.3% (42)	19.0% (8/42)
Singleton	5.4% (20)	15.0% (3/20)

MRI, magnetic resonance imaging

Table 2. Results of exome and/or genome sequencing for 371 DD/ID-affected individuals

	SNV/indel			CNV		
Assay (Affected individuals)	**Pathogenic**	**Likely pathogenic**	**VUS**	**Pathogenic**	**Likely pathogenic**	**VUS**
Exome (127)	20.4% (26)	9.4% (12)	11.0% (14)	1.6% (2)*	0% (0)	0% (0)
Genome (244)	18.0% (44)	4.1% (10)	10.2% (25)	2.0% (5)	0.4% (1)	1.2% (3)
Exome and genome (Total individuals: 371)	18.9% (70)	5.9% (22)	10.5% (39)	1.9% (7)	0.3% (1)	0.8% (3)

* Identified by microarray

Table 3: Variants with an increase in pathogenicity score due to reanalysis

Gene	Affected Individual ID(s)	Variant Info	Original Score	Updated Score	Reason(s) for Update	Evidence for upgrade
DDX3X	00075-C	NM_001356.4(DDX3X):c.745G>T (p.Glu249Ter)	VUS	Pathogenic	Publication	[37]
EBF3	00006-C	NM_001005463.2(EBF3):c.1101+1G>T	VUS	Pathogenic	GeneMatcher	Collaboration with several other groups identified patients with comparable genotypes and phenotypes
EBF3	00032-C	NM_001005463.2(EBF3):c.530C>T (p.Pro177Leu)	VUS	Pathogenic	GeneMatcher	Collaboration with several other groups identified patients with comparable genotypes and phenotypes
KIAA2022	00082-C	NM_001008537.2(KIAA2022):c.2999_3000delCT (p.Ser1000Cysfs)	VUS	Pathogenic	Publication/Personal Communication	[38]
TCF20	00078-C	NM_005650.3(TCF20):c.5385_5386delTG (p.Cys1795Trpfs)	VUS	Pathogenic	Publication	[16]
ARID2	00026-C	NM_152641.2(ARID2):c.1708delT (p.Cys570Valfs)	NR	Pathogenic	Publication	[39]
CDK13	00253-C	NM_003718.4(CDK13):c.2525A>G (p.Asn842Ser)	NR	Pathogenic	Publication	[16]
CLPB	00127-C	NM_030813.5(CLPB):c.1222A>G (p.Arg408Gly) NM_030813.5(CLPB):c.1249C>T (p.Arg417Ter)	NR	Pathogenic	Publication	[40]
FGF12	00074-C	NM_021032.4(FGF12):c.341G>A (p.R114H)	NR	Pathogenic	Publication	[41]
MTOR	00040-C	NM_004958.3(MTOR):c.4785G>A (p.Met1595Ile)	NR	Pathogenic	Publication	For Review [26]; See also [27]
MTOR	00028-C, 00028-C2	NM_004958.3(MTOR):c.5663T>G (p.Phe1888Cys)	NR	Pathogenic	Filter	In original filter, required allele count of one; this variant was present in identical twins
HDAC8	00001-C	NM_018486.2(HDAC8):c.737+1G>A	NR	Likely Pathogenic	Filter	In original filter, required depth for all members of trio was set to 10 reads; father had only 7
LAMA2	00055-C, 00055-S	NM_000426.3(LAMA2):c.715C>T (p.Arg239Cys)	NR	Likely Pathogenic	Clarification of Clinical Phenotype	Discussion with clinicians was necessary to determine that patients' phenotypes did match those observed for LAMA2
MAST1	00270-C	NM_014975.2:c.278C>T, p.Ser93Leu	NR	Likely Pathogenic	GeneMatcher	Collaboration with several other groups identified patients with comparable genotypes and phenotypes
SUV420H1	00056-C	NM_017635.3:c.2497G>T, p.Glu833X	NR	Likely Pathogenic	Publication	[16]