

1 Title

2 *Comprehensive reanalysis for CNVs in ES data from unsolved rare disease cases*
3 *results in new diagnoses*

4 Authors

5 German Demidov^{1,2*-}, Burcu Yaldiz^{3,4*}, José Garcia-Pelaez^{5,6,7*}, Elke de Boer^{3,8,9*}, Nika
6 Schuermans^{10*}, Liedewei Van de Vondel^{11,12*}, Ida Paramonov¹³, Lennart F. Johansson¹⁴,
7 Francesco Musacchia^{15,16}, Elisa Benetti¹⁷, Gemma Bullich¹³, Karolis Sablauskas^{3,18}, Sergi
8 Beltran^{13,19,20}, Christian Gilissen³, Alexander Hoischen^{3,21,22}, Stephan Ossowski^{1,2}, Richarda
9 de Voer^{3,23}, Katja Lohmann²⁴, Carla Oliveira^{5,6,7}, Ana Topf²⁵, Lisenka E.L.M. Vissers^{3,8}, the
10 Solve-RD Consortia, Steven Laurie^{13*-}

11

12 Author Contributions

13 * These authors contributed equally to this work

14 - Corresponding authors: German.Demidov@med.uni-tuebingen.de; steven.laurie@cnag.eu

15

16 The authors declare no conflicts of interest.

17

18 Affiliations

19 1. Institute of Medical Genetics and Applied Genomics, University of Tübingen, Tübingen,
20 Germany

21 2. Institute for Bioinformatics and Medical Informatics (IBMI), University of Tübingen,
22 Tübingen, Germany

23 3. Department of Human Genetics, Radboud University Medical Center, Nijmegen, The
24 Netherlands

25 4. Department of Clinical Genetics, Maastricht University Medical Center, Maastricht, The
26 Netherlands

27 5. i3S - Instituto de Investigação e Inovação em Saúde, Rua Alfredo Allen, 208, 4200-135,
28 Porto, Portugal

- 29 6. IPATIMUP - Institute of Molecular Pathology and Immunology, University of Porto,
30 Portugal
- 31 7. Faculty of Medicine, University of Porto, Porto, Portugal
- 32 8. Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen,
33 The Netherlands
- 34 9. Department of Clinical Genetics, Erasmus MC University Medical Center, Rotterdam,
35 The Netherlands
- 36 10. Center for Medical Genetics, Ghent University Hospital, Ghent, Belgium
- 37 11. Translational Neurosciences, Faculty of Medicine and Health Sciences, University of
38 Antwerp, Antwerp, Belgium
- 39 12. Laboratory of Neuromuscular Pathology, Institute Born-Bunge, University of Antwerp,
40 Antwerp, Belgium
- 41 13. Centro Nacional de Análisis Genómico (CNAG), Barcelona Science Park, 08028,
42 Barcelona, Spain
- 43 14. University of Groningen, University Medical Center Groningen, Department of Genetics,
44 Groningen, The Netherlands
- 45 15. Center for Human Technologies, Italian Institute of Technology (IIT), Genova, Italy
- 46 16. Telethon Institute for Genetics and Medicine, 80078 Pozzuoli (Napoli), Italy
- 47 17. Med Biotech Hub and Competence Center, Department of Medical Biotechnologies,
48 University of Siena, 53100 Siena, Italy
- 49 18. Institute of Data Science and Digital Technologies, Vilnius University, Vilnius, Lithuania
- 50 19. Universitat Pompeu Fabra (UPF), Barcelona, Spain
- 51 20. Departament de Genètica, Microbiologia i Estadística, Facultat de Biologia, Universitat
52 de Barcelona (UB), Barcelona, Spain
- 53 21. Radboud Institute for Molecular Life Sciences, Nijmegen, The Netherlands
- 54 22. Department of Internal Medicine and Radboud Center for Infectious Diseases (RCI),
55 Radboud University Medical Center, Nijmegen, the Netherlands
- 56 23. Research Institute for Medical Innovation, Radboud University Medical Center,
57 Nijmegen, The Netherlands
- 58 24. Institute of Neurogenetics, University of Lübeck, Ratzeburger Allee 160, 23562, Lübeck,
59 Germany
- 60 25. John Walton Muscular Dystrophy Research Centre, Translational and Clinical Research
61 Institute, Newcastle University and Newcastle Hospitals NHS Foundation Trust,
62 Newcastle upon Tyne, United Kingdom

63

64 **Abstract**

65 We report the diagnostic results of a comprehensive copy number variant (CNV) reanalysis
66 of 9,171 exome sequencing (ES) datasets from 5,757 families, including 6,143 individuals
67 affected by a rare disease (RD). The data analysed was extremely heterogeneous, having
68 been generated using 28 different exome enrichment kits, and sequenced on multiple short-
69 read sequencing platforms, by 42 different research groups across Europe partnering in the
70 Solve-RD project. Each of these research groups had previously undertaken their own
71 analysis of the ES data but had failed to identify disease-causing variants.

72

73 We applied three CNV calling algorithms to maximise sensitivity: ClinCNV, Conifer, and
74 ExomeDepth. Rare CNVs overlapping genes of interest in custom lists provided by one of
75 four partner European Reference Networks (ERN) were identified and taken forward for
76 interpretation by clinical experts in RD. To facilitate interpretation, Integrative Genomics
77 Viewer (IGV) screenshots incorporating a variety of custom-made tracks were generated for
78 all prioritised CNVs.

79

80 These analyses have resulted in a molecular diagnosis being provided for 51 families in this
81 sample, with ClinCNV performing the best of the three algorithms in identifying disease-
82 causing CNVs. We also identified pathogenic CNVs that are partially explanatory of the
83 proband's phenotype in a further 34 individuals. This work illustrates the value of reanalysing
84 ES *cold cases* for CNVs even where analyses had been undertaken previously. Crucially,
85 identification of these previously undetected CNVs has resulted in the conclusion of the
86 diagnostic odyssey for these RD families, some of which had endured decades.

87 Introduction

88 Rare diseases (RD) are defined in Europe as conditions which affect less than 1 in 2,000
89 individuals. Nevertheless, it is estimated that more than 30 million people across the
90 European Union are affected by one of ~6000-8000 different RDs^{1,2}. As 80% of RD are
91 expected to have a genetic aetiology, massively parallel sequencing approaches, in
92 particular exome sequencing (ES), have been widely applied over the last decade to identify
93 variants in DNA that cause RD. However, despite many advances in technology during this
94 period, more than half of all individuals affected by an RD remain without a molecular
95 diagnosis following such analyses, thus extending their diagnostic odyssey. While the
96 accurate detection of single nucleotide variants (SNV) and short (<50nt) insertions and
97 deletions (InDels) from ES data has become relatively robust in recent years³, the reliable
98 detection of larger variants, including copy number variants (CNVs), remains a challenge,
99 and it is likely that undetected pathogenic CNVs account for a proportion of undiagnosed
100 individuals.

101

102 CNVs comprise losses, which may be heterozygous or homozygous in autosomes, or
103 hemizygous in gonosomes, and gains of genetic material, which we refer to here as
104 *deletions* and *duplications*, respectively. Identification of CNVs from short-read ES data (*i.e.*
105 100-150nt paired-end reads) is complicated by several factors, the most important of which
106 being that read length is usually shorter than variant length, and that the boundaries of the
107 CNV, referred to as breakpoints, are unlikely to be captured directly by the enrichment
108 targets, since they represent only ~1-2% of the genome. An exacerbating factor is marked
109 variability in the enrichment process, in which targets for ~200,000 exons undergo DNA
110 hybridisation and PCR amplification prior to sequencing, both between kits, and between
111 experiments. Many methods have been developed for CNV detection from ES data, most of
112 which use comparison of depth of coverage (DoC) between the observed number of reads
113 covering a particular exon/target in a sample of interest and the normalised coverage for the

114 same exon/target in a large homogeneous reference batch of matched experimental
115 samples⁴⁻⁹. For such methods to be successful, the sequencing data needs to be as
116 homogenous as possible, particularly with respect to evenness of coverage¹⁰, which is the
117 key factor in CNV detection since it directly affects the signal-to-noise ratio.

118

119 As reviewed recently in Gordeeva *et al.*¹¹, these methods differ from each other primarily in
120 terms of the approach taken for read count normalisation, assumptions regarding read-depth
121 distribution, and the segmentation process, *i.e.* identification of the boundaries of a variant.
122 Despite application of sophisticated normalisation techniques, the correct separation of the
123 signal of true CNVs from background noise remains challenging, particularly for short CNVs
124 that only impact upon one or a few exons. This is illustrated by numerous cross-tool
125 comparisons in which the intersection of CNVs detected by different methods is limited,
126 ranging from ~1-20% concordance when three or more tools are compared across
127 samples¹²⁻¹⁴. Indeed, a recent benchmarking initiative involving sixteen tools showed that
128 the number of raw CNVs called on a single ES sample ranged from just two to over a
129 thousand¹¹, reflecting differing optimisation of algorithms for specificity or sensitivity.
130 Therefore, following identification of a list of potential CNVs, subsequent filtering steps are
131 required, including determining which CNVs are technically valid (*i.e. bona fide* biological
132 events), and whether any of the valid CNVs are of clinical relevance with respect to the
133 phenotype of the affected individual. Hence, both technical expertise and expert clinical
134 knowledge are required if disease-causing CNVs are to be correctly identified. diagnoses.

135

136 This complexity may explain why the detection of CNVs has often been omitted from
137 diagnostic ES workflows, with array comparative genome hybridisation (aCGH) continuing to
138 be the preferred method in the clinic over the last decade, despite limitations in its sensitivity
139 and resolution, particularly with respect to short CNVs. However, recent studies have
140 indicated that ES may be a suitable replacement as a first-tier diagnostic test¹⁵⁻¹⁷, with the
141 added benefit that SNVs and InDels are detected simultaneously.

142 A key goal of the EU Horizon 2020 Solve-RD project is to raise the diagnostic rate of
143 individuals with an RD for whom ES analysis and variant interpretation have previously been
144 undertaken, but without a conclusive diagnosis having been reached. This is being achieved
145 by undertaking massive pan-European data collation and complete reanalysis from raw data,
146 followed by expert technical and clinical interpretation and validation of variants¹⁸. The CNV
147 analysis conducted here, was an integral part of a larger re-analysis effort undertaken on the
148 same dataset, covering most other variant types (Laurie *et al.* under review). Here we
149 describe the workflow applied in a comprehensive reanalysis of this heterogeneous sample
150 of ES data from 9,171 individuals pertaining to 5,757 families, including 6,143 individuals
151 affected by an RD, to identify (likely) pathogenic CNVs. The ES data had been generated
152 using 28 different enrichment kits in multiple sequencing centres. Hence, to maximise
153 accuracy and sensitivity of CNV detection we applied three different algorithms, ClinCNV,
154 Conifer, and ExomeDepth, and analysed experiments in 28 different batches, comprising
155 data generated using the same enrichment kit. We filtered the raw call set, initially consisting
156 of over two million CNV calls (average of ~300 per individual), to a manageable number of 0-
157 2 potentially pathogenic rare CNVs per affected individual requiring interpretation by the
158 clinical experts who has submitted the cases to Solve-RD. This extensive endeavour has led
159 to the closure of many diagnostic odysseys, some of which had been ongoing for decades,
160 of which we provide some illustrative examples.

161

162 **Methods**

163 ***Data Collation***

164 The ES data reanalysed here comprises previously inconclusive ES experiments submitted
165 for reanalysis as part of the Solve-RD project by 42 different research groups based in
166 twelve countries across Europe, and Canada (range of 1-2,111 experiments submitted per

167 group). Each experiment was submitted via one of four European Reference Networks
168 (ERN) partnering in Solve-RD, each focusing on a particular group of RD: EURO-NMD (rare
169 neuromuscular diseases); GENTURIS (rare genetic tumour risk syndromes); ITHACA (rare
170 malformation syndromes, intellectual and other neurodevelopmental disorders); RND (rare
171 neurological diseases).

172

173 A total of 9,351 ES experiments from 9,314 individuals (6,224 affected individuals and 3,090
174 unaffected relatives) were initially submitted for reanalysis. After the removal of samples
175 sequenced with enrichment kits for which the available control cohort was not large enough
176 to allow accurate CNV identification, data from 9,171 individuals from 5,757 families was
177 analysed (see Technical Results). While 1,320 of 1,788 (74%) families from ITHACA were
178 composed of parent-child trios, facilitating identification of *de novo* mutations, only 239 of the
179 remaining 3,969 (6%) probands from other ERNs were trios. ES had been performed using
180 28 different enrichment kits (range of 4-2,078 experiments per kit), and each of the forty-two
181 research groups had followed their own DNA library preparation, target enrichment, and
182 short-read sequencing protocol in their local labs, or via external DNA sequencing providers.
183 Furthermore, each group had previously undertaken their own historic analysis and
184 interpretation of the resulting ES data to identify disease-causing variants, which has proven
185 inconclusive. The date at which the initial ES analysis and interpretation had been
186 undertaken ranged from six months to eight years prior to the experimental data being
187 submitted to Solve-RD for reanalysis, however this information was not collected
188 systematically for individual data sets.

189

190 In addition to sequencing data, a phenotypic description for each affected individual was
191 recorded in the PhenoStore module of the RD-Connect GPAP¹⁹, consisting of a minimum of
192 five Human Phenotype Ontology terms (HPO²⁰) wherever possible, and disease
193 classification using Orphanet Rare Disease Ontology (ORDO) ORPHA codes
194 (<http://www.orphadata.org/cgi-bin/index.php>), and/or OMIM identifiers²¹

195 (<https://www.omim.org/>) where appropriate, together with family pedigrees. A detailed
196 description of this data set can be found in Laurie et al, 2023 (under review).

197 ***CNV Identification***

198 Raw ES data was realigned to the hs37d5 reference genome²², using BWA-MEM²³, as
199 described in the Supplementary Materials. With the goal of maximising the probability of
200 detecting potentially disease-causing CNVs, three different algorithms which identify CNVs
201 based on DoC were applied. Two of these, Conifer⁴, and ExomeDepth⁶, have been widely
202 applied to ES data with success previously, while the third, ClinCNV, was developed recently
203 by a Solve-RD partner²⁴. Each of these tools offers the practical advantage of separating the
204 DoC calculation for each individual experiment from the CNV calling step, and thus CNVs
205 were subsequently called in batches by enrichment kit. Furthermore, each algorithm
206 provides an estimate of the likelihood that calls produced are biologically real, and the most
207 likely false positive calls were excluded based upon these metrics. As primary filters, in the
208 case of Conifer a value in excess of +/-1.75 SV-RPKM was required for a CNV call to be
209 taken forward for biological interpretation, while for ExomeDepth a Bayes Factor (BF)
210 greater than fifteen was required, and for ClinCNV a minimum log-likelihood estimation of
211 twenty was applied (see Supplementary Methods for further detail).

212 ***Call Filtering and Visualisation***

213 As the focus of Solve-RD is diagnosing RD cases, through the identification of rare variants
214 that are potentially disease-causing, any apparent CNV call observed in a region where
215 more than 1% of individuals in the whole sample had a similar type of call (*i.e.* a deletion or
216 duplication) were discarded as being too common to be clinically relevant with respect to
217 RD. Furthermore, CNVs returned for interpretation by clinical experts were restricted to
218 those which overlapped with at least part of a gene in a predefined list of curated genes of
219 interest provided by the respective ERN: EURO-NMD (n=615), GENTURIS (230), ITHACA
220 (1,944), RND (1,820). The full list of ERN curated genes is provided in **Supplementary**

221 **Table 1** and details as to how these lists were determined in Laurie et al, 2023 (under
222 review). Potential CNVs of interest were subsequently categorised into six non-redundant
223 classes to aid interpretation: Long CNVs (>500kb in length) ; Homozygous deletions;
224 Heterozygous CNVs affecting genes known to cause disorders with an autosomal dominant
225 mode of inheritance; Regions with apparent copy numbers of four or more; Gonosomal
226 CNVs ; Potential compound-heterozygous *double-hits* in the form of a CNV affecting the
227 second allele of a gene in which biallelic variants are known to be disease-causing, and in
228 which a potentially pathogenic SNV has been previously identified in Solve-RD.

229

230 To provide support for interpretation of the technical validity of CNV calls, images of regions
231 containing CNV calls were generated automatically using the Integrative Genomics Viewer
232 (IGV)²⁵. A variety of custom tracks, including call tracks for each of the three algorithms,
233 BAM DoC, and gene tracks for ERN genes of interest were incorporated, among others (see
234 Supplementary Methods).

235 ***Clinical Interpretation***

236 Further annotations to aid interpretation (**Supplementary Table 2**) were added to the results
237 using AnnotSV²⁶ (Version 3.0.7), and fully annotated CNV call sets generated for all tools
238 together with accompanying customised IGV visualisations were distributed to clinical
239 experts in each ERN for diagnostic interpretation. Each ERN prioritised calls for further
240 investigation based on their expert knowledge of underlying disease mechanisms in their
241 respective patients. Many CNV calls could be rapidly discarded based upon a lack of match
242 between the gene potentially affected and the phenotype of the affected individual, and/or
243 segregation patterns within the family. Others were rejected when visual inspection of the
244 IGV tracks indicated that they were likely false-positive calls, and thus unlikely to be *bona*
245 *fide* biological events. Where deemed necessary and when feasible, CNVs believed to be
246 diagnostically relevant were validated at local centres using orthologous approaches. The

247 final decision as to whether a CNV was determined to be pathogenic or not was taken by the
248 respective clinical experts from the ERN (see Supplementary Methods for further details).

249

250 **Results**

251 ***Technical Results***

252 Prior to the initiation of CNV calling, a minimal quality control was undertaken, which took the
253 form of requiring that data from each submitted family included at least one affected
254 individual with accompanying HPO terms. Furthermore, following alignment of sequencing
255 reads, it was required that at least 70% of the target region of the enrichment kit had a DoC
256 of ten reads. After removal of 143 experiments which did not meet these criteria, CNV calling
257 was undertaken on data from a total of 9,171 individuals from 5,757 families, of whom 6,143
258 had a rare condition. Initial investigations indicated the presence of a large variance in
259 sequencing depth both within and between the twenty-eight enrichment kit batches,
260 reflecting the heterogeneity of the sequencing data submitted to Solve-RD (**Supplementary**
261 **Figure 1**).

262

263 Following identification and removal of likely false positive calls based upon tool-specific QC
264 metrics, the removal of commonly observed events, and restriction to events overlapping
265 genes in the custom gene lists from the corresponding ERN, a total of 7,849 calls in 3,436
266 affected individuals from 3,300 families remained for interpretation (**Table 1**). The number of
267 probands with at least one CNV call to be interpreted by clinical specialists from the ERN
268 ranged from 113 for GENTURIS (33% of families), to 1,239 for ITHACA (69% of families)
269 (**Supplementary Table 3**). No CNV of interest was detected in 2,707 affected individuals
270 from the remaining 2,457 families. In addition, a further 393 pairs of potential CNV-SNV
271 *double-hit* compound heterozygous variants in 226 affected individuals were returned to
272 clinical experts for interpretation. Overall, a mean of 1.3 CNVs per proband were returned for

273 interpretation. However, as CNVs of potential interest were only identified in 55% of
274 probands, this equated to 2.4 variants per proband that required interpretation.

275

276 The total number of CNV calls in affected individuals returned for interpretation was highest
277 for ExomeDepth (n=4,205), while ClinCNV called about two-thirds of this number (2,782),
278 and Conifer approximately one-fifth (862), reflecting different predilections of the underlying
279 algorithms with respect to sensitivity and specificity of CNV detection. While Conifer and
280 ExomeDepth showed a significant bias towards calling duplications, the reverse pattern was
281 observed for ClinCNV, which identified more deletions ($p < 0.00001$ in all cases, Fisher exact
282 test; **Supplementary Table 4**). We assessed the distribution of the length of CNVs returned
283 for interpretation as identified by each tool. Notably, the average length of CNVs detected by
284 Conifer was approximately an order of magnitude larger than that of ExomeDepth, which in
285 turn was longer than that of ClinCNV. This pattern held for both duplications and deletions,
286 and again reflects differences in the way the tools identify and segment CNVs (**Figure 1**,
287 **Supplementary Table 5**).

288 ***Diagnostic Results***

289 Following expert interpretation, 105 potentially pathogenic CNVs of interest in 103 affected
290 probands were identified, of which 52 have been confirmed as disease-causing in 51
291 individuals (**Table 2**). The disease-causing CNVs included three “double-hit” instances
292 where an SNV and CNV affecting different alleles of the same gene were identified, resulting
293 in a compound heterozygous diagnosis, and one instance where two CNVs affecting
294 different genes provided a dual genetic diagnosis for a complex phenotype. A further 25
295 CNVs are regarded as pathogenic by the clinical experts, but not sufficient, to explain the full
296 phenotype observed in the affected individual, including seven complete gonosomal
297 aneuploidies (“Partially Explanatory” in **Tables 2 and 3**). A further 28 potentially pathogenic
298 CNVs were identified for which further validation is not logistically possible due to lack of
299 access to DNA and/or the patient (referred to as candidates below). While 81% (42 of 52) of

300 confirmed disease-causing CNVs are deletions, only 39% (7 of 18) of the partially
301 explanatory pathogenic CNVs are deletions, even when disregarding the gonosomal
302 duplications. Of the 28 candidate CNVs 57% (16) are deletions (**Figure 2, Table 2**).

303

304 Of the 77 confirmed pathogenic CNVs, 40 (52%) were initially identified by all three callers
305 (**Figure 2, Table 2**). However, in the case of ten of the 40, the Conifer call was subsequently
306 discarded due to it being within the applied SV-RPKM threshold, and one of the ten was also
307 discarded by the ExomeDepth workflow due to a low BF. Of the remaining 37 pathogenic
308 CNVs, 36 (97%) were identified by ClinCNV, two of which subsequently failed ClinCNV
309 quality control thresholds, while 25 (68%) were identified by ExomeDepth, five of which were
310 subsequently discarded due to a low BF. Interestingly one of the 37, a duplication in *PIEZO2*
311 was identified by Conifer alone.

312

313 ***Examples of successful new diagnoses***

314 Below we provide an example of an RD case from each of the four ERN partners in Solve-
315 RD solved through the analysis of CNVs undertaken here.

316 ERN EURO-NMD

317 This male in his thirties first came to clinical attention in his adolescence, affected by poor
318 balance, recurrent falls, and difficulty rising from the floor. Prior to this he had been able to
319 run and play sports normally. His symptoms worsened slowly over time, and he is currently
320 unable to walk or stand without assistance. He also has mild facial weakness and mildly
321 elevated serum creatine kinase. His family history is negative, having several unaffected
322 siblings. Muscle biopsy showed clear features of muscular dystrophy, and
323 immunohistochemical analysis suggested reduced expression of dystrophin. Exome
324 sequencing was initially undertaken in 2017, but no diagnosis was reached at that point.

325 As a result of reanalysis of the ES data undertaken here, a three-exon deletion affecting
326 exons 45 through 47 of the *DMD* gene was detected by both ExomeDepth and ClinCNV,
327 consistent with the suspected diagnosis of Becker Muscular Dystrophy. This hemizygous
328 deletion was subsequently confirmed using Multiplex Ligation-dependent Probe Amplification
329 (MLPA). Confirmation of the molecular diagnosis in this individual has enabled enhanced
330 genetic counselling, as any future daughter he may have would be an obligate, and possibly
331 manifesting, carrier of the CNV, thus requiring clinical management.

332 ERN GENTURIS

333 This family first came to clinical attention in 2003, meeting the criteria for hereditary diffuse
334 gastric cancer (HDGC)²⁷, as several family members had developed diffuse gastric cancers
335 prior to 30 years of age. HDGC typically results from *CDH1* loss of function^{28,29}. However,
336 Sanger sequencing of *CDH1* performed proved negative, as did subsequent investigation in
337 the form of MLPA, and ES, at which point no potentially explanatory SNVs, InDels, or CNVs
338 were identified in *CDH1*, nor other candidate genes³⁰.

339

340 Following these negative findings, the ES data was submitted to Solve-RD for two affected,
341 and four unaffected siblings. The comprehensive reanalysis of the ES data resulted in the
342 identification of a ~116kb heterozygous deletion impacting half of the *CDH1* gene (from
343 intron 7 forwards) and the start of the downstream gene *TANGO6* (as far as intron 14)
344 (g:16:68846036-68964198del) in four of the six siblings (**Supplementary Figure 2**). The
345 CNV was detected by both ClinCNV and ExomeDepth and further supported by split-reads
346 and abnormally paired reads observed in data from one of the affected individuals.
347 Visualisation in IGV, and subsequent MLPA, validated this large event. Of note, one of the
348 unaffected siblings, a female carrier in her forties, has not developed gastric cancer to date,
349 in accordance with previously reported incomplete penetrance among *CDH1* mutation
350 carriers³¹. Another of the unaffected siblings was a carrier but never developed gastric
351 cancer as a result of having undergone prophylactic total gastrectomy due to the high

352 incidence of cancer in the family. The remaining unaffected siblings were found not to
353 harbour the deletion, but unfortunately both have also already undergone prophylactic
354 gastrectomy. Nevertheless, as a result of their inclusion in Solve-RD, the family has since
355 been recontacted and enrolled in a clinical pathway of care, and their twenty-year diagnostic
356 odyssey has now come to an end. Importantly, targeted genetic testing has now been made
357 available to their offspring to avoid unnecessary prophylactic gastrectomy in subsequent
358 generations. The functional analysis and clinical implications of this CNV are described in
359 more detail in São José et al.³².

360 ERN ITHACA

361 This girl was first referred to paediatric neurology in her first year of life, presenting with
362 generalised tonic-clonic seizures. During her infancy mild global developmental delay
363 became evident, with delays in speech and language acquirement and in gross-motor skill
364 acquisition. Seizures were controlled with lamotrigine monotherapy, which could be
365 discontinued during childhood following prolonged seizure-free periods. Apart from
366 polyhydramnios, pregnancy and delivery were uncomplicated. Medical history comprised
367 constipation and eczema, and family history was unremarkable. Physical examination
368 revealed no additional phenotypic features *i.e.* no congenital anomalies, no facial
369 dysmorphisms, and no growth abnormalities. Investigations, including cerebral MRI and
370 general metabolic screening. Singleton ES was performed, followed by trio ES which
371 revealed a heterozygous *de novo* SNV of uncertain significance (VUS) in *STIP1* (*STIP1*;
372 Chr11(GRCh37):g.63961718C>T; NM_001282652.1:c.418C>T; p.(Arg140*)). Within this
373 diagnostic trajectory, no analysis dedicated to CNV detection was performed.

374

375 The systematic reanalysis of ES data reported here led to the identification of a
376 heterozygous 27kb deletion on chromosome 6p21 (chr6:31630124-31657924-DEL) in the
377 proband. This deletion was detected by all three tools, and visual inspection of sequence
378 alignment files in IGV clearly indicated the presence of the variant in the affected daughter,

379 and its absence in both parents, thus confirming that it is a *de novo* deletion. The deletion
380 fully removes *CSNK2B*, *LY6G5B* and *LY6G5C*, and its breakpoints affect *GPANK1* and
381 *ABHD16A*. *GPANK1*, *LY6G5B* and *LY6G5C* currently have no disease association, and
382 while *ABHD16A* is associated with autosomal recessive spastic paraplegia-86
383 (MIM#619735), there is no apparent second hit in *ABHD16A*, and the phenotype of the
384 proband does not comprise spastic paraplegia. *CSNK2B*, on the other hand, has recently
385 been shown to be associated with autosomal dominant Poirier-Bienvenu
386 neurodevelopmental syndrome (POBINDS; MIM#618732), in which truncating variants in
387 *CSNK2B* result in haploinsufficiency, leading to early-onset seizures and highly variable
388 impairments of intellectual functioning³³⁻³⁵. As the *de novo* deletion observed in this proband
389 results in haploinsufficiency of *CSNK2B*, and her phenotypic descriptions fits within the
390 *CSNK2B*-associated phenotypic spectrum, this 27kb deletion on chromosome 6p21 is
391 regarded as explanatory for her rare condition, thus ending a seven-year diagnostic odyssey
392 for this family.

393 ERN RND

394 This teenage female was first evaluated in paediatric neurology as a child, presenting with
395 global developmental delay, and behavioural and learning problems. Retrospectively, the
396 first symptoms had become apparent in her infancy, consisting of mild delayed development
397 of fine and gross motor skills. Additionally, she had delays in language and speech
398 development, and was diagnosed with attention deficit disorder, for which she is being
399 treated with methylphenidate and responding well. No obvious dysmorphic features were
400 observed upon physical examination, but mild hypertonia of the triceps surae, hyperreflexia,
401 kinetic tremor, mirror hand movement, and a tiptoeing gait were observed. Subsequent
402 cerebral MRI showed ventriculomegaly, corpus callosum hypoplasia, prominent cerebellar
403 folia, and thin middle cerebellar peduncles. Genetic testing, consisting of aCGH (median
404 resolution 180kb), targeted testing for Fragile X syndrome, and ES did not pinpoint a
405 molecular cause.

406

407 Systematic reanalysis of the ES data undertaken here led to the identification of a
408 heterozygous deletion of ~200kb at chromosome 4q31.1: Chr4(GRCh37):g.140187686-
409 140394334del, encompassing part of the *MGARP* gene (not known to be associated with
410 disease), and the entire *NAA15* gene, which encodes the catalytic subunit in the N-terminal
411 acetyltransferase A complex (MIM: 608000). The deletion was identified by all three tools,
412 and subsequently validated using high resolution aCGH (median resolution 60kb). Following
413 review of the prior results, the absence of recall of the variant in the initial aCGH analysis
414 was attributed to its limited resolution. The patient's mother, who had had similar learning
415 problems and has mild cognitive disability, was subsequently also found to be positive for
416 the deletion. No further family testing was possible. Echocardiography was normal in both
417 cases. Loss-of-function variants in *NAA15* and heterozygous deletion of this gene and
418 nearby genes are associated with 'Intellectual developmental disorder, autosomal dominant
419 50, with behavioural abnormalities' (MIM: 617787)^{36,37}. This disorder has the features of a
420 wide spectrum of neurodevelopmental severity and variable association of congenital
421 anomalies, thus confirming that the observed CNV was causative in this case and ending
422 this family's seven-year diagnostic odyssey.

423 **Discussion**

424 Rigorous detection of CNVs from ES requires sequencing data that has been generated as
425 uniformly as possible, in order that the test experiment can be compared against a similarly
426 generated batch of matched control samples. However, the ES data submitted to Solve-RD
427 had been generated using twenty-eight different enrichment kits and sequenced with
428 different short-read technologies to different depths of coverage, in multiple sequencing
429 centres across Europe. Hence the primary challenge encountered during this analysis was
430 data heterogeneity. Similarly, from the perspective of diagnosis, it is essential to have a clear
431 clinical description of the affected individual to be able to determine in which genes, variants,
432 if encountered, may explain the observed phenotype. This was achieved here firstly through

433 use of the HPO ontology to capture a deep phenotypic description of affected individuals
434 from the referring clinicians, and secondly using the curated set of genes of interest provided
435 by each ERN. Together these significantly reduced the search space for potentially disease-
436 causing CNVs.

437

438 The interpretation of raw CNV calls is challenging due to the initial high number of calls most
439 tools report. We applied a robust filtering strategy to remove calls that were clearly unlikely
440 to be of relevance for RD and benefited from the curated lists of genes of interest provided
441 by each ERN. Nevertheless, visual inspection of the affected region using IGV was key for
442 assessing the technical validity of remaining calls, prior to, or in parallel with, their biological
443 interpretation. It is likely that this is an aspect where an AI-based tool for automated IGV-
444 image analysis would be of significant benefit, potentially saving many hours of human
445 expert-review time. The clinical researchers representing each ERN applied their own
446 prioritisation strategy when interpreting CNV calls, according to the specific pathologic and
447 phenotypic characteristics of their patients.

448

449 When used as a first-tier analysis, CNV detection from ES has been reported to result in
450 diagnostic yields as high as 7-19%³⁸⁻⁴⁰, whereas yield The overall rate of novel diagnoses
451 reached was 0.9%, ranging from 0.6% for RND and 0.9% for ITHACA to 1.2% for
452 GENTURIS and EURO-NMD. Notably nine of the sixteen CNVs established as being
453 disease-causing in ITHACA cases could be confirmed as *de novo* mutations due to ES data
454 being available from the proband's parents. While our values are lower than those of prior
455 reports, where yield from reanalyses efforts, have resulted in increases in diagnostic yield
456 with respect to CNVs in the range of 1.6-2.0%⁴¹⁻⁴³, in those studies the prior CNV analyses
457 had largely consisted of only chromosomal microarray (CMA) analyses, which lack
458 sensitivity for short CNV events which were hence identified in the subsequent ES-based
459 CNV analyses. Our results reflect several factors: the likelihood that detailed CNV analysis

460 of the ES had been undertaken prior to submission to Solve-RD; the role that CNVs are
461 likely to have in the respective class of disease; the time passed since the initial analysis,
462 which would affect the number of genes known to be associated with a particular class of
463 disease. Interestingly, the number of genes of interest in each of the custom ERN gene lists
464 does not appear to be a factor given that GENTURIS had by far the shortest list, and RND
465 and ITHACA the longest.

466

467 There was a clear bias towards deletions *vis-à-vis* duplications being identified as
468 pathogenic with 49 of 77 (64%) confirmed pathogenic CNVs being deletions, and 42 of 52
469 (81%) disease-causing CNVs. This reflects the facts that duplications are more challenging
470 to detect, and even when detected by ES, it is invariably unclear as to whether they are
471 tandem duplications, possibly inverted, or inserted elsewhere in the genome, each of these
472 scenarios being likely to result in a different biological consequence, making interpretation
473 challenging. Furthermore, long duplications appear to be under less evolutionary constraint
474 than similarly sized deletions⁴⁴, suggesting that they are less likely to result in disease.
475 Accordingly, the ACMG guidelines for the interpretation of constitutional CNVs⁴⁵, require
476 more supporting evidence for a duplication to be confirmed as pathogenic than is required
477 for a deletion.

478

479 It is noteworthy that, in comparison with the other two tools, Conifer called very few CNVs
480 under 20kb in length, and indeed failed to successfully identify 18 of 20 deletions <20kb that
481 were determined to be disease-causing, and the remaining two fell below the calling
482 threshold. Notably, Conifer also failed to identify duplications over 1Mb in length, including
483 seven sex-chromosome aneuploidies. It is this failure at the two extremes of CNV length that
484 largely contribute to the inferior performance of Conifer. It should also be highlighted that we
485 required a Z-score in excess of +/-1.75 for a CNV called by Conifer to be returned for

486 interpretation, whereas had we used +/-1.5, Conifer would have successfully identified a
487 further eight events of the disease-causing CNVs, all but two of which were over 20kb in
488 length. ClinCNV performed best of the three callers with this highly heterogeneous dataset,
489 which is likely due to its more adaptive DoC calculation whereby it subsegments target
490 regions into 120bp tiles, significantly improving resolution, particularly for short CNVs, most
491 of which were also detected by ExomeDepth but some fell below the minimal calling
492 threshold.

493

494 In addition to cases of *de novo* dominant inheritance resolved by an individual CNV, we also
495 identified eight cases where an SNV and CNV were affecting different alleles of the same
496 gene potentially forming a disease-causing compound heterozygote. Two of these have
497 been confirmed as being explanatory for the individuals' conditions, with the remaining six
498 requiring further validation. These findings underline the importance of having all data
499 relevant to the interpretation of an affected individual's condition readily at hand, as had the
500 SNV and CNV analyses been undertaken independently, these individuals would have been
501 unlikely to have received a diagnosis. Furthermore, in one affected individual, we identified
502 two pathogenic CNVs affecting different genes, each of which explain unique features of the
503 individual's complex phenotype, *i.e.* a dual diagnosis⁴⁶. We are confident that many of the
504 CNVs that we currently classify as candidates are likely pathogenic in the affected
505 individuals, but complete follow-up has not yet been possible. The complete expert-curated
506 dataset of deletions and duplications, together with the detailed phenotypes and pedigrees,
507 and the aligned sequence files (CRAMs) are available to the entire RD community via the
508 European Genome-Phenome Archive (EGA)⁴⁷, allowing for new discoveries.

509

510 There are many reasons why a pathogenic CNV identified here may not have been found in
511 prior analysis of the ES data. Firstly, there may have been no attempt to identify CNVs by

512 the respective clinical research team, due to a lack of resources or expertise. However, we
513 know that some form of prior CNV analysis had been undertaken for the majority of affected
514 individuals analysed here. Secondly, the tool(s) applied previously for CNV detection may
515 not have identified the relevant CNV, or though identified, it may have been discarded due to
516 local quality control parameters applied e.g. approximately 10% of all the experiments
517 submitted to Solve-RD were of sufficiently poor quality such that one of the centres involved
518 in the reanalysis undertaken here would have routinely QC-failed the sample in their
519 diagnostic workflow and thus not attempted to identify CNVs. Thirdly, while the CNV may
520 have been identified, there may not have been any known association between the affected
521 gene(s) and the clinical presentation of the patient at the time of the initial analysis, resulting
522 in, at best, classification of the CNV as a variant of uncertain significance (VUS), and the
523 individual remaining undiagnosed.

524

525 We would emphasise that any observations of potential tendencies in the results presented
526 here must be interpreted prudently since this was an extremely heterogeneous dataset both
527 in terms of the breadth and the quality of the data, and in terms of the time and expertise that
528 had been applied to the interpretation of the ES data in analyses undertaken prior to
529 submission to Solve-RD. As we gather more information about the role of CNVs in RD
530 through projects that share data widely such as Solve-RD, hopefully the accuracy of CNV
531 detection will improve, and the entire process of identification and interpretation of this
532 important class of variants, from sequencing data to identification of pathogenic variants can
533 be automated, resulting in families affected by RD receiving a diagnosis sooner rather than
534 later.

535

536 ***Limitations of this work***

537 The work presented here has several clear limitations *vis-a-vis* reaching a diagnosis for
538 individuals affected by an RD. Firstly, given that the data was from ES, and that we only
539 considered events affecting one of between 230 and 1,944 genes of interest identified by
540 each of the ERNs, we will obviously miss any non-exonic events, or CNVs affecting genes
541 not in the list of genes of interest. However, undertaking this work without using gene lists
542 would result in a currently insurmountable load of data for interpretation, and novel gene
543 discovery was not the goal of the work undertaken here. However, such discoveries are
544 enabled by the sharing of data with the wider RD community via the EGA, which we hope
545 will enable more cases to be solved. Different approaches in interpretation undertaken by
546 the ERN experts may have resulted in some biologically relevant events being discarded as
547 uninteresting, which may be particularly true for duplications, for which evidence of biological
548 relevance in RD is currently relatively scarce. It is also possible that application of other tools
549 designed to find CNVs affecting only single exons, such as VarGenius-HZD⁴⁸, may have
550 allowed the identification of shorter events missed by the tools applied. With the future
551 adoption of long-read genome sequencing technologies such as those provided by Oxford
552 Nanopore Technologies and Pacific Biosciences, it is likely that the accuracy of CNV
553 detection, and hence ease of interpretation, will improve markedly.

554

555 Despite these limitations, we have successfully provided diagnoses to at least 51 families
556 who had previously undergone extensive genetic testing and in many cases multiple hospital
557 visits over many years, some even decades, without having been provided with a diagnosis.
558 Within the larger Solve-RD reanalysis of all variant types, these 51 CNVs were the second
559 most common type of disease-causing variant identified, after SNVs/InDels, contributing to
560 ~9% of the successful diagnoses Laurie et al. 2023 (under review). The ending of a
561 diagnostic odyssey has many benefits to patients and their families, beyond changes in
562 medical management and genetic counselling of relatives. It also allows better
563 understanding of disease progression, access to disease-specific online communities, and

564 psychological closure, amongst other benefits⁴⁹. The work undertaken here indicates the
565 value of comprehensive (re)-analysis of copy number variants in undiagnosed RD cases,
566 even from historic ES data, and has resulted in patients and their families being given an
567 accurate diagnosis, finally ending their diagnostic odysseys.

568 **Recommendations**

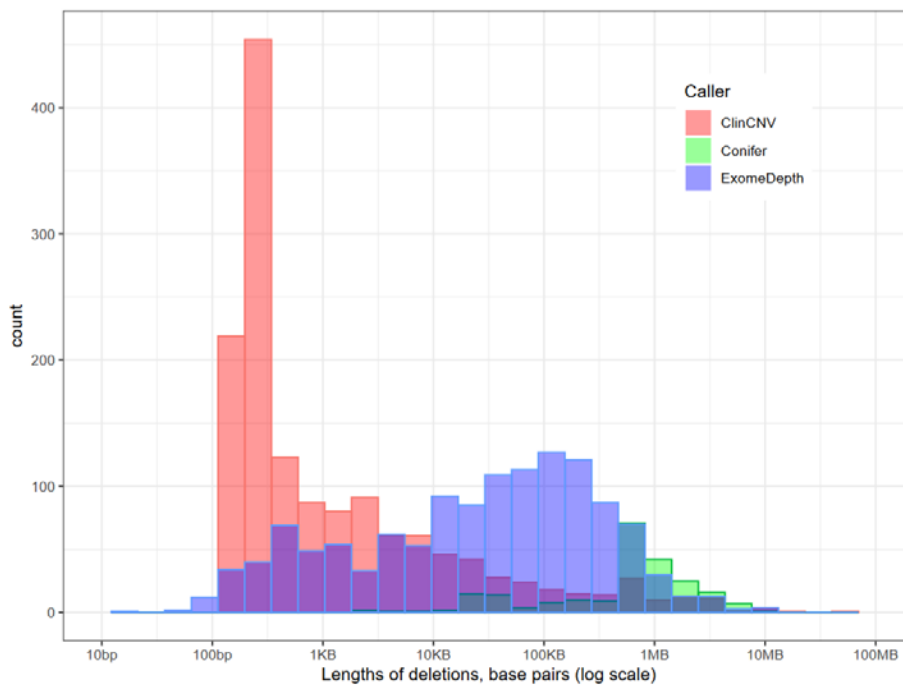
569 Based upon our findings, we suggest the following recommendations for future (re)-analyses
570 of ES data with respect to identification of disease-causing CNVs.

571

- 572 1) Know your enrichment kit. Investigate how well, and how evenly, does it capture your
573 genes of interest.
- 574 2) Choose your tools wisely. While Conifer has been shown to work with homogenous
575 datasets e.g. thousands of ES datasets generated using the same kit, in the same
576 sequencing centre, it does not perform with the heterogeneous dataset analysed
577 here. Furthermore, it identified very few CNVs <20kb in length, missing many
578 disease-causing variants.
- 579 3) Identifying regions that are commonly copy-number variant. In this way any CNVs
580 observed in such regions can be excluded from being potentially disease-causing.
- 581 4) Use an *in silico* candidate gene list when possible. This will greatly accelerate the
582 process of interpretation. If the list is very short, then any signal of a CNV in a gene
583 of interest should be examined further, since the sensitivity of tools remains low, and
584 the prior probability of the gene being variant is high. However, as lists grow longer,
585 this probability reduces, and calls will have to be filtered by quality thresholds.
- 586 5) Visualisation of CNV calls using a tool such as IGV is essential to assure that they
587 are likely to be real biological events, prior to expending time and effort on further
588 interpretation, investigation, and/or confirmation using orthologous techniques.

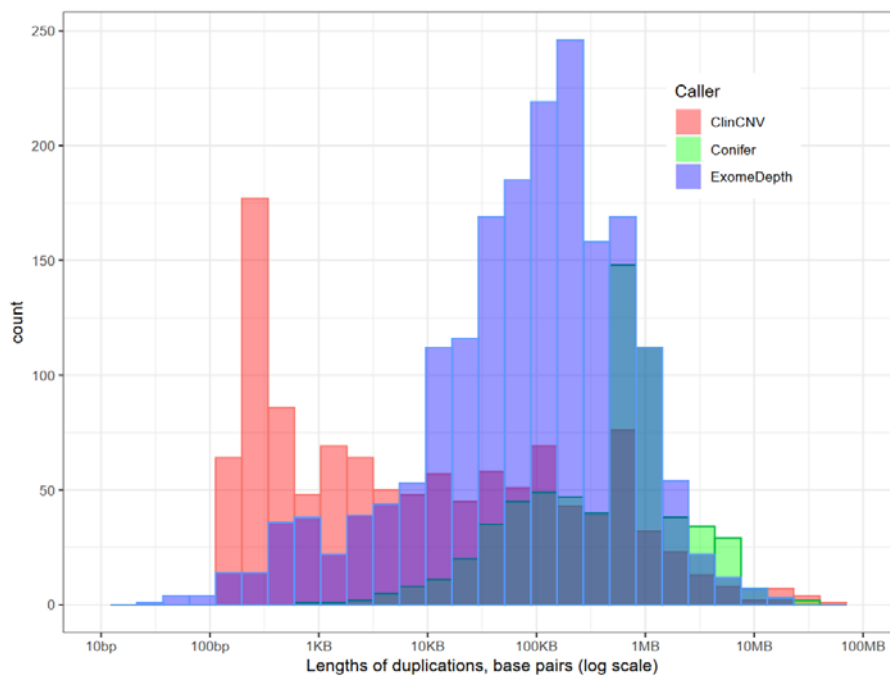
589 Figures

590 1.A Deletions



591

592 1.B Duplications



593

594 **Figure 1. Distribution of lengths of 7,849 CNV calls detected in 3,436 affected individuals, separated into**
595 **deletions (Panel A) and duplications (Panel B).** The x-axis represents the length of calls identified (\log_{10} scale),
596 and the y-axis the number of events observed. ClinCNV is represented in red. Note that the y-axis scale is
597 different in 1.A from 1.B

608 along the bottom. Where more than one gene was unaffected, it is shown as multiple, with the
609 affected chromosome indicated.

610 **Data availability and Ethics Statement**

611 Data will be deposited at EGA. Accession numbers to be provided. The family (FAM) and
612 participant (P) identifiers used in this manuscript are pseudonymised and known only to the
613 researchers involved in Solve-RD. The Ethics committee of the Eberhard Karl University of
614 Tübingen gave ethical approval for this work.

615

616 **Funding**

617 The Solve-RD project has received funding from the European Union's Horizon 2020
618 research and innovation programme under grant agreement number 779257. Data were
619 partially analysed and stored using the RD-Connect Genome-Phenome Analysis Platform,
620 which received funding from EU projects RD-Connect, Solve-RD and EJP-RD (grant
621 numbers FP7 305444, H2020 779257, H2020 825575), Instituto de Salud Carlos III (grant
622 numbers PT13/0001/0044, PT17/0009/0019; Instituto Nacional de Bioinformática, INB) and
623 ELIXIR Implementation Studies. We acknowledge support of the Spanish Ministry of
624 Economy, Industry and Competitiveness (MEIC) to the EMBL partnership, the Centro de
625 Excelencia Severo Ochoa and the CERCA Programme/Generalitat de Catalunya. We also
626 acknowledge the support of the Generalitat de Catalunya through Departament de Salut and
627 Departament d'Empresa i Coneixement and the Co-financing by the Spanish Ministry of
628 Economy, Industry and Competitiveness (MEIC) with funds from the European Regional
629 Development Fund (ERDF) corresponding to the 2014-2020 Smart Growth Operating
630 Program.

631

632 References

- 633 1. Nguengang Wakap, S. *et al.* Estimating cumulative point prevalence of rare diseases:
634 analysis of the Orphanet database. *European Journal of Human Genetics* **28**, (2020).
- 635 2. European Commission. EU research on rare diseases. [https://research-and-](https://research-and-innovation.ec.europa.eu/research-area/health/rare-diseases_en)
636 [innovation.ec.europa.eu/research-area/health/rare-diseases_en](https://research-and-innovation.ec.europa.eu/research-area/health/rare-diseases_en) (2023).
- 637 3. Poplin, R. *et al.* Scaling accurate genetic variant discovery to tens of thousands of
638 samples. *bioRxiv* 201178 (2018) doi:10.1101/201178.
- 639 4. Krumm, N. *et al.* Copy number variation detection and genotyping from exome
640 sequence data. *Genome research* **22**, 1525–1532 (2012).
- 641 5. Li, J. *et al.* CONTRA: copy number analysis for targeted resequencing. **28**, 1307–
642 1313 (2012).
- 643 6. Plagnol, V. *et al.* A robust model for read count data in exome sequencing
644 experiments and implications for copy number variant calling. *Bioinformatics* **28**,
645 2747–2754 (2012).
- 646 7. Backenroth, D. *et al.* CANOES: detecting rare copy number variants from whole
647 exome sequencing data. *Nucleic Acids Research* **42**, e97–e97 (2014).
- 648 8. Fromer, M. & Purcell, S. M. Using XHMM Software to Detect Copy Number Variation
649 in Whole-Exome Sequencing Data. *Current protocols in human genetics* **81**, 7.23.1–
650 7.23.21 (2014).
- 651 9. Demidov, G., Sturm, M. & Ossowski, S. ClinCNV: multi-sample germline CNV
652 detection in NGS data. *bioRxiv* 2022.06.10.495642 (2022)
653 doi:10.1101/2022.06.10.495642.
- 654 10. Yaldiz, B. *et al.* Twist exome capture allows for lower average sequence coverage in
655 clinical exome sequencing. *Human genomics* **17**, (2023).
- 656 11. Gordeeva, V. *et al.* Benchmarking germline CNV calling tools from exome sequencing
657 data. *Scientific Reports* **11**, 14416 (2021).
- 658 12. Tan, R. *et al.* An Evaluation of Copy Number Variation Detection Tools from Whole-
659 Exome Sequencing Data. *Human Mutation* **35**, 899–907 (2014).
- 660 13. Yao, R. *et al.* Evaluation of three read-depth based CNV detection tools using whole-
661 exome sequencing data. *Molecular Cytogenetics* **10**, 1–7 (2017).
- 662 14. Zhao, L., Liu, H., Yuan, X., Gao, K. & Duan, J. Comparative study of whole exome
663 sequencing-based copy number variation detection tools. *BMC Bioinformatics* **21**, 1–
664 10 (2020).
- 665 15. Srivastava, S. *et al.* Meta-analysis and multidisciplinary consensus statement: exome
666 sequencing is a first-tier clinical diagnostic test for individuals with
667 neurodevelopmental disorders. *Genetics in Medicine* **21**, 2413–2421 (2019).

- 668 16. Martinez-Granero, F. *et al.* Comparison of the diagnostic yield of aCGH and genome-
669 wide sequencing across different neurodevelopmental disorders. *npj Genomic*
670 *Medicine* **6**, 25 (2021).
- 671 17. Royer-Bertrand, B. *et al.* CNV Detection from Exome Sequencing Data in Routine
672 Diagnostics of Rare Genetic Disorders: Opportunities and Limitations. (2021)
673 doi:10.3390/genes12091427.
- 674 18. Zurek, B. *et al.* Solve-RD: systematic pan-European data sharing and collaborative
675 analysis to solve rare diseases. *European Journal of Human Genetics* **29**, 1325–1331
676 (2021).
- 677 19. Laurie, S. *et al.* The RD-Connect Genome-Phenome Analysis Platform: Accelerating
678 diagnosis, research, and gene discovery for rare diseases. *Human Mutation* **43**, 717–
679 733 (2022).
- 680 20. Köhler, S. *et al.* The human phenotype ontology in 2021. *Nucleic Acids Research* **49**,
681 D1207–D1217 (2021).
- 682 21. Amberger, J. S., Bocchini, C. A., Scott, A. F. & Hamosh, A. OMIM.org: leveraging
683 knowledge across phenotype-gene relationships. *Nucleic acids research* **47**, D1038–
684 D1043 (2019).
- 685 22. Auton, A. *et al.* A global reference for human genetic variation. *Nature* vol. 526 68–74
686 Preprint at <https://doi.org/10.1038/nature15393> (2015).
- 687 23. Li, H. *Aligning sequence reads, clone sequences and assembly contigs with BWA-*
688 *MEM*. <http://github.com/lh3/bwa>. (2013).
- 689 24. Demidov, G. Methods for detection of germline and somatic copy-number variants in
690 next generation sequencing data. (Universitat Pompeu Fabra, 2019).
- 691 25. Robinson, J. T., Thorvaldsdóttir, H., Wenger, A. M., Zehir, A. & Mesirov, J. P. Variant
692 review with the integrative genomics viewer. *Cancer Research* **77**, e31–e34 (2017).
- 693 26. Geoffroy, V. *et al.* AnnotSV: an integrated tool for structural variations annotation.
694 *Bioinformatics (Oxford, England)* **34**, 3572–3574 (2018).
- 695 27. Blair, V. R. *et al.* Hereditary diffuse gastric cancer: updated clinical practice guidelines.
696 *Lancet Oncol* **21**, e386–e397 (2020).
- 697 28. Guilford, P. *et al.* E-cadherin germline mutations in familial gastric cancer. *Nature* **392**,
698 402–5 (1998).
- 699 29. Oliveira, C. *et al.* Germline CDH1 deletions in hereditary diffuse gastric cancer
700 families. *Hum Mol Genet* **18**, 1545–55 (2009).
- 701 30. Fewings, E. *et al.* Germline pathogenic variants in PALB2 and other cancer-
702 predisposing genes in families with hereditary diffuse gastric cancer without CDH1
703 mutation: a whole-exome sequencing study. *Lancet Gastroenterol Hepatol* **3**, 489–
704 498 (2018).

- 705 31. Oliveira, C., Pinheiro, H., Figueiredo, J., Seruca, R. & Carneiro, F. Familial gastric
706 cancer: genetic susceptibility, pathology, and implications for management. *Lancet*
707 *Oncol* **16**, e60-70 (2015).
- 708 32. São José, C. *et al.* Combined loss of CDH1 and downstream regulatory sequences
709 drive early-onset diffuse gastric cancer and increase penetrance of hereditary diffuse
710 gastric cancer. *Gastric Cancer* (2023) doi:10.1007/s10120-023-01395-0.
- 711 33. Poirier, K. *et al.* CSNK2B splice site mutations in patients cause intellectual disability
712 with or without myoclonic epilepsy. *Human mutation* **38**, 932–941 (2017).
- 713 34. Li, J. *et al.* Germline de novo variants in CSNK2B in Chinese patients with epilepsy.
714 *Scientific reports* **9**, 17909 (2019).
- 715 35. Nakashima, M. *et al.* Identification of de novo CSNK2A1 and CSNK2B variants in
716 cases of global developmental delay with seizures. *Journal of human genetics* **64**,
717 313–322 (2019).
- 718 36. Stessman, H. A. F. *et al.* Targeted sequencing identifies 91 neurodevelopmental-
719 disorder risk genes with autism and developmental-disability biases. *Nature genetics*
720 **49**, 515–526 (2017).
- 721 37. Cheng, H. *et al.* Truncating Variants in NAA15 Are Associated with Variable Levels of
722 Intellectual Disability, Autism Spectrum Disorder, and Congenital Anomalies.
723 *American journal of human genetics* **102**, 985–994 (2018).
- 724 38. Dong, X. *et al.* Clinical exome sequencing as the first-tier test for diagnosing
725 developmental disorders covering both CNV and SNV: a Chinese cohort Diagnostics.
726 *J Med Genet* **57**, 558–566 (2020).
- 727 39. Zhai, Y., Zhang, Z., Shi, P., Martin, D. M. & Kong, | Xiangdong. Incorporation of
728 exome-based CNV analysis makes trio-WES a more powerful tool for clinical
729 diagnosis in neurodevelopmental disorders: A retrospective study. (2021)
730 doi:10.1002/humu.24222.
- 731 40. Pennings, M. *et al.* Copy number variants from 4800 exomes contribute to ~7% of
732 genetic diagnoses in movement disorders, muscle disorders and neuropathies.
733 *European Journal of Human Genetics* **31**, 654–662 (2023).
- 734 41. Pfundt, R. *et al.* Detection of clinically relevant copy-number variants by exome
735 sequencing in a large cohort of genetic disorders. *Genetics in Medicine* **19**,
736 667–675 (2016).
- 737 42. Bullich, G. *et al.* Systematic Collaborative Reanalysis of Genomic Data Improves
738 Diagnostic Yield in Neurologic Rare Diseases. *The Journal of molecular diagnostics*:
739 *JMD* **24**, 529–542 (2022).
- 740 43. Marchuk, D. S. *et al.* Increasing the diagnostic yield of exome sequencing by copy
741 number variant analysis. *PLoS One* **13**, e0209185 (2018).

- 742 44. Sudmant, P. H. *et al.* An integrated map of structural variation in 2,504 human
743 genomes. *Nature* **526**, 75–81 (2015).
- 744 45. Riggs, E. R. *et al.* Technical standards for the interpretation and reporting of
745 constitutional copy-number variants: a joint consensus recommendation of the
746 American College of Medical Genetics and Genomics (ACMG) and the Clinical
747 Genome Resource (ClinGen). *Genetics in Medicine* **22**, 245–257 (2020).
- 748 46. Posey, J. E. *et al.* Resolution of Disease Phenotypes Resulting from Multilocus
749 Genomic Variation. *New England Journal of Medicine* **376**, (2017).
- 750 47. Freeberg, M. A. *et al.* The European Genome-phenome Archive in 2021. *Nucleic
751 Acids Res* **50**, D980–D987 (2022).
- 752 48. Musacchia, F. *et al.* VarGenius-HZD Allows Accurate Detection of Rare Homozygous
753 or Hemizygous Deletions in Targeted Sequencing Leveraging Breadth of Coverage.
754 (2021) doi:10.3390/genes12121979.
- 755 49. Marshall, D. A. *et al.* The value of diagnostic testing for parents of children with rare
756 genetic diseases. *Genetics in Medicine* **21**, 2798–2806 (2019).
- 757
- 758

759 **Corporate author list**

760 ***Solve-RD consortium***

761 **EKUT:** Olaf Riess^{1,2}, Tobias B. Haack¹, Holm Graessner^{1,2}, Birte Zurek^{1,2}, Kornelia Ellwanger^{1,2}, Stephan
762 Ossowski^{1,3}, German Demidov¹, Marc Sturm¹, Julia M. Schulze-Hentrich¹, Rebecca Schüle^{1,2}, Jishu Xu^{4,5},
763 Christoph Kessler^{4,5}, Melanie Kellner^{4,5}, Matthis Synofzik^{4,5}, Carlo Wilke^{4,5}, Andreas Traschütz^{4,5}, Ludger
764 Schöls^{4,5}, Holger Hengel^{4,5}, Holger Lerche¹, Josua Kegele⁶, Peter Heutink^{4,5}

765 **RUMC:** Han Brunner^{7,9}, Hans Scheffer^{7,8}, Nicoline Hoogerbrugge^{7,10}, Alexander Hoischen^{7,10,11}, Peter A.C.
766 't Hoen^{10,12}, Lisenka E.L.M. Vissers^{7,8}, Christian Gilissen^{7,10}, Wouter Steyaert^{7,10}, Karolis Sablauskas⁷,
767 Richarda M. de Voer^{7,10}, Erik-Jan Kamsteeg⁷, Bart van de Warrenburg^{8,11}, Nienke van Os^{8,11}, Iris te Paske⁷
768 ¹⁰, Erik Janssen^{7,10}, Elke de Boer^{7,8}, Marloes Steehouwer⁷, Burcu Yaldiz⁷, Tjitske Kleefstra^{7,8}

769 **University of Leicester:** Anthony J. Brookes¹⁴, Colin Veal¹⁴, Spencer Gibson¹⁴, Vatsalya Maddi¹⁴, Mehdi
770 Mehtarizadeh¹⁴, Umar Riaz¹⁴, Greg Warren¹⁴, Farid Yavari Dizjikan¹⁴, Thomas Shorter¹⁴

771 **UNEW:** Ana Töpf¹³, Volker Straub¹³, Chiara Marini Bettolo¹⁵, Jordi Diaz Manera¹⁵, Sophie Hambleton¹⁶,
772 Karin Engelhardt¹⁶

773 **MUH:** Jill Clayton-Smith^{17,18}, Siddharth Banka^{17,18}, Elizabeth Alexander¹⁸, Adam Jackson^{17,18}

774 **DIJON:** Laurence Faivre^{19,21}, Christel Thauvin^{19,21}, Antonio Vitobello²¹, Anne-Sophie Denommé-Pichon²¹,
775 Yannis Duffourd^{21,22}, Ange-Line Bruel²¹, Christine Peyron^{24,25}, Aurore Pélissier^{24,25}

776 **CNAG-CRG:** Sergi Beltran^{26,27}, Ivo Glynne Gut^{26,27}, Steven Laurie⁶, Davide Piscia²⁶, Leslie Matalonga²⁶,
777 Anastasios Papakonstantinou²⁶, Gemma Bullich²⁶, Alberto Corvo²⁶, Marcos Fernandez-Callejo²⁶, Carles
778 Hernández²⁶, Daniel Picó²⁶, Ida Paramonov²⁶, Hanns Lochmüller²⁶

779 **EURORDIS:** Gulcin Gumus⁸, Virginie Bros-Facer²⁹

780 **INSERM-Orphanet:** Ana Rath³⁰, Marc Hanauer³⁰, David Lagorce³⁰, Oscar Hongnat³⁰, Maroua
781 Chahdil³⁰, Emeline Lebreton³⁰

782 **INSERM-ICM:** Giovanni Stevanin^{31,35}, Alexandra Durr^{31,34,36}, Claire-Sophie Davoine^{31,35}, Léna Guillot-Noel^{31,35},
783 Anna Heinzmann^{31,34,37}, Giulia Coarelli^{31,34,37}

784 **INSERM-CRM:** Gisèle Bonne³⁸, Teresinha Evangelista³⁸, Valérie Allamand³⁸, Isabelle Nelson³⁸, Rabah Ben
785 Yaou^{38,40}, Corinne Metay^{38,41}, Bruno Eymard^{38,39}, Enzo Cohen³⁸, Antonio Atalaia³⁸, Tanya Stojkovic^{38,39}

786 **Univerzita Karlova:** Milan Macek Jr.⁴², Marek Turnovec⁴², Dana Thomasová⁴², Radka Pourová
787 Kremlíková⁴², Vera Franková⁴², Markéta Havlovicová⁴², Petra Lišková^{43,44}, Pavla Doležalová⁴⁵

788 **EMBL-EBI:** Helen Parkinson⁴⁶, Thomas Keane⁴⁶, Mallory Freeberg⁴⁶, Coline Thomas⁴⁶, Dylan Spalding⁴⁶

789 **Jackson Laboratory:** Peter Robinson⁴⁷, Daniel Danis⁴⁷

790 **KCL:** Glenn Robert⁴⁸, Alessia Costa⁴⁹, Christine Patch^{49,50}

791 **UCL-IoN:** Mike Hanna⁵¹, Henry Houlden⁵², Mary Reilly⁵¹, Jana Vandrovцова⁵², Stephanie Efthymiou⁵²,
792 Heba Morsy⁵², Elisa Calì⁵², Francesca Magrinelli⁵³, Sanjay M. Sisodiya⁵⁴, Jonathan Rohrer⁵⁵

- 793 **UCL-ICH**, Francesco Muntoni^{56, 57}, Irina Zaharieva⁵⁶, Anna Sarkozy⁵⁶
- 794 **Universiteit Antwerpen**: Vincent Timmerman^{58, 59}, Jonathan Baets^{60, 61}, Geert de Vries^{59, 60}, Jonathan De
795 Winter⁵⁹⁻⁶¹, Danique Beijer⁵⁸⁻⁶⁰, Peter de Jonghe^{59, 61}, Liedewei Van de Vondel⁵⁸⁻⁶⁰, Willem De Ridder⁵⁹⁻⁶¹,
796 Sarah Weckhuysen^{60, 62}
- 797 **Uni Naples/Telethon UDP**: Vincenzo Nigro^{63, 64}, Margherita Mutarelli^{64, 65}, Manuela Morleo⁶⁴, Michele
798 Pinelli⁶⁴, Alessandra Varavallo⁶⁴, Sandro Banfi^{61, 64}, Annalaura Torella⁶³, Francesco Musacchia^{63, 64}, Giulio
799 Piluso⁶³
- 800 **UNIFE**: Alessandra Ferlini⁶⁶, Rita Selvatici⁶⁶, Francesca Gualandi⁶⁶, Stefania Bigoni⁶⁶, Rachele Rossi⁶⁶,
801 Marcella Neri⁶⁶
- 802 **UKB**: Stefan Aretz^{67, 68}, Isabel Spier^{67, 68}, Anna Katharina Sommer⁶⁷, Sophia Peters⁶⁷
- 803 **IPATIMUP**: Carla Oliveira⁶⁹⁻⁷¹, Jose Garcia-Pelaez^{69, 70, 72}, Rita Barbosa-Matos^{69, 70, 73}, Celina São José^{69, 70, 72},
804 Marta Ferreira^{69, 70, 74}, Irene Gullo^{69-71, 75}, Susana Fernandes⁷⁶, Luzia Garrido⁷⁵, Pedro Ferreira^{69, 70, 77}, Fátima
805 Carneiro^{69-71, 75}
- 806 **UMCG**: Morris A Swertz⁷⁸, Lennart Johansson⁷⁸, Joeri K van der Velde⁷⁸, Gerben van der Vries⁷⁸, Pieter B
807 Neerinx⁷⁸, David Ruvolo⁷⁸, Kristin M Abbott⁷⁹, Wilhemina S Kerstjens Frederikse^{79, 80}, Eveline Zonneveld-
808 Huijssoon^{79, 81}, Dieuwke Roelofs-Prins⁷⁸, Marielle van Gijn^{79, 81}
- 809 **Charité**: Sebastian Köhler⁸²
- 810 **SHU**: Alison Metcalfe^{48, 83}
- 811 **APHP**: Alain Verloes^{84, 85}, Séverine Drunat^{84, 85}, Delphine Heron^{86, 87}, Cyril Mignot^{86, 88}, Boris Keren⁸⁶, Jean-
812 Madeleine de Sainte Agathe⁸⁶
- 813 **CHU Bordeaux**: Caroline Rooryck⁸⁹, Didier Lacombe⁸⁹, Aurelien Trimouille⁹⁰
- 814 **Spain UDP**: Manuel Posada De la Paz⁹¹, Eva Bermejo Sánchez⁹¹, Estrella López Martín⁹¹, Beatriz
815 Martínez Delgado⁹¹, F. Javier Alonso García de la Rosa⁹¹
- 816 **Ospedale Pediatrico Bambino Gesù, Rome**: Andrea Ciolfi⁹², Bruno Dallapiccola⁹², Simone Pizzi⁹²,
817 Francesca Clementina Radio⁹², Marco Tartaglia⁹²
- 818 **University of Siena**: Alessandra Renieri⁹³⁻⁹⁵, Simone Furini^{93, 94}, Chiara Fallerini^{93, 94}, Elisa Benetti^{93, 94}
- 819 **Semmelweis University Budapest**: Peter Balicza⁹⁶, Maria Judit Molnar⁹⁶
- 820 **University of Ljubljana**, Ales Maver⁹⁷, Borut Peterlin⁹⁷
- 821 **University of Lübeck**: Alexander Münchau⁹⁸, Katja Lohmann⁹⁹, Rebecca Herzog^{98, 100}, Martje Pauly^{98, 99}
- 822 **Val d'Hebron Barcelona**: Alfons Macaya^{101, 102}, Ana Cazorro-Gutiérrez¹⁰¹, Belén Pérez-Dueñas¹⁰¹, Francina
823 Munell¹⁰¹, Clara Franco Jarava^{103, 104}, Laura Batlle Masó^{105, 106}, Anna Marcé-Grau¹⁰¹, Roger Colobran^{103, 104, 107}
- 824 **Hospital Sant Joan de Déu Barcelona**: Andrés Nascimento Osorio¹⁰⁸, Daniel Natera de Benito¹⁰⁸
- 825 **University of Freiburg**: Hanns Lochmüller¹⁰⁹⁻¹¹¹, Rachel Thompson¹¹¹, Kiran Polavarapu¹¹¹, Bodo
826 Grimbacher¹¹²⁻¹¹⁶

- 827 **University of Oxford:** David Beeson¹¹⁷, Judith Cossins¹¹⁷
- 828 **Folkhälsan Research Centre:** Peter Hackman¹¹⁸, Mridul Johari¹¹⁸, Marco Savarese¹¹⁸, Bjarne Udd¹¹⁸⁻¹²⁰
- 829 **University of Cambridge:** Rita Horvath¹²¹, Patrick F. Chinnery^{121, 122}, Thiloka Ratnaike¹²³, Fei Gao¹²¹,
830 Katherine Schon^{121, 124}
- 831 **Catalan Institute of Oncology, Barcelona:** Gabriel Capella¹²⁵, Laura Valle¹²⁵
- 832 **KU Munich:** Elke Holinski-Feder¹²⁶, Andreas Laner¹²⁷, Verena Steinke-Lange¹²⁶
- 833 **TU Dresden:** Evelin Schröck¹²⁸, Andreas Rump^{128, 129}
- 834 **Koç University:** Ayşe Nazlı Başak¹³⁰
- 835 **Ghent University Hospital:** Dimitri Hemelsoet^{131, 132}, Bart Dermaut¹³²⁻¹³⁴, Nika Schuermans¹³²⁻¹³⁴, Bruce
836 Poppe¹³²⁻¹³⁴, Hannah Verdin¹³³
- 837 **University Hospital Meyer, Florence:** Davide Mei¹³⁵, Annalisa Vetro¹³⁵, Simona Balestrini^{135, 136}, Renzo
838 Guerrini¹³⁵
- 839 **KU Leuven:** Kristl Claeys^{137, 138}
- 840 **LUMC:** Gijs W.E. Santen¹³⁹, Emilia K. Bijlsma¹³⁹, Mariette J.V. Hoffer¹³⁹, Claudia A.L. Ruivenkamp¹³⁹
- 841 **Ludwig Boltzmann Institute for Rare and Undiagnosed Diseases, Vienna:** Kaan Boztug¹⁴⁰⁻¹⁴⁴, Matthias
842 Haimeil¹⁴⁰⁻¹⁴²
- 843 **Institute of Pathology and Genetics, Gosselies, Belgium:** Isabelle Maystadt^{145, 146}
- 844 **Technical University Munich:** Isabell Cordts¹⁴⁷, Marcus Deschauer¹⁴⁷
- 845 **Neurology/Neurogenetics Laboratory University of Crete, Heraklion, Crete, Greece:** Ioannis
846 Zaganas¹⁴⁸, Evgenia Kokosalı¹⁴⁸, Mathioudakis Lambros¹⁴⁸, Athanasios Evangelıou¹⁴⁹, Martha Spilioti¹⁵⁰,
847 Elisabeth Kapaki¹⁵¹, Mara Bourbouli¹⁵¹
- 848 **IRCCS G. Gaslini:** Pasquale Striano^{152, 153}, Federico Zara^{153, 154}, Antonella Riva^{153, 154}, Michele Iacomino^{154, 155},
849 Paolo Uva¹⁵⁵, Marcello Scala^{152, 153}, Paolo Scudieri^{153, 154}
- 850 **Cliniques universitaires Saint-Luc (CUSL):** Maria-Roberta Cilio¹⁵⁶, Evelina Carpancea¹⁵⁶, Chantal
851 Depondt¹⁵⁷, Damien Lederer¹⁵⁸, Yves Sznajer¹⁵⁹, Sarah Duerinckx¹⁶⁰, Sandrine Mary¹⁵⁸
- 852 **Institute of Human Genetics, University Hospital Essen:** Christel Depienne^{161, 162}, Andreas Roos^{111, 163, 164}
- 853 **University of Luxembourg:** Patrick May¹⁶⁵
- 854 **Affiliations**
- 855 1. Institute of Medical Genetics and Applied Genomics, University of Tübingen, Tübingen, Germany.
- 856 2. Centre for Rare Diseases, University of Tübingen, Tübingen, Germany.
- 857 3. NGS Competence Center Tübingen (NCCT), University of Tübingen, Tübingen, Germany.

- 858 4. Department of Neurodegeneration, Hertie Institute for Clinical Brain Research (HIH), University of
859 Tübingen, Tübingen, Germany.
- 860 5. German Center for Neurodegenerative Diseases (DZNE), Tübingen, Germany.
- 861 6. Department of Neurology and Epileptology, Hertie Institute for Clinical Brain Research (HIH),
862 University of Tübingen, Tübingen, Germany.
- 863 7. Department of Human Genetics, Radboud University Medical Center, Nijmegen, The Netherlands.
- 864 8. Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center,
865 Nijmegen, The Netherlands.
- 866 9. Department of Clinical Genetics, Maastricht University Medical Centre, Maastricht, the
867 Netherlands.
- 868 10. Radboud Institute for Molecular Life Sciences, Nijmegen, The Netherlands.
- 869 11. Department of Internal Medicine and Radboud Center for Infectious Diseases (RCI), Radboud
870 University Medical Center, Nijmegen, the Netherlands.
- 871 12. Center for Molecular and Biomolecular Informatics, Radboud University Medical Center,
872 Nijmegen, the Netherlands.
- 873 13. Department of Neurology, Radboud University Medical Center, Nijmegen, The Netherlands.
- 874 14. Department of Genetics and Genome Biology, University of Leicester, Leicester, UK.
- 875 15. John Walton Muscular Dystrophy Research Centre, Translational and Clinical Research Institute,
876 Newcastle University and Newcastle Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK.
- 877 16. Primary Immunodeficiency Group, Translational and Clinical Research Institute, Newcastle
878 University and Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK.
- 879 17. Division of Evolution, Infection and Genomics, School of Biological Sciences, Faculty of Biology,
880 Medicine and Health, University of Manchester, Manchester M13 9WL, UK.
- 881 18. Manchester Centre for Genomic Medicine, St Mary's Hospital, Manchester University Hospitals
882 NHS Foundation Trust, Health Innovation Manchester, Manchester M13 9WL, UK.
- 883 19. Dijon University Hospital, Genetics Department, Dijon, France.
- 884 20. Dijon University Hospital, Centre of Reference for Rare Diseases: Development disorders and
885 malformation syndromes, Dijon, France.
- 886 21. Inserm - University of Burgundy-Franche Comté, UMR1231 GAD, Dijon, France.
- 887 22. Dijon University Hospital, FHU-TRANSLAD, Dijon, France.
- 888 23. Dijon University Hospital, GIMI institute, Dijon, France.
- 889 24. University of Burgundy-Franche Comté, Dijon Economics Laboratory, Dijon, France.
- 890 25. University of Burgundy-Franche Comté, FHU-TRANSLAD, Dijon, France.

- 891 26. CNAG-CRG, Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and
892 Technology, Baldiri Reixac 4, Barcelona 08028, Spain.
- 893 27. Universitat Pompeu Fabra (UPF), Barcelona, Spain.
- 894 28. EURORDIS-Rare Diseases Europe, Sant Antoni Maria Claret 167 - 08025 Barcelona, Spain.
- 895 29. EURORDIS-Rare Diseases Europe, Plateforme Maladies Rares, 75014 Paris, France.
- 896 30. INSERM, US14 - Orphanet, Plateforme Maladies Rares, 75014 Paris, France.
- 897 31. Institut National de la Santé et de la Recherche Medicale (INSERM) U1127, Paris, France.
- 898 32. Centre National de la Recherche Scientifique, Unité Mixte de Recherche (UMR) 7225, Paris,
899 France.
- 900 33. Unité Mixte de Recherche en Santé 1127, Université Pierre et Marie Curie (Paris 06), Sorbonne
901 Universités, Paris, France.
- 902 34. Institut du Cerveau - ICM, Paris, France.
- 903 35. Ecole Pratique des Hautes Etudes, Paris Sciences et Lettres Research University, Paris, France.
- 904 36. Centre de Référence de Neurogénétique, Hôpital de la Pitié-Salpêtrière, Assistance Publique-
905 Hôpitaux de Paris (AP-HP), Paris, France.
- 906 37. Hôpital de la Pitié-Salpêtrière, Assistance Publique-Hôpitaux de Paris (AP-HP), Paris, France.
- 907 38. Sorbonne Université, Inserm, Institut de Myologie, Centre de Recherche en Myologie, F-75013
908 Paris, France.
- 909 39. AP-HP, Centre de Référence de Pathologie Neuromusculaire Nord, Est, Ile-de-France, Institut de
910 Myologie, G.H. Pitié-Salpêtrière, F-75013 Paris, France.
- 911 40. Institut de Myologie, Equipe Bases de données, G.H. Pitié-Salpêtrière, F-75013 Paris, France.
- 912 41. AP-HP, Unité Fonctionnelle de Cardiogénétique et Myogénétique Moléculaire et Cellulaire, G.H.
913 Pitié-Salpêtrière, F-75013 Paris, France.
- 914 42. Department of Biology and Medical Genetics, Charles University Prague-2nd Faculty of Medicine
915 and University Hospital Motol, Prague, Czech Republic.
- 916 43. Department of Paediatrics and Inherited Metabolic Disorders, First Faculty of Medicine, Charles
917 University and General University Hospital in Prague, Prague, Czech Republic.
- 918 44. Department of Ophthalmology, First Faculty of Medicine, Charles University and General
919 University Hospital in Prague, Prague, Czech Republic.
- 920 45. Centre for Paediatric Rheumatology and Autoinflammatory Diseases, Department of Paediatrics
921 and Inherited Metabolic Disorders, 1st Faculty of Medicine, Charles University and General
922 University Hospital in Prague, Czech Republic.
- 923 46. European Bioinformatics Institute, European Molecular Biology Laboratory, Wellcome Genome
924 Campus, Hinxton, Cambridge, United Kingdom.

- 925 47. Jackson Laboratory for Genomic Medicine, Farmington, CT 06032, USA.
- 926 48. Florence Nightingale Faculty of Nursing, Midwifery & Palliative Care, King's College, London, UK.
- 927 49. Society and Ethics Research, Connecting Science, Wellcome Genome Campus,
928 Hinxton, UK.
- 929 50. Genomics England, Queen Mary University of London, Dawson Hall, EC1M 6BQ, London, UK.
- 930 51. MRC Centre for Neuromuscular Diseases and National Hospital for Neurology and Neurosurgery,
931 UCL Queen Square Institute of Neurology, London, UK.
- 932 52. Department of Neuromuscular Diseases, UCL Queen Square Institute of Neurology, London, UK.
- 933 53. Department of Clinical and Movement Neurosciences, UCL Queen Square Institute of Neurology,
934 University College London, WC1N 3BG.
- 935 54. Department of Clinical and Experimental Epilepsy, UCL Queen Square Institute of Neurology,
936 London, UK.
- 937 55. Dementia Research Centre, Department of Neurodegenerative Disease, UCL Queen Square
938 Institute of Neurology, London, UK.
- 939 56. Dubowitz Neuromuscular Centre, UCL Great Ormond Street Hospital, London, UK.
- 940 57. NIHR Great Ormond Street Hospital Biomedical Research Centre, London, United Kingdom.
- 941 58. Peripheral Neuropathy Research Group, University of Antwerp, Antwerp, Belgium.
- 942 59. Laboratory of Neuromuscular Pathology, Institute Born-Bunge, University of Antwerp,
943 Antwerpen, Belgium.
- 944 60. Translational Neurosciences, Faculty of Medicine and Health Sciences, University of Antwerp,
945 Belgium.
- 946 61. Neuromuscular Reference Centre, Department of Neurology, Antwerp University Hospital,
947 Antwerpen, Belgium.
- 948 62. VIB-CMN, Applied and Translational Neurogenomics Group.
- 949 63. Dipartimento di Medicina di Precisione, Università degli Studi della Campania "Luigi Vanvitelli",
950 Napoli, Italy.
- 951 64. Telethon Institute of Genetics and Medicine, Pozzuoli, Italy.
- 952 65. Istituto di Scienze Applicate e Sistemi Intelligenti "E.Caianiello" - ISASI -CNR.
- 953 66. Unit of Medical Genetics, Department of Medical Sciences, University of Ferrara, Italy.
- 954 67. Institute of Human Genetics, Medical Faculty, University of Bonn, Bonn, Germany.
- 955 68. Center for Hereditary Tumor Syndromes, University Hospital Bonn, Bonn, Germany.

- 956 69. i3S - Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Portugal.
- 957 70. IPATIMUP - Institute of Molecular Pathology and Immunology of the University of Porto,
958 Portugal.
- 959 71. Faculty of Medicine, University of Porto, Portugal.
- 960 72. Doctoral Programme in Biomedicine, Faculty of Medicine, University of Porto, Portugal.
- 961 73. Doctoral Programme in BiotechHealth, School of Medicine and Biomedical Sciences, University of
962 Porto, Portugal.
- 963 74. Doctoral Programme in Computer Science, Faculty of Sciences, University of Porto, Portugal.
- 964 75. CHUSJ, Centro Hospitalar e Universitário de São João, Porto, Portugal.
- 965 76. Department of Genetics, Faculty of Medicine, University of Porto, Portugal.
- 966 77. Faculty of Sciences, University of Porto, Portugal.
- 967 78. Department of Genetics, Genomics Coordination Center, University Medical Center Groningen,
968 University of Groningen, Groningen, The Netherlands.
- 969 79. Department of Genetics, University Medical Center Groningen, University of Groningen,
970 Groningen, The Netherlands.
- 971 80. ERN-GENTURIS.
- 972 81. ERN-RITA: European Reference Network for Immunodeficiency, Autoinflammatory, Autoimmune
973 and Paediatric Rheumatic diseases, Utrecht, Netherlands.
- 974 82. Ada Health GmbH, Karl-Liebknecht-Str. 1, 10178 Berlin, Germany.
- 975 83. College of Health, Well-being and Life-Sciences, Sheffield Hallam University, Sheffield, UK.
- 976 84. Dept of Genetics, Assistance Publique-Hôpitaux de Paris - Université de Paris, Robert DEBRE
977 University Hospital, 48 bd SERURIER, Paris, France.
- 978 85. INSERM UMR 1141 "NeuroDiderot", Hôpital Robert DEBRE, Paris, France.
- 979 86. Department of Genetics, Assistance Publique-Hôpitaux de Paris - Sorbonne Université, Pitié-
980 Salpêtrière University Hospital, 83 Boulevard de l'Hôpital, Paris, France.
- 981 87. Reference center of rare diseases "intellectual disability of rare causes", Paris, France.
- 982 88. Institut du Cerveau (ICM), UMR S 1127, Inserm U1127, CNRS UMR 7225, Sorbonne Université,
983 75013, Paris, France.
- 984 89. Univ. Bordeaux, MRGM INSERM U1211, CHU de Bordeaux, Service de Génétique Médicale , F-
985 33000 Bordeaux, France.
- 986 90. Laboratoire de Génétique Moléculaire, Service de Génétique Médicale, CHU Bordeaux – Hôpital
987 Pellegrin, Place Amélie Raba Léon, 33076 Bordeaux Cedex, France.

- 988 91. Institute of Rare Diseases Research, Spanish Undiagnosed Rare Diseases Cases Program
989 (SpainUDP) & Undiagnosed Diseases Network International (UDNI), Instituto de Salud Carlos III,
990 Madrid, Spain.
- 991 92. Molecular Genetics and Functional Genomics, Ospedale Pediatrico Bambino Gesù, IRCCS, Rome,
992 Italy.
- 993 93. Med Biotech Hub and Competence Center, Department of Medical Biotechnologies, University of
994 Siena, Italy.
- 995 94. Medical Genetics, University of Siena, Italy.
- 996 95. Genetica Medica, Azienda Ospedaliero-Universitaria Senese, Italy.
- 997 96. Institute of Genomic Medicine and Rare Diseases, Semmelweis University, Budapest, Hungary.
- 998 97. Clinical Institute of Genomic Medicine, University Medical Centre Ljubljana, Slovenia.
- 999 98. Institute of Systems Motor Science, University of Lübeck, Ratzeburger Allee 160, 23562, Lübeck,
1000 Germany.
- 1001 99. Institute of Neurogenetics, University of Lübeck, Ratzeburger Allee 160, 23562, Lübeck,
1002 Germany.
- 1003 100. Department of Neurology, University Hospital Schleswig Holstein, Ratzeburger Allee 160,
1004 23562, Lübeck, Germany.
- 1005 101. Pediatric Neurology Research Group, Vall d'Hebron Research Institute, Universitat Autònoma
1006 de Barcelona, Barcelona, Spain.
- 1007 102. Institute of Neuroscience, Universitat Autònoma de Barcelona, Barcelona, Spain.
- 1008 103. Diagnostic Immunology Research Group, Vall d'Hebron Research Institute (VHIR), Barcelona,
1009 Spain.
- 1010 104. Immunology Division, Genetics Department. Vall d'Hebron University Hospital (HUVH),
1011 Barcelona, Spain.
- 1012 105. Infection in Immunocompromised Pediatric Patients Research Group, Vall d'Hebron Research
1013 Institute (VHIR), Barcelona, Spain.
- 1014 106. Pediatric Infectious Diseases and Immunodeficiencies Unit, Vall d'Hebron University Hospital
1015 (HUVH), Barcelona, Spain.
- 1016 107. Immunology Unit. Department of Cell Biology, Physiology and Immunology. Autonomous
1017 University of Barcelona (UAB), Bellaterra, Spain.
- 1018 108. Neuromuscular Disorders Unit, Department of Pediatric Neurology. Hospital Sant Joan de Déu,
1019 Barcelona, Spain
- 1020 109. Department of Neuropediatrics and Muscle Disorders, Medical Center, Faculty of Medicine,
1021 University of Freiburg, Freiburg, Germany.

- 1022 110. Centro Nacional de Análisis Genómico (CNAG-CRG), Center for Genomic Regulation, Barcelona
1023 Institute of Science and Technology (BIST), Barcelona, Spain.
- 1024 111. Children's Hospital of Eastern Ontario Research Institute, University of Ottawa, Ottawa, Canada.
- 1025 112. Institute for Immunodeficiency, Center for Chronic Immunodeficiency (CCI), Medical Center,
1026 Faculty of Medicine, Albert-Ludwigs-University of Freiburg, Germany.
- 1027 113. Clinic of Rheumatology and Clinical Immunology, Center for Chronic Immunodeficiency (CCI),
1028 Medical Center, Faculty of Medicine, Albert-Ludwigs-University of Freiburg, Germany.
- 1029 114. DZIF – German Center for Infection Research, Satellite Center Freiburg, Germany.
- 1030 115. CIBSS – Centre for Integrative Biological Signalling Studies, Albert-Ludwigs University, Freiburg,
1031 Germany.
- 1032 116. RESIST – Cluster of Excellence 2155 to Hanover Medical School, Satellite Center Freiburg,
1033 Germany.
- 1034 117. Nuffield Department of Clinical Neurosciences, University of Oxford, UK.
- 1035 118. Folkhälsan Research Centre and Medicum, University of Helsinki, Helsinki, Finland.
- 1036 119. Tampere Neuromuscular Center, Tampere, Finland.
- 1037 120. Vasa Central Hospital, Vaasa, Finland.
- 1038 121. Department of Clinical Neurosciences, University of Cambridge, Cambridge, UK.
- 1039 122. Medical Research Council Mitochondrial Biology Unit, University of Cambridge, Cambridge, UK.
- 1040 123. Department of Paediatrics, University of Cambridge, Cambridge, UK.
- 1041 124. East Anglian Medical Genetics Service, Cambridge University Hospitals NHS Foundation Trust,
1042 Cambridge, UK.
- 1043 125. Bellvitge Biomedical Research Institute (IDIBELL), Barcelona, Spain.
- 1044 126. Medizinische Klinik und Poliklinik IV – Campus Innenstadt, Klinikum der Universität München,
1045 Munich, Germany.
- 1046 127. MGZ - Medical Genetics Center, Munich, Germany.
- 1047 128. Institute of Clinical Genetics, University Hospital Carl Gustav Carus, Technical University
1048 Dresden, Dresden, Germany.
- 1049 129. Center for Personalized Oncology, University Hospital Carl Gustav Carus, Technical University
1050 Dresden, Dresden, Germany.
- 1051 130. Koç University, School of Medicine, Translational Medicine Research Center, KUTTAM-NDAL
1052 Istanbul Turkey.
- 1053 131. Dpt. of Neurology, Ghent University Hospital, Ghent, Belgium.

- 1054 132. Program for Undiagnosed Rare Diseases (UD-PrOZA), Ghent University Hospital, Ghent,
1055 Belgium.
- 1056 133. Center for Medical Genetics, Ghent University Hospital, Ghent, Belgium.
- 1057 134. Department of Biomolecular Medicine, Faculty of Medicine and Health Sciences, Ghent
1058 University, Ghent, Belgium.
- 1059 135. Neuroscience Department, Children's Hospital A. Meyer-University of Florence, 50139,
1060 Florence, Italy.
- 1061 136. Department of Clinical and Experimental Epilepsy, UCL Queen Square Institute of Neurology,
1062 and Chalfont Centre for Epilepsy, Gerrard Cross, UK.
- 1063 137. Department of Neurology, University Hospitals Leuven, Leuven, Belgium.
- 1064 138. Laboratory for Muscle Diseases and Neuropathies, Department of Neurosciences, and Leuven
1065 Brain Institute (LBI), KU Leuven - University of Leuven, Leuven, Belgium.
- 1066 139. Department of Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands.
- 1067 140. Ludwig Boltzmann Institute for Rare and Undiagnosed Diseases, Vienna, Austria.
- 1068 141. St. Anna Children's Cancer Research Institute (CCRI), Vienna, Austria.
- 1069 142. CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna,
1070 Austria.
- 1071 143. Department of Pediatrics and Adolescent Medicine, Medical University of Vienna, Vienna,
1072 Austria.
- 1073 144. St. Anna Children's Hospital, Department of Pediatrics and Adolescent Medicine, Medical
1074 University of Vienna, Vienna, Austria.
- 1075 145. Centre de Génétique Humaine, Institut de Pathologie et de Génétique, Gosselies, Belgium.
- 1076 146. Département de Médecine, Université de Namur (Unamur), Namur, Belgique.
- 1077 147. Department of Neurology, Klinikum rechts der Isar, Technical University Munich, Munich,
1078 Germany.
- 1079 148. Neurology / Neurogenetics Laboratory University of Crete, Heraklion, Crete, Greece.
- 1080 149. Aristotle University of Thessaloniki, Thessaloniki, Greece.
- 1081 150. 1st Department of Neurology, Aristotle University of Thessaloniki, University General Hospital
1082 of Thessaloniki, AHEPA, Thessaloniki, Greece.
- 1083 151. Neurochemistry and Biomarker Unit, 1st Department of Neurology, School of Medicine,
1084 National and Kapodistrian University of Athens, Eginition Hospital, Athens, Greece.
- 1085 152. Pediatric Neurology and Muscular Disease Unit, IRCCS Istituto Giannina Gaslini, Genoa, Italy.

- 1086 153. Department of Neurosciences, Rehabilitation, Ophthalmology, Genetics, Maternal and Child
1087 Health, University of Genoa, Genoa, Italy.
- 1088 154. Unit of Medical Genetics, IRCCS Istituto Giannina Gaslini, Genoa, Italy.
- 1089 155. Clinical Bioinformatics, IRCCS Istituto Giannina Gaslini, Genoa, Italy.
- 1090 156. Pediatric Neurology Department, Saint-Luc University Hospital, Université Catholique de
1091 Louvain, Brussels, Belgium.
- 1092 157. Neurology Department, Erasme Hospital, Université Libre de Bruxelles , Bruxelles, Belgium.
- 1093 158. Institute of Pathology and Genetics, Charleroi, Belgium.
- 1094 159. Human Genetics Department, Saint-Luc University Hospital, Université Catholique de Louvain,
1095 Brussels, Belgium.
- 1096 160. Institute of Interdisciplinary Research in Human and Molecular Biology, Human Genetics,
1097 IRIBHM, Université Libre de Bruxelles, Brussels, Belgium.
- 1098 161. Institute of Human Genetics, University Hospital Essen, University Duisburg-Essen, Essen,
1099 Germany.
- 1100 162. Institut du Cerveau et de la Moelle épinière (ICM), Sorbonne Université, UMR S 1127, Inserm
1101 U1127, CNRS UMR 7225, F-75013 Paris, France.
- 1102 163. Department of Pediatric Neurology, Developmental Neurology and Social Pediatrics, Children's
1103 Hospital University of Essen, Essen, Germany.
- 1104 164. Department of Neurology, Heimer Institute for Muscle Research, University Hospital
1105 Bergmannsheil, Ruhr-University Bochum, 44789 Bochum, Germany.
- 1106 165. Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette,
1107 Luxembourg.